The Misspecified Cramér-Rao Bound and its Application to Scatter Matrix Estimation in Complex Elliptically Symmetric Distributions

Stefano Fortunati[1,2], Fulvio Gini[1,2], Maria S. Greco[1,2]

[1]Dipartimento di Ingegneria dell'Informazione, University of Pisa
Via G. Caruso 16, 56122 Italy

[2]CNIT RaSS (Radar and Surveillance System) National Laboratory, Pisa, Italy

E-mails: {stefano.fortunati,f.gini,m.greco}@iet.unipi.it

Abstract- *This paper focuses on the application of recent results on lower bounds under misspecified models to the estimation of the scatter matrix of Complex Elliptically Symmetric (CES) distributed random vectors. Buildings upon the original works of Q. H. Vuong and Richmond-Horowitz, a lower bound, named Misspecified Cramér-Rao bound (MCRB), for the error covariance matrix of any unbiased (in a proper sense) estimator of a deterministic parameter vector under misspecified models is introduced. Then, we show how to apply these results to the problem of estimating the scatter matrix of CES distributed data under data mismodeling. In particular, the performance of the maximum likelihood (ML) estimator are compared, under mismatched conditions, with the MCRB and with the classical CRB in some relevant study cases.*

Keywords: *Maximum Likelihood estimator, Matrix estimation, Complex Elliptically Symmetric distribution, misspecified model, Cramér-Rao Lower Bound, Kullback-Leibler divergence, Huber sandwich matrix.*

I. Introduction

The problem of estimating a deterministic parameter vector from a set of acquired data is ubiquitous in signal processing applications. A fundamental assumption underlying most estimation problems is that the true data model and the model assumed to derive an estimation algorithm are the same, that is, the model is correctly specified. However, a certain amount of mismatch is often inevitable in practice. Among others, the model mismatch can be due to an imperfect knowledge of the true data model or to the need to fulfill some operative constraints on the estimation algorithm (processing time, simple hardware implementation, and so on). In the statistical and econometric literature, much attention has been devoted to the behavior of the Maximum Likelihood (ML) estimator under mismatched conditions ([3, 4, 5, 6], and recently [7]). In particular, Huber [3] and White [4] have shown that the asymptotic distribution of the ML estimator under misspecified models is a Gaussian one whose mean value is the minimizer (also called *pseudo-true parameter vector* in [1]) of the Kullback-Leibler (KL) divergence between the true and the assumed data distributions and the covariance matrix is given by the so-called Huber "sandwich" matrix. A milestone on the misspecification analysis is the very comprehensive book [6]. It provides an excellent and insightful discussion about statistical inference in the presence of distributional misspecification, with a focus on estimation and hypotheses testing problems. A different mismodeling related to the dynamic of the acquired data has been investigated in [8]. In particular, in [8] the asymptotic performance of the ML estimator and of the generalized likelihood ratio test (GLRT) is derived under the assumption of independent identically distribution (i.i.d.) samples, when in the actual model the data vector are dependent.

In conjunction with the asymptotic analysis, a question that naturally arises is if it is possible to establish a lower bound on the error covariance matrix of a certain class of mismatched estimators. When the parametric model is correctly specified, a few of such lower bounds exist; one of these is the well-known Cramér-Rao Bound (CRB). Recent works attempt to generalize the Cramér-Rao inequality in the presence of model misspecification. In [9], a Bayesian bound of the Ziv-Zakai type has been derived under model mismatch conditions restricted to misparameterized zero mean complex Gaussian distributions. More recently, Richmond and Horowitz [2] derived a covariance inequality for deterministic complex parameter vector in the presence of model misspecification. Moreover, in [2] a generalization to the mismatched case of the Slepian-Bangs formula for the evaluation of the CRB for multivariate complex Gaussian distributed observations is also derived. To the best of our knowledge, [2] represents the first attempt to introduce in the Signal Processing community an organic framework for deriving a covariance inequality of the Cramér-Rao type in the presence of model mismatch. However, the proof proposed in [2] limits the applicability of the derived misspecified CRB (MCRB) to a very restricted class of estimators. In an unrecognized (at least among the Signal Processing community) working paper [1], Q. H. Vuong had proposed the same bound derived in [2] but with a different proof that allows for its applicability to a wider class of estimators, i.e. the class of unbiased (in a proper sense) estimators. Finally, to conclude the review of the literature on misspecified bound, we refer to the recent paper [10], where the authors adopted a different definition of unbiasedness and a different score function than in [1] and [2].

The aim of this paper is twofold: in the first part, we provide a concise but comprehensive review of the main findings about the Misspecified Cramér-Rao Bound (MCRB) and the *Missmatched Maximum Likelihood* (MML) estimator. In the second part, we discuss the use of this bound first through an illustrative example, i.e. the estimation of the variance of a set of one-dimensional Gaussian data under misspecified mean value, and then through its application in a classical radar signal processing problem: the estimation of the disturbance covariance (scatter) matrix for adaptive detection algorithms. We put this classical radar problem in the more general context of the estimation of the scatter matrix in the Complex Elliptically Symmetric (CES) distribution family.

In the following, a formal description of the estimation problem under mismatched conditions is provided. Let $\mathbf{x}_m \in \mathbb{C}^N$ be a *N*-dimensional random vector representing the outcome of a random experiment (i.e. the observation vector) whose probability density function (p.d.f.) is known to belong to a family $\mathcal{P}$. A *structure T* is a set of hypotheses, which implies a unique p.d.f. in $\mathcal{P}$ for $\mathbf{x}_m$. Such p.d.f. is indicated with $p_X(\mathbf{x}_m;T)$ ([11], [12]). The set of all the a priori possible structures is called a *model*. We assume that the p.d.f. of the random vector $\mathbf{x}_m$ has a parametric representation, i.e. we assume that every structure T is parameterized by a *d*-dimensional vector $\boldsymbol{\tau}$ and that the model is described by a compact subspace $\mathrm{T} \subset \mathbb{R}^d$.

The common assumption underlying any practical estimation problem is the perfect knowledge of the p.d.f. $p_X\left(\mathbf{x};\bar{\boldsymbol{\theta}}\right)$ that characterizes the i.i.d. observations, $\mathbf{x} = \left\{\mathbf{x}_m\right\}_{m=1}^M$, except for the value of the parameter vector $\bar{\boldsymbol{\theta}} \in \mathrm{T}$. However, a certain amount of mismatch between the true p.d.f. of the observations and the p.d.f. that we assume to derive an estimator of the parameters of interest is always present. Specifically, suppose that the true parametric p.d.f. of the observations $p_X\left(\mathbf{x};\bar{\boldsymbol{\theta}}\right)$ and the assumed p.d.f. $f_X\left(\mathbf{x};\boldsymbol{\theta}\right)$ belong to two (generally different) families of p.d.f.'s, $\mathcal{P}$ and $\mathcal{F}$. Since in practical situations the true model is unknown, i.e. we have no prior information on the particular parameterization of the true distribution, in the following we refer to $p_X\left(\mathbf{x};\bar{\boldsymbol{\theta}}\right)$ only as $p_X\left(\mathbf{x}\right)$ in order to highlight the fact the neither the model, nor the true parameter vector $\bar{\boldsymbol{\theta}}$ are accessible by a mismatched estimator [2].

Suppose then that the (possibly complex) *M* measurement vectors are sampled from a particular p.d.f. belonging to $\mathcal{P}$, i.e. $\mathbf{x}_m \sim p_X(\mathbf{x}_m)$, for *m*=1,2,…,*M*. Suppose now that the true distribution $p_X\left(\mathbf{x}\right)$ does not belong to $\mathcal{F}$. In the rest of this paper, we indicate this mismatch between the true and assumed p.d.f.'s as *misspecified model*.

The rest of the paper is organized as follows. In Sect. II, the main theoretical results on the MCRB and the MML estimator are reviewed and discussed. In Sect. III, a simple example useful to better clarify the theoretical findings of Sect. II and how they should be applied is provided. Sect. IV focuses on the application of the MML estimator and of the MCRB to the estimation of the scatter matrix in the CES distribution family. Sect. V summarizes our conclusions.

II. A lower bound in the presence of misspecified models: the MCRB

The aim of this section is to provide an organic view of the findings in [1] and [2]. Starting from [1], we first provide a list of regularity conditions that are not only a fundamental prerequisite for the derivation of the MCRB, but allow to better understand the nature and the usefulness of the bound. Then, we provide the expression of the MCRB and the class of estimators to which it applies (Theorem 4.1 [1]). A discussion on the main differences between the general proof in [1] and the one given in [2] is also provided. Finally, we conclude this section by introducing the MML estimator, its asymptotic properties and their link with the MCRB [3], [4].

II.A Regular models

Let $\mathbf{x} = \left\{\mathbf{x}_m\right\}_{m=1}^M \in \mathbb{C}^N$ be a set of i.i.d. *N*-dimensional random vector and let $p_X(\mathbf{x})$ the true p.d.f. of $\mathbf{x}$. Let $\mathcal{F} = \left\{ f_{\boldsymbol{\theta}} \,\middle|\, f_X(\mathbf{x};\boldsymbol{\theta}) \text{ is a pdf } \forall \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p \right\}$ be a family of parametric p.d.f.s that possibly does not contain $p_X(\mathbf{x})$.

Assumption A1: For every $\boldsymbol{\theta} \in \Theta$, the functions $\left|\ln f_X(\mathbf{x};\boldsymbol{\theta})\right|$, $\left|\partial \ln f_X(\mathbf{x};\boldsymbol{\theta})/\partial\theta_i\right|$ and $\left|\partial^2 \ln f_X(\mathbf{x};\boldsymbol{\theta})/\partial\theta_i\partial\theta_j\right|$, $i,j = 1,\ldots,p$ are dominated by a function $m(\mathbf{x})$ independent of $\boldsymbol{\theta}$ and square-integrable with respect to $p_X(\mathbf{x})$.

Assumption A2: (a) The function $E_p\left\{\ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right)\right\} = \int \ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right) p_X(\mathbf{x}_m) d\mathbf{x}_m$ has a unique maximum on Θ at an interior point $\boldsymbol{\theta}_0$. (b) The matrix $\mathbf{A}_{\boldsymbol{\theta}_0}$ whose entries are

$$\left[\mathbf{A}_{\boldsymbol{\theta}_0}\right]_{ij} \triangleq \left[E_p\left\{\nabla_{\boldsymbol{\theta}_0}\nabla_{\boldsymbol{\theta}_0}^T \ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}_0\right)\right\}\right]_{ij} = E_p\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right)\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\} \quad (1)$$

is non-singular. Note that $\nabla_{\boldsymbol{\theta}_0} u(\boldsymbol{\theta}_0)$ indicates the gradient (column) vector of the scalar function *u* evaluated in $\boldsymbol{\theta}_0$. This can be recognized also as the identifiability condition (see [11, 12, 13]) for $\boldsymbol{\theta}_0$. The interior point $\boldsymbol{\theta}_0$ can be equivalently seen as the point that minimizes the Kullback-

Leibler divergence between the true distribution $p_X(\mathbf{x}_m)$ and the assumed distribution $f_X(\mathbf{x}_m;\boldsymbol{\theta})$ [3], [4]:

$$\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min}\left\{D\left(p\|f_{\boldsymbol{\theta}}\right)\right\} = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min}\left\{-E_p\left\{\ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right)\right\}\right\}, \tag{2}$$

where

$$D\left(p\|f_{\boldsymbol{\theta}}\right) \triangleq E_p\left\{\ln\left(\frac{p_X(\mathbf{x})}{f_X(\mathbf{x};\boldsymbol{\theta})}\right)\right\} = \int \ln\left(\frac{p_X(\mathbf{x})}{f_X(\mathbf{x};\boldsymbol{\theta})}\right) p_X(\mathbf{x})d\mathbf{x}. \tag{3}$$

Assumption A3: There exists a neighborhood Γ of $\boldsymbol{\theta}_0$ such that for every $\boldsymbol{\theta}\in\Gamma$ the functions $(f_X(\mathbf{x};\boldsymbol{\theta}_0))^{-1}\left|\partial \ln f_X(\mathbf{x};\boldsymbol{\theta})/\partial\theta_i\right|, i=1,\ldots,p$ are dominated by a function $m(\mathbf{x})$ independent of $\boldsymbol{\theta}$ and square-integrable with respect to $p_X(\mathbf{x})$.

Assumptions A1 and A3 essentially allow differentiation under the integral sign of the expectation of any random variable or vector with finite variance that we will encounter in the rest of the paper. Assumption A2 ensures the existence and the uniqueness of the so-called *pseudo-true parameters vector* $\boldsymbol{\theta}_0$. As we will see soon, $\boldsymbol{\theta}_0$ plays a key role both in the definition of the MCRB and of the MML.

Definition 1 *(regular models)* [1]: A parametric model $\mathcal{F}$ is regular with respect to (w.r.t.) a p.d.f. $p_X(\mathbf{x})$ if Assumptions A1-A3 hold. It is regular w.r.t. a family $\mathcal{P}$ if it is regular w.r.t. every p.d.f. in $\mathcal{P}$. It is regular if it is regular w.r.t. every p.d.f. in $\mathcal{F}$.

II.B MS-unbiased estimators and the Misspecified Cramér-Rao Bound

After setting the necessary regularity conditions, a covariance inequality in the presence of misspecified regular models can be defined. First, the concept of *misspecified unbiasedness*, in short MS-unbiasedness, has to be introduced.

Definition 2 *(MS-unbiasedness)* [1]: Let $\mathcal{P}$ be a family of p.d.f.s w.r.t. which the (misspecified) parametric model $\mathcal{F}$ is regular. Let $\hat{\boldsymbol{\theta}}(\mathbf{x})$ be an estimator derived under the misspecified model $\mathcal{F}$ from the i.i.d. observations $\mathbf{x}=\{\mathbf{x}_m\}_{m=1}^M$. Then, $\hat{\boldsymbol{\theta}}(\mathbf{x})$ is an MS-unbiased estimator of $\boldsymbol{\theta}_0$ if and only if:

$$E_p\left\{\hat{\boldsymbol{\theta}}(\mathbf{x})\right\} = \int \hat{\boldsymbol{\theta}}(\mathbf{x}) p_X(\mathbf{x})d\mathbf{x} = \boldsymbol{\theta}_0, \quad \forall p_X(\mathbf{x})\in\mathcal{P}. \tag{4}$$

It is easy to show that this definition is consistent with the classical definition of unbiasedness. In fact, when the model $\mathcal{F}$ is correctly specified, there exists a $\bar{\boldsymbol{\theta}}\in\Theta$ such that $p_X(\mathbf{x}) = f_X(\mathbf{x};\bar{\boldsymbol{\theta}})$ for every $\mathbf{x}$. Then, from Assumption A2, $\bar{\boldsymbol{\theta}}=\boldsymbol{\theta}_0$ and finally eq. (4) reduces to $E_{f_{\bar{\boldsymbol{\theta}}}}\left\{\hat{\boldsymbol{\theta}}(\mathbf{x})\right\} = \int \hat{\boldsymbol{\theta}}(\mathbf{x}) f_X(\mathbf{x};\bar{\boldsymbol{\theta}})d\mathbf{x} = \bar{\boldsymbol{\theta}}$ that is exactly the standard definition of unbiasedness.

At this point, a lower bound in the presence of (regular) misspecified models can be introduced.

Theorem 1 *(The Misspecified Cramér-Rao Bound, MCRB)* [1], [2]: *Let $\mathcal{F}$ be a parametric model. Let $\mathcal{P}(\mathcal{F})$ be the family of all p.d.f.s w.r.t. which $\mathcal{F}$ is regular. Suppose that $\mathcal{P}(\mathcal{F})$ is not empty. Let $\hat{\boldsymbol{\theta}}(\mathbf{x})$ be an MS-unbiased estimator derived under the misspecified model $\mathcal{F}$ from the i.i.d. observations $\mathbf{x}=\{\mathbf{x}_m\}_{m=1}^M$. Then, for every $p_X(\mathbf{x})$ in $\mathcal{P}(\mathcal{F})$:*

$$\mathbf{C}\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\boldsymbol{\theta}_0\right) \geq \frac{1}{M}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1} \triangleq \mathrm{MCRB}\left(\boldsymbol{\theta}_0\right) \tag{5}$$

where

$$\mathbf{C}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\boldsymbol{\theta}_0\right) \triangleq E_p\left\{\left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0\right)\left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0\right)^T\right\} \tag{6}$$

is the error covariance matrix of $\hat{\boldsymbol{\theta}}(\mathbf{x})$, the matrix $\mathbf{A}_{\boldsymbol{\theta}_0}$ has been defined in eq. (1) and $\mathbf{B}_{\boldsymbol{\theta}_0}$ is a matrix whose entries are defined as:

$$\left[\mathbf{B}_{\boldsymbol{\theta}_0}\right]_{ij} \triangleq \left[E_p\left\{\nabla_{\boldsymbol{\theta}_0}\ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}_0\right)\nabla_{\boldsymbol{\theta}_0}^T \ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}_0\right)\right\}\right]_{ij} = E_p\left\{\left.\frac{\partial \ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right)}{\partial\theta_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\cdot\left.\frac{\partial \ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right)}{\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\}. \tag{7}$$

Following [2], *we refer to the right side of eq. (7) as the Misspecified Cramér-Rao Bound (MCRB).*

The proof of this Theorem can be found in [1]. It can be noted that the hypothesis that $\mathcal{P}(\mathcal{F})$ is not empty is not so strong. In fact, it requires that there exists at least one p.d.f. $p_X(\mathbf{x}_m)$ for which, from Assumption A2, the point $\boldsymbol{\theta}_0$ exists [1]. In the next sections, we provide three examples in which it is possible to evaluate $\boldsymbol{\theta}_0$, and then the MCRB applies. Other relevant signal processing problems in which the pseudo-true vector $\boldsymbol{\theta}_0$ can be evaluated are discussed in [2].

In this case, it is easy to verify that the MCRB is consistent with the classical CRB. As for the unbiasedness, when the model $\mathcal{F}$ is correctly specified, $p_X(\mathbf{x}) = f_X(\mathbf{x};\bar{\boldsymbol{\theta}})$ for some $\bar{\boldsymbol{\theta}}\in\Theta$. Then, the matrices $-\mathbf{A}_{\bar{\boldsymbol{\theta}}}$ and $\mathbf{B}_{\bar{\boldsymbol{\theta}}}$ are equal and correspond to the classical Fisher Information Matrix (FIM), and finally:

$$\mathbf{C}_{f_{\bar{\boldsymbol{\theta}}}}\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\bar{\boldsymbol{\theta}}\right) \geq \frac{1}{M}\mathbf{A}_{\bar{\boldsymbol{\theta}}}^{-1}\mathbf{B}_{\bar{\boldsymbol{\theta}}}\mathbf{A}_{\bar{\boldsymbol{\theta}}}^{-1} = -\frac{1}{M}\mathbf{A}_{\bar{\boldsymbol{\theta}}}^{-1} = \frac{1}{M}\mathbf{B}_{\bar{\boldsymbol{\theta}}} = -\frac{1}{M}\left(E_{f_{\bar{\boldsymbol{\theta}}}}\left\{\nabla_{\bar{\boldsymbol{\theta}}}\nabla_{\bar{\boldsymbol{\theta}}}^T \ln f_X\left(\mathbf{x}_m;\bar{\boldsymbol{\theta}}\right)\right\}\right)^{-1}$$
$$= \frac{1}{M}\left(E_{f_{\bar{\boldsymbol{\theta}}}}\left\{\nabla_{\bar{\boldsymbol{\theta}}} \ln f_X\left(\mathbf{x}_m;\bar{\boldsymbol{\theta}}\right)\nabla_{\bar{\boldsymbol{\theta}}}^T \ln f_X\left(\mathbf{x}_m;\bar{\boldsymbol{\theta}}\right)\right\}\right)^{-1}, \tag{8}$$

which represents the classical Cramér-Rao inequality for an unbiased estimator.

Remark 1: The statement and the proof of Theorem 1 given in [1] consider only the case of *real* parameter space, i.e. $\Theta \subseteq \mathbb{R}^p$. However, as shown in [2], the derivation can be easily extended to the complex case, i.e. when $\Theta \subseteq \mathbb{C}^p$. This is because all the p.d.f.s are *real functions of complex variables* (**x** and **θ**), so we do not need sophisticated holomorphic calculus to generalize the derivatives w.r.t. a complex parameter vector **θ**. Insightful procedures, useful to generalize the Cramér-Rao inequality in the complex case, are discussed in [14], [15] and [16].

Remark 2: In order to evaluate the MCRB of (5), the knowledge of the true p.d.f. $p_X(\mathbf{x})$ is required. However, this has not to be seen as a limitation of its applicability. Think for example to the common situation in which one knows that the true data distribution is given by an involved function that does not admit an easy analytical tractability (e.g. it is impossible to derive an ML estimator). In these cases, one usually decides to assume a simpler model, e.g. a Gaussian distribution, introducing a mismatch but, on the other hand, gaining the possibility to derive a simple (mismatched) estimator. Then, the MCRB can be evaluated since the true model is known and it can be used to evaluate the potential performance loss due to the mismatch between the assumed and the true model. Another useful application of the MCRB is the prediction of possible weaknesses (i.e. breakdown of the estimation performance) of the system under peculiar conditions. In particular, given an assumed model for the data, one can be interested in evaluating the performance loss in the presence of a certain number of "true" possible data distributions that the system can undergo. We note, in passing, that in all the situations in which the true p.d.f. is known but is not possible to evaluate in closed form the expectation operator involved in the definition of the matrices $\mathbf{A}_{\boldsymbol{\theta}_0}$ in eq. (1) and $\mathbf{B}_{\boldsymbol{\theta}_0}$ in eq. (7), the MCRB can be approximated by means of Monte Carlo simulations.

Before passing to introduce the MML estimator, in the following we provide a brief discussion about the difference between the results obtained in [2] and the general derivation of the MCRB provided in [1]. Even if the final mismatched covariance inequality assumes exactly the same expression, the proof in [1] is actually more general, while the one provided in [2] relies on the first order Taylor expansion of the estimation error term (see eq. 41 in [2]). This derivation leads to define a restricted class of estimators for which the MCRB of (5) applies. In particular, this class of estimators can be defined by the following two properties:

1. The expected value w.r.t. the true distribution is the same for all the estimators in the class and is equal to $E_p\left\{\hat{\boldsymbol{\theta}}(\mathbf{x})\right\} = \boldsymbol{\mu}$,
2. The correlation matrix $\boldsymbol{\Xi}_{\boldsymbol{\theta}}$ between the estimation error and the score function $\boldsymbol{\eta}_{\boldsymbol{\theta}}(\mathbf{x})$, i.e.

$$\boldsymbol{\Xi}_{\boldsymbol{\theta}} = E_p\left\{\left(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\mu}_p\right)\boldsymbol{\eta}_{\boldsymbol{\theta}}(\mathbf{x})^T\right\} \tag{9}$$

must be equal to some matrix function $\mathbf{M}(\boldsymbol{\theta})$, such that $\boldsymbol{\Xi}_{\boldsymbol{\theta}} = \pm\mathbf{M}(\boldsymbol{\theta})$ for all the estimators in the class. Since, in order to define explicitly the MCRB, the score function used in [2] is $\boldsymbol{\eta}_{\boldsymbol{\theta}}(\mathbf{x}) = \nabla_{\boldsymbol{\theta}} \ln f_X\left(\mathbf{x};\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}} D\left(p \| f_{\boldsymbol{\theta}}\right)$ it turns out that the correlation matrix $\boldsymbol{\Xi}_{\boldsymbol{\theta}}$ must be equal to $\pm\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}$.

As shown in [2], there is at least an estimator that *asymptotically* satisfies constrains 1) and 2). This estimator is exactly the MML estimator that we introduce in the next session. However, in general, it would be very difficult to characterize explicitly a class of estimators that satisfy these two constraints. The advantage on the proof in [1] w.r.t. the one in [2] is the fact that it shows that the inequality in (5) holds for all the MS-unbiased estimators and not only for those that satisfy 1) and 2).

II.C The Mismatched Maximum Likelihood Estimator

In this section, we consider the MML estimator derived in [3] and [4] as:

$$\hat{\boldsymbol{\theta}}_{MML}\left(\mathbf{x}\right) = \underset{\boldsymbol{\theta}\in\Theta}{\arg\max} \ln f_X\left(\mathbf{x};\boldsymbol{\theta}\right) = \underset{\boldsymbol{\theta}\in\Theta}{\arg\max} \sum_{m=1}^{M} \ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right), \tag{10}$$

where $\mathbf{x}_m \sim p_X(\mathbf{x}_m)$. It can be shown (see [3] and [4]) that the MML estimator converges *almost surely* (*a.s.*) to the $\boldsymbol{\theta}_0$ introduced in eq. (2), i.e. the vector that minimizes the Kullback-Leibler (KL) divergence between $p_X(\mathbf{x}_m)$ and $f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right)$:

$$\hat{\boldsymbol{\theta}}_{MML}\left(\mathbf{x}\right) \underset{M\to\infty}{\overset{a.s.}{\to}} \boldsymbol{\theta}_0, \tag{11}$$

Under similar regularity conditions to the ones given in Section II.A, in [3] and [4], the asymptotic normality of the MML estimator is proved. This result can be summarized in the following Theorem (see [3] and [4] for the proof):

Theorem 2 ([3], [4]): *Under suitable regularity conditions, it can be proved that*

$$\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{MML}\left(\mathbf{x}\right) - \boldsymbol{\theta}_0\right) \underset{M\to\infty}{\overset{d.}{\sim}} \mathcal{N}\left(\mathbf{0}, \mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\right), \tag{12}$$

where $\underset{M\to\infty}{\overset{d.}{\sim}}$ *indicates the convergence in distribution and the matrices* $\mathbf{A}_{\boldsymbol{\theta}_0}$ *and* $\mathbf{B}_{\boldsymbol{\theta}_0}$ *have been defined in eqs. (1) and (7)* *respectively . The asymptotic covariance matrix* $\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}$ *is generally called Huber's "sandwich" covariance.*

Theorem 1 and Theorem 2 highlight an interesting fact: the MML estimator is asymptotically MS-unbiased and its error covariance matrix asymptotically achieves the MCRB. The similarity with the classical (matched) estimation framework is now clear: the MML estimator is the counterpart of the ML estimator in the presence of misspecified models, as the MCRB is the counterpart of the classical (matched) CRB. In must be noted however that, while in the classical matched case, the convergence and the unbiasedness of the ML estimator is defined w.r.t. the true parameter vector $\bar{\boldsymbol{\theta}}$, in the mismatched case the convergence and the MS-unbiasedness of the MML is always defined w.r.t. the pseudo-true parameter vector $\boldsymbol{\theta}_0$ in eq. (2). The next section will provide some insights about this important point.

II.D A particular case: the MCRB as a bound on the Mean Square Error (MSE)

In this section, we focus on a particular mismatched case: the unknown parameter space T of the true model is the same of the parameter space Θ of the assumed model, i.e. $\mathrm{T} \equiv \Theta \subset \mathbb{R}^p$. More formally, suppose that the true parametric p.d.f. of the observations $p_X(\mathbf{x}) \equiv p_X(\mathbf{x};\bar{\boldsymbol{\theta}})$ and the assumed p.d.f. $f_X(\mathbf{x};\boldsymbol{\theta})$ belong to two (generally different) families of p.d.f.'s, $\mathcal{P}$ and $\mathcal{F}$, that can be parameterized by using the same parameter space Θ:

$$\mathcal{P} = \left\{ p_{\boldsymbol{\theta}} \,\middle|\, p_X(\mathbf{x};\boldsymbol{\theta}) \text{ is a pdf } \forall \boldsymbol{\theta} \in \Theta \right\}, \quad \mathcal{F} = \left\{ f_{\boldsymbol{\theta}} \,\middle|\, f_X(\mathbf{x};\boldsymbol{\theta}) \text{ is a pdf } \forall \boldsymbol{\theta} \in \Theta \right\}. \tag{11}$$

Even if this is only a particular case of the theory developed in the previous sections, this type of mismatch allows us to more deeply understand the nature of the MCRB and of the MML estimator. In particular, if condition (11) is satisfied, we can directly compare the MCRB and the MML estimator with their classical (matched) counterparts, i.e. the CRB and the ML estimator. This can be done since the pseudo-true parameter vector $\boldsymbol{\theta}_0$ belongs to the same parameter space of the true model $\bar{\boldsymbol{\theta}}$, then the difference vector $\mathbf{r} \triangleq \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ is well-defined. It is good to underline that, even in this particular case, $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ may be different, i.e. the vector **r** is in general different from a zero-vector. To best of our knowledge, general conditions under which $\bar{\boldsymbol{\theta}} \equiv \boldsymbol{\theta}_0$ are not available in the literature. The vector **r** indicates the distance between the convergence point $\bar{\boldsymbol{\theta}}$ of the classical ML estimator if the true p.d.f. $p_X(\mathbf{x};\bar{\boldsymbol{\theta}})$ were perfectly known and the convergence point $\boldsymbol{\theta}_0$ of the MML estimator when the mismatched p.d.f. $f_X(\mathbf{x};\boldsymbol{\theta})$ satisfies the condition in (11) . Moreover, using **r**, a bound on the *Mean Square Error* (MSE) on the estimation of $\bar{\boldsymbol{\theta}}$ in the presence of mismatched models can be established. To do this, for any MS-unbiased mismatched estimator, the Error Covariance Matrix in eq. (6) can be rewritten as:

$$\begin{aligned}\mathrm{MSE}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\bar{\boldsymbol{\theta}}\right) &\triangleq E_p\left\{\left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0+\boldsymbol{\theta}_0-\bar{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0+\boldsymbol{\theta}_0-\bar{\boldsymbol{\theta}}\right)^T\right\} \\ &= \mathbf{C}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\boldsymbol{\theta}_0\right) - 2E_p\left\{\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0\right\}\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)^T + \left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}}-\right. \\ &= \mathbf{C}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\boldsymbol{\theta}_0\right) + \mathbf{r}\mathbf{r}^T,\end{aligned} \tag{12}$$

where $E_p\left\{\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0\right\}=0$ from the MS-unbiasedness assumption. A similar expansion of the MSE can be found in [2] (see eq. 70). Finally, by substituting the covariance inequality in (5) in eq. (12), we can obtain a misspecified bound on the MSE of $\bar{\boldsymbol{\theta}}$ as:

$$\mathrm{MSE}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\bar{\boldsymbol{\theta}}\right) \geq \frac{1}{M}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1} + \mathbf{r}\mathbf{r}^T. \tag{13}$$

Moreover, if the condition in (11) is satisfied, the concept of consistency can be extended also to MS-unbiased mismatched estimators. In particular, we define as *consistent* an MS-unbiased mismatched estimator if, as the number of data vectors M goes to infinity, it converges to the *true* parameter vector $\bar{\boldsymbol{\theta}}$, i.e., $\hat{\boldsymbol{\theta}}(\mathbf{x}) \underset{M\to\infty}{\overset{a.s.}{\to}} \boldsymbol{\theta}_0 = \bar{\boldsymbol{\theta}}$.

The mismatched MSE inequality in (13) and the concept of consistency for MS-unbiased mismatched estimators can be very useful to compare in a very intuitive and self-explicative manner the nature of the MCRB and of the MML estimator as we will see in the next two sections.

III. A simple example: estimation of variance under misspecification of the mean

In order to clarify the use of the MCRB and the MML estimator, a simple example is described in the following (see also [17] for other examples). The problem is to estimate the variance of Gaussian data in the presence of misspecified mean value (e.g. we erroneously assume that the data are zero mean). Let us assume to have a set of M i.i.d. scalar observations $\mathbf{x} = \{x_m\}_{m=1}^M$, distributed according to a Gaussian p.d.f. with mean value μ_X and variance σ_X^2, i.e. $p_X(x_m) \triangleq p_X(x_m;\bar{\theta}) \equiv \mathcal{N}(\mu_X,\sigma_X^2)$. It is well-known that, given the observation vector **x**, the ML estimator of the variance is given by $\hat{\theta}_{ML}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^M (x_m-\mu_X)^2$, where $x_m \sim \mathcal{N}(\mu_X,\sigma_X^2)$. Suppose now that the assumed Gaussian p.d.f. is $f_X(x_m;\theta) = \mathcal{N}(\mu,\theta)$, so we misspecify the mean value. It can be noted that in this simple example, the true unknown model $p_X(x_m)$ and the assumed model $f_X(x_m;\theta)$

admit the same parameterization, so this example falls in the particular case addressed in Section II.D. Following eq. (10), the MML estimator for the variance of the data is given by $\hat{\theta}_{MML}(\mathbf{x}) = \arg\max_{\theta\in\Theta} \sum_{m=1}^{M} \ln f_X(x_m;\theta)$, where:

$$\ln f_X(x_m;\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\theta - \frac{1}{2\theta}(x_m-\mu)^2 . \qquad (14)$$

It is immediate to show that the MML estimator is given by:

$$\hat{\theta}_{MML}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M}(x_m-\mu)^2 . \qquad (15)$$

In this case, the KL divergence between $p_X(x_m)$ and $f_X(x_m;\theta)$ can be expressed as [18]:

$$D(p\|f_{\theta}) = \frac{(\mu_X-\mu)^2}{2\theta} + \frac{1}{2}\left(\frac{\sigma_X^2}{\theta} - 1 - \ln\frac{\sigma_X^2}{\theta}\right). (16)$$

By taking the derivative with respect to θ and by setting equal to zero the resulting expression, we get:

$$\frac{\partial D(p\|f_\theta)}{\partial\theta}\Big|_{\theta=\theta_0} = \frac{-(\mu_X-\mu)^2-\sigma_X^2}{2\theta^2} + \frac{1}{2\theta}\Big|_{\theta=\theta_0} = 0 \quad (17)$$

Hence, we get $\theta_0 = \sigma_X^2 + (\mu_X-\mu)^2 \neq \bar{\theta}$. Eq. (17) shows that the MML does not converge to the true variance, unless $\mu = \mu_X$, i.e. when there is no model mismatch. This means that the MML estimator is not consistent. From the scalar version of eq. (4), the mean value of the MML estimator with respect to the true distribution $p_X(\mathbf{x};\bar{\theta})$ is:

$$\mu = E_p\left\{\hat{\theta}_{MML}(\mathbf{x})\right\} = \sigma_X^2 + (\mu_X-\mu)^2 = \theta_0 . \qquad (18)$$

Hence, the MML estimator is MS-unbiased and the MCRB can be evaluated as shown in (5). By evaluating the first and the second derivative of the $\ln f_X(x_m;\theta)$ and after some simple calculation, the matrices (that in this case are scalars) $\mathbf{A}_{\theta_0}$ of eq. (1) and $\mathbf{B}_{\theta_0}$ eq. (7) are obtained:

$$A_{\theta_0} = -\frac{1}{2\theta_0^2},$$
$$B_{\theta_0} = \frac{3\sigma_X^4 + 6\sigma_X^2(\mu_X-\mu)^2 + (\mu_X-\mu)^4 - \theta_0^2}{4\theta_0^4} . \qquad (19)$$

Finally, from (13), we have that:

$$\mathrm{MCRB}(\bar{\theta}) = \mathrm{MCRB}(\sigma_X^2) = \frac{2\sigma_X^4}{M} + \frac{4\sigma_X^2(\mu_X-\mu)^2}{M} + (\mu_X-\mu)^4 . \qquad (20)$$

It is well-known that the CRLB for this estimation problem is given by $\mathrm{CRB}(\sigma_X^2) = 2\sigma_X^4/M$. Hence, $\mathrm{MCRB}(\sigma_X^2) \geq \mathrm{CRB}(\sigma_X^2)$, i.e. the MCRB is always greater or equal than the CRLB in the present study case. When $\mu = \mu_X$, i.e. we correctly specify the mean value, then $\theta_0 = \bar{\theta} = \sigma_X^2$ and $\mathrm{MCRB}(\sigma_X^2) = \mathrm{CRB}(\sigma_X^2)$.

IV. MCRB for the estimation of the scatter matrix in the family of CES distributions

In this section, we use the MCRB to investigate the problem of estimating the $N \times N$ scatter matrix of Complex Elliptically Symmetric (CES) distributed data, given *M* i.i.d. realizations of the *N*-dimensional data vector **x**, in the presence of data mismodelling. CES distributions constitute a wide family of distributions such as the complex Gaussian, Cauchy, generalized Gaussian, and compound Gaussian, which in turn includes the Gaussian distribution, the *K*-distribution, and the complex *t*-distribution [22]. The CES distributions are widely applied in many areas, such as radar, sonar, and communications [19, 20, 21, 22].

A complex *N*-dimensional random vector $\mathbf{x}_m$ is CES distributed, in shorthand notation $\mathbf{x}_i \sim CE_N(\boldsymbol{\gamma},\boldsymbol{\Sigma},h)$, if its p.d.f. is of the form:

$$p_X(\mathbf{x}_m) = c_{N,h}|\boldsymbol{\Sigma}|^{-1} h\left((\mathbf{x}_m-\boldsymbol{\gamma})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x}_m-\boldsymbol{\gamma})\right), \qquad (21)$$

where h is the density generator, $c_{N,h}$ is a normalizing constant, $\boldsymbol{\gamma} \triangleq E_p\{\mathbf{x}_m\}$ and $\boldsymbol{\Sigma}$ is the normalized (or shape) covariance matrix, also called *scatter matrix*, such that $\mathrm{tr}(\boldsymbol{\Sigma}) = N$. In particular, if $\mathbf{M} \triangleq E_p\{(\mathbf{x}_m-\boldsymbol{\gamma})(\mathbf{x}_m-\boldsymbol{\gamma})^H\}$ is the covariance matrix of the vector $\mathbf{x}_m$, then $\boldsymbol{\Sigma} = N/\mathrm{tr}(\mathbf{M})\cdot\mathbf{M}$. It is important to observe that, for some CES distributions, the unnormalized covariance matrix **M** does not exist, but the scatter matrix $\boldsymbol{\Sigma}$ is still well defined. Based upon the Stochastic Representation Theorem [19] any $\mathbf{x}_m \sim CE_N(\boldsymbol{\gamma},\boldsymbol{\Sigma},h)$ with $rank(\boldsymbol{\Sigma}) = k \leq N$ admits the stochastic representation $\mathbf{x}_m =_d \boldsymbol{\gamma} + R\mathbf{P}\mathbf{u}$, where the non-negative random variable (r.v.) $R \triangleq \sqrt{Q}$, the so-called *modular variate*, is a real, non-negative random variable, **u** is a *k*-dimensional vector uniformly distributed on the unit hyper-sphere with *k*-1 topological dimensions such that $\mathbf{u}^H\mathbf{u} = 1$, R and $\mathbf{u}$ are independent and $\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}^H$ is a factorization of $\boldsymbol{\Sigma}$, where **P** is a *N*x*k* matrix and $rank(\mathbf{P}) = k$. In the following derivations, we assume that $\boldsymbol{\Sigma}$ is full-rank, then $rank(\mathbf{P}) = rank(\boldsymbol{\Sigma}) = N$, and that it is real. For the CES distributions, the term $\sigma_X^2 \triangleq E\{Q\}/N$ can be interpreted as the statistical power of the random vector $\mathbf{x}_m$, i.e. the covariance matrix **M** and the scatter matrix $\boldsymbol{\Sigma}$ are linked by $\mathbf{M} = \sigma_X^2\boldsymbol{\Sigma}$. In general, the density generator itself depends of some additional parameters. For example, the complex *t*-distribution is completely characterized if its mean vector $\boldsymbol{\gamma}$, its scatter matrix $\mathbf{\Sigma}$ and its shape and scale parameters, λ and η respectively, are perfectly known [19]. Suppose now to have the following type of misspecification: the true distribution is a CES distribution with mean value $\bar{\boldsymbol{\gamma}}$ and scatter matrix $\bar{\mathbf{\Sigma}}$ i.e. $p_X(\mathbf{x}) = CE_N(\bar{\boldsymbol{\gamma}}, \bar{\mathbf{\Sigma}}, h)$, while the assumed distribution is a complex *t*-distribution $f_X(\mathbf{x};\boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta} = \left[\boldsymbol{\gamma}^T \quad \mathrm{vecs}(\mathbf{\Sigma})^T \quad \lambda \quad \eta\right]^T$, where the vecs-operator is the "symmetric" counterpart of the standard vec-operator that maps a symmetric $N \times N$ matrix $\mathbf{\Sigma}$ in a $N(N+1)/2$-dimensional vector whose entries are the elements of the lower (or upper) triangular sub-matrix of $\mathbf{\Sigma}$. Then, we could apply the MML estimator in eq. (10) and the MCRB in (5), provided that the pseudo-true parameters $\boldsymbol{\theta}_0$ exists. Of course, if the assumed model were a CES distribution different from the complex *t*-distribution, the parameter vector $\boldsymbol{\theta}$ need to be recast in order to take into account the additional parameters that characterize the given CES distribution. We left this general problem to future work and focus on the particular mismatched case discussed in Section II.D. In particular, since in many scenarios (e.g. radar and sonar) the mean value of the data can be considered null, we assume $\bar{\boldsymbol{\gamma}} = \boldsymbol{\gamma} = \mathbf{0}$. Moreover, we assume that all the characteristic parameters of the assumed CES distribution are known, except the elements of the scatter matrix $\mathbf{\Sigma}$. Then, the parameter vector that parameterizes an assumed zero-mean CES distribution can be defined as $\boldsymbol{\theta} = \mathrm{vecs}(\mathbf{\Sigma})$.

In the following, we assume that both the true distribution $p_X(\mathbf{x}_m)$ and the assumed distribution $f_X(\mathbf{x}_m;\boldsymbol{\theta})$ belong to the zero-mean CES distribution class:

$$p_X(\mathbf{x}_m) \triangleq p_X(\mathbf{x}_m;\bar{\mathbf{\Sigma}}) = c_{N,h}\left|\bar{\mathbf{\Sigma}}\right|^{-1} h\left(\mathbf{x}_m^H \bar{\mathbf{\Sigma}}^{-1}\mathbf{x}_m\right), \qquad (22)$$

$$f_X(\mathbf{x}_m;\boldsymbol{\theta}) \triangleq f_X(\mathbf{x}_m;\mathbf{\Sigma}) = c_{N,g}\left|\mathbf{\Sigma}\right|^{-1} g\left(\mathbf{x}_m^H \mathbf{\Sigma}^{-1}\mathbf{x}_m\right), \qquad (23)$$

where $\bar{\boldsymbol{\theta}} = \mathrm{vecs}(\bar{\mathbf{\Sigma}})$, $\boldsymbol{\theta} = \mathrm{vecs}(\mathbf{\Sigma})$, h is the density generator of the true p.d.f., and g is the density generator of the assumed p.d.f.. In the following, we analyze two different scenarios.

Case Study 1 Assumed p.d.f.: complex Normal; true p.d.f.: *t*-student.

We assume a complex Gaussian model for the data, i.e. we assume that each i.i.d. complex vector of the available dataset $\mathbf{x} = \{\mathbf{x}_m\}_{m=1}^M$ is distributed according to a complex Normal multivariate p.d.f., which also belongs to the CES family:

$$f_X(\mathbf{x}_m;\boldsymbol{\theta}) \triangleq f_X(\mathbf{x}_m;\mathbf{\Sigma}) = \frac{1}{\left(\pi\sigma^2\right)^N \left|\mathbf{\Sigma}\right|}\exp\left(-\frac{\mathbf{x}_m^H \mathbf{\Sigma}^{-1}\mathbf{x}_m}{\sigma^2}\right). \qquad (24)$$

The covariance matrix $\mathbf{M} = E\left\{\mathbf{x}_m\mathbf{x}_m^H\right\} = \sigma^2\mathbf{\Sigma}$ in this case exists. However, the true data are distributed according to another CES distribution, the complex *t*-distribution:

$$p_X(\mathbf{x}_m;\bar{\boldsymbol{\theta}}) \triangleq p_X(\mathbf{x}_m;\bar{\mathbf{\Sigma}}) = \frac{1}{\pi^N\left|\bar{\mathbf{\Sigma}}\right|}\frac{\Gamma(N+\lambda)}{\Gamma(\lambda)}\left(\frac{\lambda}{\eta}\right)^{\lambda}\left(\frac{\lambda}{\eta} + \mathbf{x}_m^H\bar{\mathbf{\Sigma}}^{-1}\mathbf{x}_m\right)^{-(N+\lambda)},$$ (25)

where λ is the shape parameter and η is the scale parameter characterizing the model [19], [22].

The assumption of a complex Normal model is motivated by the fact that the MML estimator of the scatter matrix can be easily derived to be the well-known Sample Covariance Matrix (SCM), $\hat{\mathbf{M}}_{MML} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_m\mathbf{x}_m^H$, so we get:

$$\hat{\mathbf{\Sigma}}_{MML} = \frac{\hat{\mathbf{M}}_{MML}}{\sigma^2} = \frac{1}{M\sigma^2}\sum_{m=1}^{M}\mathbf{x}_m\mathbf{x}_m^H, \qquad (26)$$

where the power σ^2 is assumed to be a priori known. As first step, we evaluate the matrix that minimizes the KL divergence between $p_X(\mathbf{x}_m;\bar{\mathbf{\Sigma}})$, considered here as a generic element of the CES family, and $f_X(\mathbf{x}_m;\mathbf{\Sigma})$ (the complex Normal p.d.f.). This matrix is the convergence point of the MML estimator in eq. (26). The differential of the KL divergence with respect to $\mathbf{\Sigma}$ is given by [23]:

$$\begin{aligned}\partial D\left(p \| f_{\mathbf{\Sigma}}\right) &= -E_p\left\{\partial \ln f_X(\mathbf{x}_m;\mathbf{\Sigma})\right\} = -E_p\left\{\partial \ln\left|\mathbf{\Sigma}\right|^{-1} + \partial \ln g\left(\mathbf{x}_m^H\mathbf{\Sigma}^{-1}\mathbf{x}_m\right)\right\} \\ &= \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\partial\mathbf{\Sigma}\right) + \mathrm{tr}\left(E_p\left\{\frac{d\ln g(Q_{\mathbf{\Sigma}})}{dQ_{\mathbf{\Sigma}}}\mathbf{\Sigma}^{-1}\mathbf{x}_m\mathbf{x}_m^H\mathbf{\Sigma}^{-1}\partial\mathbf{\Sigma}\right\}\right),\end{aligned} \qquad (27)$$

where:

$$Q_{\mathbf{\Sigma}} \triangleq \mathbf{x}_m^H\mathbf{\Sigma}^{-1}\mathbf{x}_m \qquad (28)$$

The last equality in eq. (27) follows directly from the same calculus given in [21] and [24]. Since the assumed distribution $f_X(\mathbf{x}_m;\mathbf{\Sigma})$ is a complex Normal distribution, $g(Q_{\mathbf{\Sigma}}) = \exp(-Q_{\mathbf{\Sigma}}/\sigma^2)$ and $\frac{d\ln g(Q_{\mathbf{\Sigma}})}{dQ_{\mathbf{\Sigma}}} = -\frac{1}{\sigma^2}$. By substituting this result in eq. (27), we get:

$$\partial D\left(p \| f_{\mathbf{\Sigma}}\right) = \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\partial\mathbf{\Sigma}\right) - \frac{1}{\sigma^2}\mathrm{tr}\left(E_p\left\{\mathbf{\Sigma}^{-1}\mathbf{x}_m\mathbf{x}_m^H\mathbf{\Sigma}^{-1}\partial\mathbf{\Sigma}\right\}\right) = \mathrm{tr}\left(\left[\mathbf{\Sigma}^{-1} - \frac{\sigma_X^2}{\sigma^2}\mathbf{\Sigma}^{-1}\right.\right. \qquad (29)$$

where we used the property $E_p\left\{\mathbf{x}_m\mathbf{x}_m^H\right\} = \sigma_X^2\bar{\mathbf{\Sigma}}$. Then, following the standard rules of matrix calculus [23], the derivative of the KL divergence w.r.t. $\mathbf{\Sigma}$ is:

$$\frac{\partial}{\partial \mathbf{\Sigma}} D\left(p \| f_{\Sigma}\right) = \mathbf{\Sigma}^{-1} - \frac{\sigma_X^2}{\sigma^2} \mathbf{\Sigma}^{-1} \bar{\mathbf{\Sigma}} \mathbf{\Sigma}^{-1} . \quad (30)$$

Finally, by posing the derivative in eq. (30) equal to zero, we obtain that the matrix $\mathbf{\Sigma}_0$ that minimizes the KL divergence is:

$$\mathbf{\Sigma}_0 = \frac{\sigma_X^2}{\sigma^2} \bar{\mathbf{\Sigma}} . \quad (31)$$

Eq. (31) shows that the MML estimator converges *a.s.* to a scaled version of the true scatter matrix, $\hat{\mathbf{\Sigma}}_{MML}(\mathbf{x}) \underset{M\to\infty}{\overset{a.s.}{\to}} \mathbf{\Sigma}_0 = \left(\sigma_X^2/\sigma^2\right)\bar{\mathbf{\Sigma}}$, so it is not consistent. It is consistent only when the two powers of the assumed and true p.d.f.'s are equal. The mean value of the MML estimator with respect to the true distribution is:

$$\boldsymbol{\mu} = E_p\left\{\hat{\mathbf{\Sigma}}_{MML}(\mathbf{x})\right\} = \frac{\sigma_X^2}{\sigma^2} \bar{\mathbf{\Sigma}} = \mathbf{\Sigma}_0 . \quad (32)$$

Hence, the MML estimator is MS-unbiased. Given the MS-unbiasedness of the proposed MML estimator, we can now evaluate the MCRB. In [24] the MCRB on the estimation of the scatter matrix was evaluated for two CES distributions, the complex-*t* and the Generalized Gaussian, when the assumed misspecified distribution is a complex Normal p.d.f.. Here, we assume that the true distribution is a complex-*t* distribution with p.d.f. given in eq. (25).

Before providing the expression of the MCRB, some considerations on a reasonable choice of the true distribution parameters, λ and η, have to be made. In fact, the power $\sigma_X^2 \triangleq E_p\left\{Q_{\bar{\Sigma}}\right\}/N$ is function of these two parameters. In fact, by applying the Stochastic Representation Theorem, we have that $Q_{\bar{\Sigma}}$ has an *F*-distribution [19] such that

$$p_{Q_{\bar{\Sigma}}}(q) = \frac{1}{B(N,\lambda)} q^{N-1} \left(\frac{\lambda}{\eta}\right)^{\lambda} \left(\frac{\lambda}{\eta} + q\right)^{-(N+\lambda)}, \quad (33)$$

where $B(N,\lambda) = \frac{\Gamma(N)\Gamma(\lambda)}{\Gamma(N+\lambda)} = \frac{(N-1)!\Gamma(\lambda)}{\Gamma(N+\lambda)}$. In this case, we have:

$$\sigma_X^2 = \frac{E_p\left\{Q_{\bar{\Sigma}}\right\}}{N} = \frac{\lambda}{\eta(\lambda-1)}, \quad (34)$$

$$E_p\left\{Q_{\bar{\Sigma}}^2\right\} = \sigma_X^4 \frac{N(N+1)(\lambda-1)}{(\lambda-2)}, \quad \lambda > 2 . \quad (35)$$

In order to focus on the impact of the mismatch due to the differences between the density generator, we assume that $\sigma_X^2 = \sigma^2$, so that $\mathbf{\Sigma}_0 = \bar{\mathbf{\Sigma}}$, this guarantees that the MML estimator is consistent and we choose λ and η accordingly.

A compact expression for the MCRB for two distributions in the CES family is given in Appendix A. Then, following the results in [24] and by applying eq. (A.10), the MCRB can be expressed as:

$$\mathrm{MCRB}\left(\bar{\boldsymbol{\theta}}\right) = \frac{1}{M} \mathbf{D}_N^{\dagger} \left[\frac{1}{(\lambda-2)} \mathrm{vec}\left(\bar{\mathbf{\Sigma}}\right) \mathrm{vec}\left(\bar{\mathbf{\Sigma}}\right)^T + \frac{(\lambda-1)}{(\lambda-2)} \bar{\mathbf{\Sigma}} \otimes \bar{\mathbf{\Sigma}}\right] \left(\mathbf{D}_N^{\dagger}\right)^T . \quad (36)$$

where $\mathbf{D}_N$ is the so-called Duplication matrix of order N ([25], [26], [27]). The duplication matrix is implicitly defined as the unique $N^2 \times N(N+1)/2$ matrix that satisfies the following equality: $\mathbf{D}_N \mathrm{vecs}(\mathbf{A}) = \mathrm{vec}(\mathbf{A})$ for any symmetric matrix $\mathbf{A}$. The symbol † denotes the Moore-Penrose pseudo-inverse. Moreover, using the expression of the Fisher Information Matrix (FIM) for *t*-distributed data evaluated in [21] and the properties of the vec and vecs operators, the duplication matrix $\mathbf{D}_N$ and of the Kronecker product $\otimes$ ([25], [26], [27],[28]), the CRLB can be expressed as:

$$\mathrm{CRLB}\left(\bar{\boldsymbol{\theta}}\right) = \frac{1}{M} \mathbf{D}_N^{\dagger} \left[\frac{N+\lambda+1}{\lambda(N+\lambda)} \mathrm{vec}\left(\bar{\mathbf{\Sigma}}\right) \mathrm{vec}\left(\bar{\mathbf{\Sigma}}\right)^T + \frac{N+\lambda+1}{N+\lambda} \bar{\mathbf{\Sigma}} \otimes \bar{\mathbf{\Sigma}}\right] \left(\mathbf{D}_N^{\dagger}\right)^T , \quad (37)$$

It can be proved that the matrix inequality $\mathrm{MCRB}\left(\bar{\boldsymbol{\theta}}\right) \geq \mathrm{CRLB}\left(\bar{\boldsymbol{\theta}}\right)$ holds true for any value of the parameters in the present study case (we do not report here the details for lack of space). In the following, we describe some simulation results that confirm this finding.

For the sake of comparison, in the following figures we report, along with the MSE of the MML, the MCRB and the CRLB, also the MSE of the robust (unconstrained) Tyler's estimator ([29], [30], [31], [32]). Tyler's estimator has been derived in the context of the CES distribution as the most robust estimator in min-max sense [32]. In particular, Tyler's estimator can be obtained as the recursive solution of the following (unconstrained) fixed-point (FP) matrix equation:

$$\mathbf{\Sigma} = \frac{N}{M} \sum_{m=1}^{M} \frac{\mathbf{x}_m \mathbf{x}_m^H}{\mathbf{x}_m^H \mathbf{\Sigma}^{-1} \mathbf{x}_m} . \quad (38)$$

To solve eq. (38), we use the following iterative approach:

$$\begin{cases} \hat{\mathbf{\Sigma}}_T^{(0)} = \hat{\mathbf{\Sigma}}_{MML} \\ \hat{\mathbf{\Sigma}}_T^{(k+1)} = \dfrac{N}{M} \displaystyle\sum_{m=1}^{M} \frac{\mathbf{x}_m \mathbf{x}_m^H}{\mathbf{x}_m^H \left(\hat{\mathbf{\Sigma}}_T^{(k)}\right)^{-1} \mathbf{x}_m}, \quad k = 0, \ldots, K \end{cases} \quad (39)$$

It can be noted that, unlike the recursive procedure proposed in [29], in (39) there is not a normalization constraint on the trace of $\hat{\mathbf{\Sigma}}_T^{(k)}$. It must be noted that the MCRB in (5) does not apply to the Tyler's estimator since $\hat{\mathbf{\Sigma}}_T^{(k)}$ has not been derived under any assumed CES distribution.

In order to have a global performance index (i.e. an index that is able to take into account the errors made in the estimation of all the covariance entries) we define ε as:

$$\varepsilon \triangleq \frac{\left\| E_p\left\{ \left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\bar{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\bar{\boldsymbol{\theta}}\right)^T \right\} \right\|_F}{\left\|\bar{\mathbf{\Sigma}}\right\|_F}, \quad (40)$$

where $\hat{\boldsymbol{\theta}} = \text{vecs}(\hat{\mathbf{\Sigma}})$, $\hat{\mathbf{\Sigma}}$ is an estimate of the true covariance matrix $\bar{\mathbf{\Sigma}}$, $\bar{\boldsymbol{\theta}} = \text{vecs}(\bar{\mathbf{\Sigma}})$ and $\|\mathbf{A}\|_F = \sqrt{\text{tr}\left(\mathbf{A}^T\mathbf{A}\right)}$ is the Frobenius norm. Fig. 1 shows the behavior of this global performance index for the MML and Tyler's estimators as a function of the shape parameter λ. As performance bounds, the following quantities are plotted:

$$\varepsilon_{\text{HB}} \triangleq \frac{\left\|\text{MCRB}\left(\bar{\boldsymbol{\theta}}\right)\right\|_F}{\left\|\bar{\mathbf{\Sigma}}\right\|_F}, \ \varepsilon_{\text{CRLB}} \triangleq \frac{\left\|\text{CRLB}\left(\bar{\boldsymbol{\theta}}\right)\right\|_F}{\left\|\bar{\mathbf{\Sigma}}\right\|_F} \quad (41)$$

The true covariance matrix is assumed to be $[\mathbf{\Sigma}]_{i,j} = \rho^{|i-j|}$. The value of the one-lag coefficient is $\rho = 0.9$, the number of observations vectors is M=3N. To calculate the global performance indices ε_{MML} and ε_{Tyler} of the estimators, we run 10^5 Monte Carlo trials. As expected, for high values of λ the MCRB and the CRLB tend to be equal, since for $\lambda \to \infty$ the *t*-student p.d.f. tends to a complex Gaussian p.d.f., and the matched and mismatched models coincide. Moreover, as $\lambda \to \infty$, the MML estimator converges to the ML estimator, and then it attains the CRLB. This is not the case for Tyler's estimator that suffers from "robustness losses", i.e. it is robust but not optimal when the data tends to be Gaussian distributed ($\lambda \to \infty$). In Fig. 2, ε_{MCRB}, ε_{CLRB}, ε_{MML} and ε_{Tyler} are compared as a function of the number of available data M, for $\lambda = 3$. In this case, Tyler's estimator has better estimation performance than the MML estimator, thanks to its robustness [19]. For completeness, in Fig. 3 we investigate the performance of the MML and of Tyler's estimator as function of the one-lag coefficient ρ for $\lambda = 3$, N=8 and M=3N. As expected, Tyler's estimator achieves better estimation performance for all the values of ρ. Finally, it can be noted that the MCRB is not applicable to Tyler's estimator since it is not based on any misspecified data distribution, therefore its RMSE sometimes falls below the MCRB. On the other hand, since Tyler's estimator is an unbiased estimator of $\bar{\mathbf{\Sigma}}$ (at least in its unconstrained version) the CRB applies.

Case Study 2 Assumed p.d.f.: Generalized Gaussian; true p.d.f.: *t*-student.

As in the previous example, we assume that the true distribution is a complex-*t* distribution but unlike the previous case, we assume a complex Generalized Gaussian (GG) distribution for the data. The MML estimator, then, is the ML estimator for the GG data. Unlike the SCM (i.e. the ML estimator for Gaussian data), the ML estimator for GG data cannot be expressed with an explicit equation but has to be defined through a fixed-point equation. In this section, we first discuss some properties of the MML estimator (in particular, bias and consistency in the mismatched sense), and then we evaluate the relevant MCRB. In this case study, the true distribution has the same p.d.f. given in eq. (25), while the assumed p.d.f. is the Generalized Gaussian distribution:

$$f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right) \triangleq f_X\left(\mathbf{x}_m;\mathbf{\Sigma}\right) = \frac{\beta\Gamma\left(N\right)b^{-N/\beta}}{\pi^N\Gamma\left(N/\beta\right)}\frac{1}{|\mathbf{\Sigma}|}\exp\left(-\frac{\left(\mathbf{x}_m^H\mathbf{\Sigma}^{-1}\mathbf{x}_m\right)^\beta}{b}\right), \quad (42)$$

where β is the shape parameter and b is the scale parameter [19] that are assumed to be known. In this case, the MML estimator (i.e. the ML estimator for GG data) is the solution of the following fixed-point matrix equation [19], [21], [33]:

$$\hat{\mathbf{\Sigma}}_{MML} = \frac{1}{M}\sum_{m=1}^{M}\varphi\left(\mathbf{x}_m^H\hat{\mathbf{\Sigma}}_{MML}^{-1}\mathbf{x}_m\right)\mathbf{x}_m\mathbf{x}_m^H = H_M\left(\hat{\mathbf{\Sigma}}_{MML}\right) \quad (43)$$

where the function φ is given by $\varphi(t) = \frac{\beta}{b}t^{\beta-1}$. Following Theorem 6 in [19], it can be shown that, for every (symmetric and positive-definite) starting matrix $\mathbf{\Sigma}^{(0)}$, the recursive version of eq. (43) converges to $\hat{\mathbf{\Sigma}}_{MML}$, i.e. $\hat{\mathbf{\Sigma}}^{k+1} = H_M\left(\hat{\mathbf{\Sigma}}^k\right) \underset{k\to\infty}{\to} \hat{\mathbf{\Sigma}}_{MML}$ if and only if $\beta \in (0,1)$. For $\beta > 1$, i.e. when the tails of the GG distribution are lighter than the one of the Normal distribution, the recursive estimator of the scatter matrix is no longer reliable. In fact, for $\beta > 1$, the conditions on $\varphi(t)$ that guarantee the existence and the uniqueness of the estimator are not satisfied. Theorem 5 in [30] can be used to prove that, for $\beta \in (0,1)$ we have $\hat{\mathbf{\Sigma}}_{MML} \underset{M\to\infty}{\overset{a.s.}{\to}} \mathbf{\Sigma}_0$, i.e. the MML estimator $\hat{\mathbf{\Sigma}}_{MML}$ converges with probability 1 to $\mathbf{\Sigma}_0$. From the theory previously discussed, the limiting value $\mathbf{\Sigma}_0$ must be the matrix that minimizes the KL divergence between $p_X\left(\mathbf{x}_m\right)$ and $f\left(\mathbf{x}_m;\mathbf{\Sigma}\right)$. In order to evaluate $\mathbf{\Sigma}_0$, we can apply eq. (27), where, in this case, the density generator is the one of the GG distribution, i.e. $l(t) = \exp\left(-t^\beta/b\right)$. After some calculation, we get:

$$\partial D\left(p \| f_{\Sigma}\right)=\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma}\right)-\frac{\beta}{b} \operatorname{tr}\left(E_{p}\left\{\left(\mathbf{x}_{m}^{H} \boldsymbol{\Sigma}^{-1} \mathbf{x}_{m}\right)^{\beta-1} \boldsymbol{\Sigma}^{-1} \mathbf{x}_{m} \mathbf{x}_{m}^{H} \boldsymbol{\Sigma}^{-1}\right\} \partial \boldsymbol{\Sigma}\right). \quad (44)$$

By applying the Stochastic Representation Theorem, we have that $\mathbf{x}_i =_d \sqrt{Q_{\bar{\Sigma}}}\mathbf{T}\mathbf{u}$, where $Q_{\bar{\Sigma}} \triangleq \mathbf{z}^H \bar{\boldsymbol{\Sigma}}^{-1}\mathbf{z}$, $\bar{\boldsymbol{\Sigma}} = \mathbf{T}\mathbf{T}^H$ is a factorization of the shape $\bar{\boldsymbol{\Sigma}}$, $\mathbf{u}$ is a N-dimensional vector uniformly distributed on the unit hyper-sphere with N-1 topological dimensions such that $\mathbf{u}^H\mathbf{u}=1$ and $E\left\{\mathbf{u}\mathbf{u}^H\right\} = N^{-1}\mathbf{I}$. Then, eq. (44) can be rewritten as:

$$\partial D\left(p \| f_{\Sigma}\right)=\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma}\right)-\frac{\beta E\left\{Q_{\bar{\Sigma}}^{\beta}\right\}}{b} \operatorname{tr}\left(E_{p}\left\{\left(\mathbf{u}^{H} \mathbf{T}^{H} \boldsymbol{\Sigma}^{-1} \mathbf{T} \mathbf{u}\right)^{\beta-1} \boldsymbol{\Sigma}^{-1} \mathbf{T} \mathbf{u} \mathbf{u}^{H} \mathbf{T}^{H} \boldsymbol{\Sigma}^{-1}\right\} \partial \boldsymbol{\Sigma}\right) \quad (45)$$

where $E\left\{Q_{\bar{\Sigma}}^{\beta}\right\}$ can be evaluated explicitly by using the integral in ([34, p. 315, n. 194.3]) as:

$$E\left\{Q_{\bar{\Sigma}}^{\beta}\right\}=\left(\frac{\lambda}{\eta}\right)^{\beta} \frac{\Gamma(\beta+N) \Gamma(\lambda-\beta)}{\Gamma(N) \Gamma(\lambda)}. \quad (46)$$

From eq. (45), setting to zero the derivative of the KL divergence w.r.t. $\boldsymbol{\Sigma}$ leads to:

$$\boldsymbol{\Sigma}^{-1}-\beta b^{-1} E\left\{Q_{\bar{\Sigma}}^{\beta}\right\} E\left\{\left(\mathbf{u}^{H} \mathbf{T}^{H} \boldsymbol{\Sigma}^{-1} \mathbf{T} \mathbf{u}\right)^{\beta-1} \boldsymbol{\Sigma}^{-1} \mathbf{T} \mathbf{u} \mathbf{u}^{H} \mathbf{T}^{H} \boldsymbol{\Sigma}^{-1}\right\}\Big|_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_{0}}=\mathbf{0}. \quad (47)$$

Through some standard matrix manipulation, we get:

$$\mathbf{T}^{-1} \boldsymbol{\Sigma}_{0} \mathbf{T}^{-H}=\beta b^{-1} E\left\{Q_{\bar{\Sigma}}^{\beta}\right\} E\left\{\left(\mathbf{u}^{H} \mathbf{T}^{H} \boldsymbol{\Sigma}_{0}^{-1} \mathbf{T} \mathbf{u}\right)^{\beta-1} \mathbf{u} \mathbf{u}^{H}\right\}. \quad (48)$$

Now, we assume that the solution of eq. (48) is a scaled version of the true shape matrix, i.e. $\boldsymbol{\Sigma}_0 = \delta\bar{\boldsymbol{\Sigma}}$. Then, we have $\delta\mathbf{I} = \frac{\beta}{bN}E\left\{Q_{\bar{\Sigma}}^{\beta}\right\}\delta^{-\beta+1}\mathbf{I}$, so that

$$\delta=\frac{\lambda}{\eta}\left(\frac{\beta}{b N} \frac{\Gamma(\beta+N) \Gamma(\lambda-\beta)}{\Gamma(N) \Gamma(\lambda)}\right)^{1 / \beta}. \quad (49)$$

Then, the matrix that minimizes the KL divergence is given by:

$$\boldsymbol{\Sigma}_{0}=\frac{\lambda}{\eta}\left(\frac{\beta}{b N} \frac{\Gamma(\beta+N) \Gamma(\lambda-\beta)}{\Gamma(N) \Gamma(\lambda)}\right)^{1 / \beta} \bar{\boldsymbol{\Sigma}} \triangleq \delta \bar{\boldsymbol{\Sigma}}. \quad (50)$$

Since $\boldsymbol{\Sigma}_0$ is a scaled version of the true scatter matrix, the MML estimator is not consistent in general. As shown in [19] and [30], for the estimator in eq. (43), the following asymptotic relation holds:

$$E_{p}\left\{\varphi\left(\mathbf{x}_{m}^{H} \boldsymbol{\Sigma}_{f, MML}^{-1} \mathbf{x}_{m}\right) \mathbf{x}_{m} \mathbf{x}_{m}^{H}\right\}=\gamma \bar{\boldsymbol{\Sigma}}. \quad (51)$$

Eq. (51) can be used to evaluate the bias of the MML estimator in the mismatched sense. The mean value of the MML estimator with respect to the true distribution g is:

$$\boldsymbol{\mu}=E_{p}\left\{\hat{\boldsymbol{\Sigma}}_{MML}(\mathbf{x})\right\} \overset{M \rightarrow \infty}{=} \gamma \bar{\boldsymbol{\Sigma}}, \quad (52)$$

where the scalar term γ can be evaluated by solving the following integral equation [19]:

$$E\left\{\varphi\left(\frac{Q_{\Sigma}}{\gamma}\right) \frac{Q_{\Sigma}}{\gamma}\right\}=N. \quad (53)$$

Given $\varphi(t) = \beta t^{\beta-1}/b$ and by using the integral in eq. (46), γ can be evaluated as:

$$\gamma=\frac{\lambda}{\eta}\left(\frac{\beta}{b N} \cdot \frac{\Gamma(\beta+N) \Gamma(\lambda-\beta)}{\Gamma(N) \Gamma(\lambda)}\right)^{1 / \beta}=\delta. \quad (54)$$

Hence, the MML estimator is (asymptotically) MS-unbiased, i.e. the mean value $\boldsymbol{\mu}$ tends to the matrix that minimizes the KL divergence $\boldsymbol{\mu} = \delta\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_0$. However, the MML is not consistent since it converges to a scaled version of the true scatter matrix. As before, we select the parameter values in such a way that the estimator is consistent. In other words, we choose a set of shape and scale parameters of the assumed and the true distributions such that $\delta = 1$, and then $\boldsymbol{\mu} = \boldsymbol{\Sigma}_0 = \bar{\boldsymbol{\Sigma}}$. To have $\delta = 1$, a possible choice of the scale parameter η of the t-distribution and the scale parameter b of the GG distribution is:

$$\eta=\frac{\lambda}{\lambda-1} \quad \text{and} \quad b=\frac{\Gamma(\beta+N) \Gamma(\lambda-\beta)}{\Gamma(N) \Gamma(\lambda)}\left(\frac{\lambda}{\eta}\right)^{\beta} \frac{\beta}{N}.$$

With this choice of the parameters, we can compare the estimation performance of the MML estimator with the MCRB. As before, the MCRB can be evaluated using the compact expression for $\mathbf{A}_{\bar{\theta}}^{-1}\mathbf{B}_{\bar{\theta}}\mathbf{A}_{\bar{\theta}}^{-1}$ derived in Appendix A, eq. (A.10). The density generator for the GG distribution is $g(t) = \exp\left(-t^{\beta}/b\right)$, hence we have:

$$\frac{\partial \ln g\left(Q_{\Sigma}\right)}{\partial Q_{\Sigma}}=-\frac{\beta}{b} Q_{\Sigma}^{\beta-1} \quad \text{and} \quad \frac{\partial^{2} \ln g\left(Q_{\Sigma}\right)}{\partial Q_{\Sigma}^{2}}=-\frac{\beta(\beta-1)}{b} Q_{\Sigma}^{\beta-2}.$$

In order to evaluate the term B_1, B_2, A_1 and A_2 in eqs. (A.2), (A.3), (A.6) and (A.7) respectively, the integral in eq. (46) is needed. In particular, we have:

$$E\left\{Q \frac{\partial \ln g(Q)}{\partial Q}\right\}=-\frac{\beta}{b}\left(\frac{\lambda}{\eta}\right)^{\beta} \frac{\Gamma(\beta+N) \Gamma(\lambda-\beta)}{\Gamma(N) \Gamma(\lambda)}, \quad (55)$$

$$E\left\{Q^2\left(\frac{\partial \ln g(Q)}{\partial Q}\right)^2\right\}=\frac{\beta^2}{b^2}\left(\frac{\lambda}{\eta}\right)^{2\beta}\frac{\Gamma(2\beta+N)\Gamma(\lambda-2\beta)}{\Gamma(N)\Gamma(\lambda)}, \tag{56}$$

$$E\left\{Q^2\frac{\partial^2 \ln g(Q)}{\partial Q^2}\right\}=-\frac{\beta(\beta-1)}{b}\left(\frac{\lambda}{\eta}\right)^{\beta}\frac{\Gamma(\beta+N)\Gamma(\lambda-\beta)}{\Gamma(N)\Gamma(\lambda)}, \tag{57}$$

with $\beta<\lambda/2$. Finally, the MCRB is evaluated using eq. (A.10), while the CRLB is given in eq. (37). In the following, we compare the RMSE of the MML estimator and of Tyler's estimator given in eq. (38) with the square root of the MCRB and with the square root of the CRLB. Both the iterations to derive the MML and Tyler's estimators are initialized using the SCM estimate. As before, the value of the one-lag coefficient is $\rho=0.9$, the number of secondary vectors is M=3N. To calculate the MSE of the estimators, we run 10^5 Monte Carlo trials. MSE, MCRB and CRLB are compared in terms of the global performance indices: ε_{MML}, ε_{Tyler}, ε_{MCRB}, and ε_{CRLB} previously defined. The simulation results concern two different scenarios: the quasi-Gaussian scenario, where λ=50 and the super-Gaussian scenario, where λ=3 (λ is the shape parameter of the *t*-distribution).

Super-Gaussian Case (λ=3). In the following, we describe the results for the super-Gaussian case, i.e. the true *t*-distribution has heavier tails than a Normal distribution. As shown in Fig. 4, unlike the quasi-Gaussian case, the MML estimator achieves better performance (here in terms of ε_{MML}) when β tends to 0, i.e. when the assumed GG distribution has heavier tails than a Normal distribution. Figs. 5 and 6, show the behavior of ε_{MML}, ε_{Tyler}, ε_{MCRB}, and ε_{CRLB} as function of the number M of available data vectors, for β=0.8 and β=0.1, respectively. As we can see, slightly better performance can be observed when the shape parameter of the assumed GG distribution is set to be equal to β=0.1. Similar considerations can be carried out by looking at the plots of ε_{MML}, ε_{Tyler}, ε_{MCRB} and ε_{CRLB} as a function of the one-lag correlation coefficient ρ (not reported here for lack of space). In all the analyzed scenarios, the MSE, the MCRB, and the CRB get worse when the clutter one-lag correlation coefficient increases.

Quasi-Gaussian scenario (λ=50). Compared to the super-Gaussian case, as expected the MML estimator and the MCRB have an opposite behavior with respect to the choice of the shape parameter of the assumed GG distribution β. In fact, with λ=50, the true *t*-distribution is very close to the Normal distribution, so the performance of the MML estimator increases as β tends to 1, i.e. the MML estimator tends to the SCM. We evaluated the ε_{MML}, ε_{FP}, ε_{MCRB}, and ε_{CRLB} as a function of the number M of available data vectors, for β=0.8 and β=0.1 (the plots not reported here for lack of space). The progress of the global indices are very similar, but a slight increase of the performance can be observed when β=0.8.

VI. Conclusions

In practical applications, a certain amount of mismatch between the true and the assumed data distribution is often inevitable. The behavior of the ML estimator under data mismodeling, i.e. the MML estimator, has been deeply investigated in the statistical literature, but little attention has been devoted to the relevant performance bound. In this paper, we present a review of the results obtained in [1] and [2] on a general covariance inequality for any MS-unbiased estimator of a deterministic parameter vector under data mismodeling and we show how to apply these results to the problem of estimating the scatter matrix of Complex Elliptically Symmetric (CES) distributed random vectors under data mismodeling. Two relevant case studies are discussed. In the first one, the true distribution is a complex-*t* distribution while the assumed distribution has complex Normal p.d.f.. In the second one, the true distribution is still a complex-*t* but the assumed distribution is a Generalized Gaussian. These two numerical case studies allowed us to quantify the "mismatch losses" in the estimation of the scatter matrix. Future work will explore a possible generalization of the MCRB to the class of Bayesian Bounds for random parameter estimation.

Appendix A: Compact expression for the MCRB in the CES family

In this Appendix, we derive a compact expression useful to evaluate the Huber Bound for the scatter matrix estimation in the family of CES distribution. This expression follows directly from the results obtained in [24]. We assume that both the true distribution $p_X(\mathbf{x})$ (that implicitly depends on the true scatter matrix $\bar{\boldsymbol{\Sigma}}$, then according to the notation used before, $\bar{\boldsymbol{\theta}}=\mathrm{vecs}(\bar{\boldsymbol{\Sigma}})$) and the assumed distribution $f_{\boldsymbol{\Sigma}}$ belong to the zero-mean CES distribution class, as shown in eqs. (22) and (23). Moreover, we define $Q \triangleq \mathbf{x}^H\boldsymbol{\Sigma}^{-1}\mathbf{x}$ as in eq. (28).

Compact expression for the matrix $\mathbf{B_0}$

In [24] the matrix $\mathbf{B_0}$ has been obtained element-by-element as:

$$\begin{aligned}
\left[\mathbf{B_0}\right]_{ij} &= E\left\{\frac{\partial \ln f_X(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \ln f_X(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_j}\right\} \\
&= \left(1+\frac{2}{N}E\left\{Q\frac{\partial \ln g(Q)}{\partial Q}\right\}+\frac{1}{N(N+1)}E\left\{Q^2\left(\frac{\partial \ln g(Q)}{\partial Q}\right)^2\right\}\right)\mathrm{tr}\big(\boldsymbol{\Sigma} \\
&\quad +\frac{1}{N(N+1)}E\left\{Q^2\left(\frac{\partial \ln g(Q)}{\partial Q}\right)^2\right\}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}_i\boldsymbol{\Sigma}^{-1}\mathbf{A}_j\right),
\end{aligned} \tag{A.1}$$

where $\mathbf{A}_i=\partial\boldsymbol{\Sigma}/\partial\theta_i$ is a symmetric 0-1 matrix.

For notation simplicity, we define:

$$B_1 = 1 + \frac{2}{N} E\left\{ Q \frac{\partial \ln g(Q)}{\partial Q} \right\} + \frac{1}{N(N+1)} E\left\{ Q^2 \left(\frac{\partial \ln g(Q)}{\partial Q} \right)^2 \right\}$$

, (A.2)

$$B_2 = \frac{1}{N(N+1)} E\left\{ Q^2 \left(\frac{\partial \ln g(Q)}{\partial Q} \right)^2 \right\}. \quad \text{(A.3)}$$

By using the properties of the vec operator, of the Duplication matrix $\mathbf{D}_N$ and of the Kronecker product [25] [26], we have:

$$\mathbf{B}_{\boldsymbol{\theta}} = \mathbf{D}_N^T \left[B_1 \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^T + B_2 \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right] \mathbf{D}_N \quad \text{(A.4)}$$

Compact expression for the matrix $\mathbf{A}_{\boldsymbol{\theta}}$

In [24] the matrix $\mathbf{A}_{\boldsymbol{\theta}}$ has been obtained element-by-element as:

$$\left[\mathbf{A}_{\boldsymbol{\theta}}\right]_{ij} = E\left\{ \frac{\partial^2 \ln f_X(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\} =$$

$$= \left(1 + \frac{2}{N} E\left\{ Q \frac{\partial \ln g(Q)}{\partial Q} \right\} + \frac{1}{N(N+1)} E\left\{ Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2} \right\} \right) \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}_i \boldsymbol{\Sigma}^{-1}\mathbf{A}_j\right) +$$

$$+ \frac{1}{N(N+1)} E\left\{ Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2} \right\} \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}_i\right) \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{A}_j\right)$$

(A.5)

For notation simplicity, we define:

$$A_2 = 1 + \frac{2}{N} E\left\{ Q \frac{\partial \ln g(Q)}{\partial Q} \right\} + \frac{1}{N(N+1)} E\left\{ Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2} \right\},$$

(A.6)

$$A_1 = \frac{1}{N(N+1)} E\left\{ Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2} \right\}. \quad \text{(A.7)}$$

Finally, as for the matrix $\mathbf{B}$, we have:

$$\mathbf{A}_{\boldsymbol{\theta}} = \mathbf{D}_N^T \left[A_1 \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^T + A_2 \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \right] \mathbf{D}_N \,. \text{(A.8)}$$

By using the Sherman-Morrison formula, we can express the inverse of the matrix $\mathbf{A}$ as follows:

$$\mathbf{A}_{\boldsymbol{\theta}}^{-1} = \mathbf{D}_N^{\dagger} \left[\frac{1}{A_2} \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} - \frac{A_1}{A_2 (A_2 + N A_1)} \mathrm{vec}(\boldsymbol{\Sigma}) \mathrm{vec}(\boldsymbol{\Sigma})^T + \right] \left(\mathbf{D}_N^{\dagger}\right)^T$$

. (A.9)

Compact expression for the MCRB, $\mathrm{MCRB}(\boldsymbol{\theta}) = \mathbf{A}_{\boldsymbol{\theta}}^{-1} \mathbf{B}_{\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}}^{-1}$ (with $\mathbf{r}=\mathbf{0}$)

$$\mathrm{MCRB}(\boldsymbol{\theta}) = \frac{1}{M} \mathbf{A}_{\boldsymbol{\theta}}^{-1} \mathbf{B}_{\boldsymbol{\theta}} \mathbf{A}_{\boldsymbol{\theta}}^{-1}$$

$$= \frac{1}{M} \mathbf{D}_N^{\dagger} \left[\frac{B_2}{A_2^2} \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + \left(\frac{B_1}{A_2^2} - \frac{2A_1 (B_2 + N B_1)}{A_2 (A_2 + N A_1)} + \frac{2 N A_1^2 (B_2 + N B_1)}{A_2^2 (A_2 + N A_1)^2} \right) \mathrm{vec}(\boldsymbol{\Sigma}) \mathrm{v}$$

(A.10)

References

[1] Q. H. Vuong, “Cramér-Rao bounds for misspecified models,” *Working paper 652, Division of the Humanities and Social Sciences, Caltech*, October 1986. Available at: https://www.hss.caltech.edu/content/cramer-rao-bounds-misspecified-models.

[2] C. D. Richmond, L. L. Horowitz, "Parameter Bounds on Estimation Accuracy Under Model Misspecification," *IEEE Trans. on Signal Processing*, vol. 63, no. 9, pp. 2263-2278, May 1, 2015.

[3] P. J. Huber, “The behavior of Maximum Likelihood Estimates under Nonstandard Conditions,” *Proc. of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*. Berkley: University of California Press, 1967.

[4] H. White, “Maximum likelihood estimation of misspecified models”, *Econometrica* vol. 50, pp. 1-25, January 1982.

[5] H. White, “Consequences and detection of misspecified nonlinear regression models,” *Journal of the American Statistical Association*, vol. 76, pp. 419-433, 1981.

[6] H. White, *Estimation, Inference and Specification Analysis*, Econometric Society Monographs, Cambridge University Press, August 1996.

[7] T. B. Fomby, R. C. Hill, *Maximum-Likelihood Estimation of Misspecified Models: Twenty Years Later*, Kidington, Oxford, UK, Elsevier Ltd, 2003.

[8] Y. Noam, J. Tabrikian, "Marginal Likelihood for Estimation and Detection Theory," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 3963-3974, Aug. 2007.

[9] W. Xu, A.B. Baggeroer, K.L. Bell, "A bound on mean-square estimation error with background parameter mismatch," *IEEE Transactions on Information Theory*, vol.50, no.4, pp.621,632, April 2004

[10] C. Fritsche, U. Orguner, E. Ozkan, F. Gustafsson, "On the Cramér-Rao lower bound under model mismatch," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015*, pp. 3986-3990, 19-24 April 2015.

[11] T. J. Rothenberg, “Identification in parametric models,” *Econometrica*, vol. 39, no. 3, pp. 577-591, May 1971.

[12] R. Bowden, “The theory of parametric identification,” *Econometrica*, vol. 41, no. 6, pp. 1069-1074, November 1973.

[13] S. Fortunati, F. Gini, M. S. Greco, A. Farina, A. Graziano, S. Giompapa, “On the Identifiability Problem in the presence of Random Nuisance Parameters”, *Signal Processing* vol. 92, pp. 2545-2551, 2012.

[14] P.J. Schreier, L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge University Press 2010.
[15] A. K. Jagannatham, B. D. Rao, "Cramér-Rao lower bound for constrained complex parameters," *IEEE Signal Process. Lett.*, vol. 11, no. 11, Nov. 2004.
[16] T. Menni, E. Chaumette, P. Larzabal, J.P. Barbot, "New Results on Deterministic Cramér–Rao Bounds for Real and Complex Parameters", *IEEE Transactions on Signal Processing*, vol 60, no. 3, pp. 1032-1049, 2012.
[17] S. Fortunati, M. S. Greco, F. Gini, "A lower bound for the Mismatched Maximum Likelihood Estimator," *IEEE Radar Conf. 2015*, Arlington, USA, May 11-15, 2015.
[18] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, NY, 1991.
[19] E. Ollila, D.E. Tyler, V. Koivunen, V.H. Poor, "Complex Elliptically Symmetric Distributions: Survey, New Results and Applications," *IEEE Trans. on Signal Processing*, Vol. 60, No. 11, pp.5597-5625, 2012.
[20] C. D. Richmond, "PDF's, confidence regions, and relevant statistics for a class of sample covariance-based array processors," *IEEE Trans. on Signal Processing*, Vol. 44 , No. 7 , pp. 1779-1793, 1996.
[21] M. S. Greco, F. Gini, "Cramér-Rao Lower Bounds on Covariance Matrix Estimation for Complex Elliptically Symmetric Distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6401-6409, December 2013.
[22] K. J. Sangston, F. Gini, M. Greco, "Coherent radar detection in heavy-tailed compound-Gaussian clutter", *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 42, No.1, pp. 64-77, 2012.
[23] J. R. Magnus, H. Neudecker, "Matrix Differential Calculus with Applications to Simple, Hadamard, and Kronecker Products," *Journal of Mathematical Psychology*, vol. 29, pp. 414-492, 1985.
[24] M. S. Greco, S. Fortunati, F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions," *Signal Processing*, vol. 104, pp. 381-386, November 2014.
[25] J. R. Magnus, H. Neudecker, "The commutation matrix: some properties and applications," *The Annals of Statistics*, vol. 7, pp. 381-394, 1979.
[26] J. R. Magnus, H. Neudecker, "The elimination matrix: some lemmas and applications," *SIAM Journal on Algebraic and Discrete Methods*, vol. 1, pp. 422-499, 1980.
[27] J. R. Magnus, H. Neudecker, "Symmetry, 0-1 matrices and Jacobians: a review," *Econometric Theory*, vol. 2, pp. 157-190, 1986.
[28] K. B. Petersen, M. S. Pedersen, *The Matrix Cookbook*, 2012 [http://matrixcookbook.com].
[29] F. Gini, M. S. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847-1859, December 2002.
[30] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The Annals of Statistics*, vol. 4, no. 1, pp. 51-67, January 1976.
[31] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, P. Larzabal, "Covariance Structure Maximum-Likelihood Estimates in Compound Gaussian Noise: Existence and Algorithm Analysis," *IEEE Transactions on Signal Processing*, vol.56, no.1, pp.34,48, Jan. 2008.
[32] D. Tyler, "A distribution-free *M*-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234-251, January 1987.
[33] F. Pascal, L. Bombrun, J.-Y. Tourneret, Y. Berthoumieu, "Parameter Estimation For Multivariate Generalized Gaussian Distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5960-5971, Dec. 2013.
[34] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, Seventh Edition, 2007.

Figures

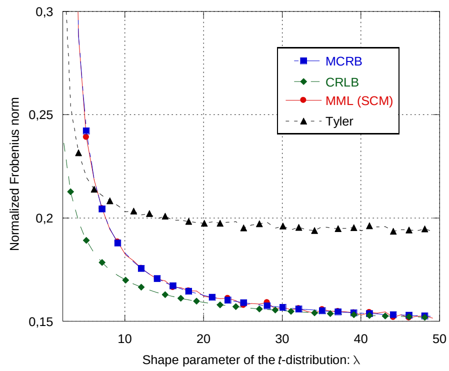

Figure 1 – Comparison among the normalized Frobenius norm of the MSE of the MML and of the Tyler's estimator, the MCRB and the CRLB as function of the shape parameter of the *t*-distribution (ρ=0.9, N=8, M=3N).

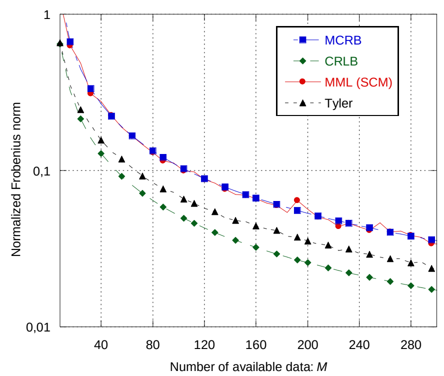

Figure 2 – Comparison among the normalized Frobenius norm of the MSE of the MML and of the Tyler's estimator, the MCRB and the CRLB as function of the available data (ρ=0.9, N=8, λ=3).

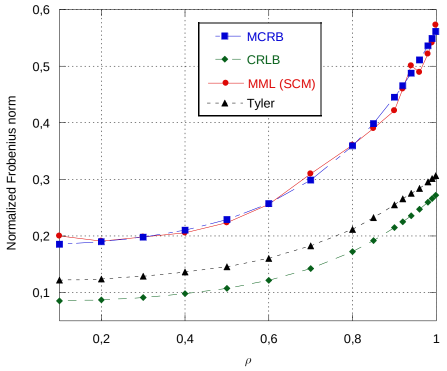

Figure 3 – Comparison among the normalized Frobenius norm of the MSE of the MML and of the Tyler's estimator, the MCRB and the CRLB as function of the correlation parameter ρ (λ=3, N=8, M=3N).

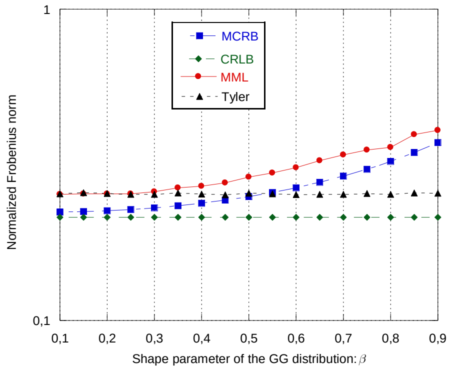

Figure 4 – Comparison among the normalized Frobenius norm of the MSE of the MML and the Tyler's estimator, the MCRB and the CRLB as a function of the shape parameter of the *GG* distribution (ρ=0.9, N=8, M=3N, λ=3).

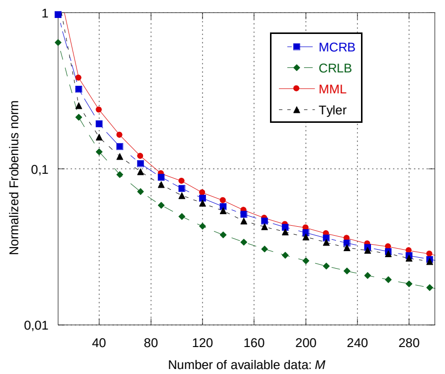

Figure 5 – Comparison among the normalized Frobenius norm of the MSE of the MML and of the Tyler's estimator, the MCRB and the CRLB as a function of the available data (ρ=0.9, N=8, β=0.8, λ=3).

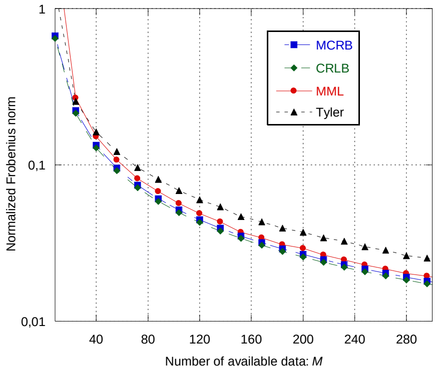

Figure 6 – Comparison among the normalized Frobenius norm of the MSE of the MML and of the Tyler's estimator, the MCRB and the CRLB as function of the available data (ρ=0.9, N=8, β=0.1, λ=3).

Authors' bios

Stefano FORTUNATI graduated in telecommunication engineering and received the PhD at the University of Pisa, Italy, in 2008 and 2012 respectively. In 2012, he joined the Department of Ingegneria dell'Informazione of the University of Pisa, where he is currently working as PostDoc researcher. From Sept. 2012 to Nov. 2012 and from Sept. 2013 to Nov. 2013, he was a visiting researcher at the CMRE NATO Research Center in La Spezia, Italy. His research interests encompasses different areas of the statistical signal processing: estimation and detection theory, target detection in non-Gaussian noise, robust detection and estimation theory, performance bounds, Compressed Sensing theory with applications to radar and sonar systems.

Fulvio GINI (Fellow IEEE) received the Doctor Engineer (cum laude) and the Research Doctor degrees in electronic engineering from the University of Pisa, Italy, in 1990 and 1995 respectively. In 1993 he joined the Department of Ingegneria dell'Informazione of the University of Pisa, where he become Associate Professor in 2000 and he is Full Professor since 2006. From July 1996 through January 1997, he was a visiting researcher at the Department of Electrical Engineering, University of Virginia, Charlottesville. He is an Associate Editor for the IEEE Transactions on Aerospace and Electronic Systems and for the Elsevier Signal Processing journal. He has been AE for the Transactions on Signal Processing (2000–06) and a Member of the EURASIP JASP Editorial Board. He has been the Editor-in-Chief of the Hindawi International Journal on Navigation and Observation (IJNO). He is the Area Editor for the Special issues of the IEEE Signal Processing Magazine. He was co-recipient of the 2001 IEEE AES Society's Barry Carlton Award for Best Paper. He was recipient of the 2003 IEE Achievement Award for outstanding contribution in signal processing and of the 2003 IEEE AES Society Nathanson Award to the Young Engineer of the Year. He has been a Member of the Signal Processing Theory and Methods (SPTM) Technical Committee (TC) of the IEEE Signal Processing Society and of the Sensor Array and Multichannel (SAM) TC for many years. He is a Member of the Board of Directors (BoD) of the EURASIP Society, the Award Chair (2006-2012) and the EURASIP President for the years 2013-2016. He was the Technical co-Chair of the 2006 EURASIP Signal and Image Processing Conference (EUSIPCO), Florence, Italy, September 2006, of the 2008 Radar Conference, Rome, Italy, May 2008, and of the IEEE CAMSAP 2015 workshop, to be held in Cancun, Mexico in December 2015. He was the General co-Chair of the 2nd Workshop on Cognitive Information Processing (CIP2010), of the IEEE ICASSP 2014, held in Florence in May 2014, and of the CoSeRa 2015 workshop on compressive sensing in radar, held in Pisa in June 2015. He was the guest co-editor of the special section of the Journal of the IEEE SP Society on Special Topics in Signal Processing on "Adaptive Waveform Design for Agile Sensing and Communication" (2007), guest editor of the special section of the IEEE Signal Processing Magazine on "Knowledge Based Systems for Adaptive Radar Detection, Tracking and Classification" (2006), guest co-editor of the two special issues of the EURASIP Signal Processing journal on "New trends and findings in antenna array processing for radar" (2004) and on "Advances in Sensor Array Processing (in memory of Alex Gershman)" (2013). He is co-editor and author of the book "Knowledge Based Radar Detection, Tracking and Classification" (2008) and of the book "Waveform Diversity and Design" (2012). His research interests include modeling and statistical analysis of radar clutter data, non-Gaussian signal detection and estimation, parameter estimation and data extraction from multichannel interferometric SAR data. He authored or co-authored 8 book chapters, about 120 journal papers and more than 150 conference papers.

Maria S. GRECO graduated in Electronic Engineering in 1993 and received the Ph.D. degree in Telecommunication Engineering in 1998, from University of Pisa, Italy. From December 1997 to May 1998 she joined the Georgia Tech Research Institute, Atlanta, USA as a visiting research scholar where she carried on research activity in the field of radar detection in non-Gaussian background. In 1993 she joined the Dept. of Information Engineering of the University of Pisa, where she is Associate Professor since Dec. 2011. She's IEEE fellow since Jan. 2011 and she was co-recipient of the 2001 IEEE Aerospace and Electronic Systems Society's Barry Carlton Award for Best Paper and recipient of the 2008 Fred Nathanson Young Engineer of the Year award for contributions to signal processing, estimation, and detection theory. In May and June 2015 she visited as invited Professor the Université Paris-Sud, CentraleSupélec, Paris, France. She has been general-chair, technical program chair and organizing committee member of many international conferences over the last 10 years. She is lead guest editor of the special issue on "Advanced Signal Processing for Radar Applications" of the IEEE Journal on Special Topics of Signal Processing, December 2015, she was guest co-editor of the special issue of the Journal of the IEEE Signal Processing Society on Special Topics in Signal Processing on "Adaptive Waveform Design for Agile Sensing and Communication," published in June 2007 and lead guest editor of the special issue of International Journal of Navigation and Observation on" Modelling and Processing of Radar Signals for Earth Observation published in August 2008. She's Associate Editor of IET Proceedings – Sonar, Radar and Navigation, Editor-in-Chief of the IEEE Aerospace and Electronic Systems Magazine, member of the Editorial Board of the Springer Journal of Advances in Signal Processing (JASP), Senior Editorial board member of IEEE Journal on Selected Topics of Signal Processing (J-STSP), member of the IEEE Signal Array Processing (SAM) Technical Committees. She's also member of the IEEE AES and IEEE SP Board of Governors and Chair of the IEEE AESS Radar Panel. She's as well SP Distinguished Lecturer for the years 2014-2015, AESS Distinguished Lecturer for the years 2015-2016 and member of the IEEE Fellow Committee. Maria is co-author of many tutorials on "Radar Clutter Modeling" and "Advanced Radar Detection", presented at the 2005 IEEE International Radar Conference (May 2005, Arlington, USA), 2008 IEEE Radar Conference (May 2008, Rome, Italy) , 2012 IEEE Radar Conference (May 2012, Atlanta, USA), 2014 International Radar

Conference (October 2014, Lille, France) , 2015 IET Radar Conference (October 2015, Hangzhou, China) and at the 2015 IEEE International Radar Conference (May 2015, Washington DC, USA). Her general interests are in the areas of statistical signal processing, estimation and detection theory. In particular, her research interests include clutter models, coherent and incoherent detection in non-Gaussian clutter, CFAR techniques, radar waveform diversity and bistatic/mustistatic active and passive radars. She co-authored many book chapters and more than 160 journal and conference papers.