Parameter Bounds Under Misspecified Models for Adaptive Radar Detection

Stefano Fortunati, Fulvio Gini, and Maria S. Greco

Dipartimento di Ingegneria dell'Informazione, University of Pisa

via G. Caruso 16, 56122 Pisa - Italy

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

Table of contents

1 List of symbols and functions

- $\mathbf{x} = \left\{\mathbf{x}_m\right\}_{m=1}^{M}$: set of the available observation vectors assumed to be independent and identically distributed (iid).

- $p_X(\mathbf{x}) \square\, p_X(\mathbf{x}; \boldsymbol{\tau})$: true joint probability density function (pdf) possibly parameterized by a real (deterministic and unknown) *true* vector $\boldsymbol{\tau} \in \mathrm{T} \subset \square^{\,p}$.

- $f_X(\mathbf{x}) \square\, f_\theta(\mathbf{x}) \square\, f_X(\mathbf{x}; \boldsymbol{\theta})$: Assumed joint pdf parameterized by a real (deterministic and unknown) vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \square^{\,d}$.

- $\hat{\boldsymbol{\theta}}(\mathbf{x})$: estimator of the deterministic parameter vector $\boldsymbol{\theta}$ based on the data $\mathbf{x}$.

- $\nabla_\theta u(\boldsymbol{\theta}) \square\, \left(\partial u/\partial \theta_1 \quad \cdots \quad \partial u/\partial \theta_d\right)^T$: gradient (column) vector of the scalar function u. With the notation $\nabla_{\theta_0} u(\boldsymbol{\theta}_0)$, we define the gradient of the function u evaluated at $\boldsymbol{\theta}_0$.

- $\mathbf{U}_\theta = \nabla_\theta^T \mathbf{u}(\boldsymbol{\theta})$: Jacobian matrix of the vector function $\mathbf{u}$. With the notation $\mathbf{U}_{\theta_0}$, we define the Jacobian matrix of the function $\mathbf{u}$ evaluated at $\boldsymbol{\theta}_0$.

- $E_p\{\mathbf{u}\} = \int \mathbf{u}(\mathbf{x}) p_X(\mathbf{x}) d\mathbf{x}$: Expectation operator of the (scalar or vector) function $\mathbf{u}$ with respect to a pdf $p_X(\mathbf{x})$.

- $D\left(p_X \| f_X\right) = \int p_X(\mathbf{x}) \ln\left(\dfrac{p_X(\mathbf{x})}{f_X(\mathbf{x})}\right) d\mathbf{x}$: Kullback-Leibler divergence between $p_X(\cdot)$ and $f_X(\cdot)$.

- vec($\mathbf{A}$): The vec-operator transforms a $N \times N$ matrix $\mathbf{A}$ into a vector by stacking the columns of the matrix one underneath the other.

- vecs(**A**): The vecs-operator denotes the $N(N+1)/2 \times 1$ vector that is obtained from vec(**A**) by eliminating all supradiagonal elements of **A**. For notation simplicity, $N(N+1)/2 \coloneqq l$.

- **D**$_N$: Duplication matrix of order N. The duplication matrix is implicitly defined as the unique $N^2 \times l$ matrix that satisfies the following equality $\mathbf{D}_N \text{vecs}(\mathbf{A}) = \text{vec}(\mathbf{A})$ for any symmetric matrix **A**.

- $\otimes$: Kronecker product.

- $\times$: Cartesian product.

- $\mathbf{A}^{\dagger} = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{A}^T$ is the Moore-Penrose pseudo-inverse of a matrix **A**.

2. 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 adopted 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).

The first fundamental result on the general theory of the estimation under misspecification was provided by Huber in his seminal paper [1] on the statistical analysis of the Maximum Likelihood (ML) estimator under mismatched condition. This work was further developed by White in [2] and [3], where the term "Quasi Maximum Likelihood" (QML) estimator was introduced. In particular, Huber and White have shown that the asymptotic distribution of the ML estimator under

misspecified models is a Gaussian distribution. Further, the mean value of the ML estimator 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, whereas the covariance matrix is given by the so-called Huber "sandwich" matrix. For the sake of clarity, in the rest of this chapter, we refer to the ML estimator under mismatched conditions as the Mismatched Maximum Likelihood (MML) estimator. A milestone on the theoretical misspecification analysis is the book in [4] that covers all developments in this field from both the statistical and econometrical points of view. This book provides an excellent and insightful discussion about statistical inference in the presence of distributional misspecification, with a focus on estimation and hypotheses testing problems.

In conjunction with the asymptotic analysis of the ML estimator, a question that naturally arises is whether it is possible to establish a lower bound on the error covariance matrix of a certain class of mismatched estimators. When the true parametric model is specified, few of such lower bounds exist; one of these is the well-known Cramér-Rao lower Bound (CRB). In a pioneering working paper [5], Q. H. Vuong firstly proposed a generalization of the CRB to the estimation problem under misspecified models.

Quite surprisingly and despite of the wide variety of potential applications, these fundamental results, disseminated in the statistical and econometrical literature, have remained largely unrecognized by the Signal Processing community for many years. Only recently, the findings about the estimation theory under misspecification have been rediscovered and applied to different signal processing problems. In a recent book [6], the main results on the ML estimator under misspecified models have been reported. A different mismodeling related to the dynamic of the acquired data has been investigated in [7]. In particular, the asymptotic performance of the ML estimator and of the generalized likelihood ratio test (GLRT) is derived under the assumption of

independent identically distribution (iid) samples, when these samples are correlated in the actual model. Other recent works attempt to generalize the Cramér-Rao inequality in the presence of model misspecification. In [8], a Bayesian bound of the Ziv-Zakai type has been derived under model mismatch conditions restricted to misparameterized zero mean complex Gaussian distributions. On the same line, in [70], [71] and [72], a Ziv-Zakai bound in the presence of model misspecification has been discussed and applied to the Time-of-Arrival (TOA) estimation problem. Recently, Richmond and Horowitz (first in [9] and then in [10]) derived a covariance inequality for deterministic complex parameter vector in the presence of model misspecification and introduced the term misspecified Cramér-Rao Bound (MCRB). Moreover, in [10] and in [74], a generalization to the mismatched case of the Slepian and Bangs formulae for the evaluation of the bound for multivariate complex Gaussian distributed observations is also derived (see Appendices A and B at the end of this chapter) and applied to the classical Direction Of Arrival (DOA) estimation problem. The application of the MCRB to the DOA estimation problem is also discussed in [11]. To the best of our knowledge, [10] represents the first attempt to introduce an organic framework for deriving a covariance inequality of the Cramér-Rao type in the presence of model mismatch to the Signal Processing Community. More recently, Richmond and Basu have extended the work in [10] to the Bayesian estimation framework ([12], [73]). Finally, the recent paper [13] is pertinent to misspecified bounds, where the authors adopted a different definition of unbiasedness and a different score function than in [5] and [10].

The aim of this chapter is twofold: in the first part, we provide a comprehensive review of the main findings about the MCRB and the MML estimator for deterministic parameter estimation. Two toy examples are also provided in order to clarify the main theoretical concepts. In the second part, we discuss the application of the MCRB and of the MML estimator to a practical radar signal processing problem: the estimation of the disturbance covariance (scatter) matrix for adaptive radar detection ([14], [15]). We recast this classical radar problem in the more general context of the

estimation of the scatter matrix in the Complex Elliptically Symmetric (CES) distribution family [16].

The rest of the chapter is organized as follows. Section 3 provides the formal description of the general deterministic estimation problem under model misspecification. In Section 4, the main theoretical results on the MCRB and the MML estimator are reviewed and discussed. In Section 5, two simple toy examples are described to better clarify the theoretical findings of Section 4 and how they should be applied. Section 6 focuses on the application of the MML estimator and of the MCRB to the estimation of the scatter matrix in the CES distribution family, while Section 7 discusses the application of the MML scatter matrix estimator in adaptive radar detection problems. Section 8 summarizes our conclusions.

3. Problem statement and motivations

In the following, a formal description of the estimation problem under mismatched conditions is provided. Let $\mathbf{x}_m \in \square^N$ be a N-dimensional random vector representing the outcome of a random experiment (i.e. the observation vector) with cumulative distribution function (cdf) $P_X(\mathbf{x}_m)$. In the remainder of the chapter, we assume that $P_X(\mathbf{x}_m)$ has a relevant probability density function (pdf) $p_X(\mathbf{x}_m)$, and we use, with a small abuse of definition, the term "distribution" always to indicate the relevant pdf. Assume that the true pdf of $\mathbf{x}_m$ is known to belong to a family P. A *structure T* is a set of hypotheses, which implies a unique pdf in P for $\mathbf{x}_m$. Such pdf is indicated with $p_X(\mathbf{x}_m;T)$ ([17], [18]). The set of all the a priori possible structures for p_X is called a *model*. We assume that the pdf 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 $\mathrm{T} \subset \square^p$.

The common assumption underlying any practical estimation problem is the perfect knowledge of the (joint) pdf $p_X(\mathbf{x};\boldsymbol{\tau})$ that characterizes the iid observations, $\mathbf{x} = \{\mathbf{x}_m\}_{m=1}^{M}$, except for the value of the parameter vector $\boldsymbol{\tau} \in \mathrm{T}$. However, a certain amount of mismatch between the true pdf of the observations and the pdf assumed to derive an estimator of the parameters of interest is always present. Specifically, suppose that the true parametric pdf of the observations, $p_X(\mathbf{x};\boldsymbol{\tau})$, and the assumed pdf, $f_X(\mathbf{x};\boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \Theta \subset \square^d$, belong to two (generally different) families of pdf's, P and F, that are parameterized by two possibly different parameters spaces T and Θ:

$$P = \{p_X \,|\, p_X(\mathbf{x};\boldsymbol{\tau}) \text{ is a pdf } \forall \boldsymbol{\tau} \in \mathrm{T}\}, \quad F = \{f_X \,|\, f_X(\mathbf{x};\boldsymbol{\theta}) \text{ is a pdf } \forall \boldsymbol{\theta} \in \Theta\}.$$

It is worth noting that the true parameter space T and the assumed parameter space Θ may be completely different and/or have a different dimensions.

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 $p_X(\mathbf{x};\boldsymbol{\tau})$ only as $p_X(\mathbf{x})$ in order to highlight the fact the neither the model, nor the true parameter vector $\boldsymbol{\tau}$ is accessible by a mismatched estimator ([10], [15]).

Suppose then that the *M* (possibly complex) iid measurement vectors are sampled from a particular pdf belonging to P, i.e. $\mathbf{x}_m \square\, p_X(\mathbf{x}_m)$, for *m*=1,2,…,*M*. Due to a lack of knowledge or to the need to fulfill some computational requirements, a parametric pdf $f_X(\mathbf{x};\boldsymbol{\theta})$, belonging to the family F, is assumed for the dataset $\mathbf{x}$. In this case a possible inference algorithm, e.g., an estimation algorithm, may be based on a *misspecified data model*, i.e., on the assumed pdf $f_X(\mathbf{x};\boldsymbol{\theta})$ and not on the true pdf $p_X(\mathbf{x})$. The question that arises is to how will the statistical properties of an estimator (e.g. convergence, consistency, efficiency) defined in the classical estimation framework change in this mismatched scenario? This is the main topic of the next section.

4 A generalization of the deterministic estimation theory under model misspecification

The aim of this section is to provide an organic view of the findings in [1], [2], [3], [5], [10] and [15]. Starting from [5], we first provide a list of regularity conditions that are not only a fundamental prerequisite for the derivation of the MCRB but also allow better understanding of the nature and the usefulness of this bound. Then, we provide the expression of the MCRB and the class of estimators to which it applies (Theorem 4.1 [5]). Finally, we conclude this section by introducing the MML estimator, its asymptotic properties and their link with the MCRB.

4.1 Regular models

As stated before, let $\mathbf{x} = \{\mathbf{x}_m\}_{m=1}^{M}$ be a set of iid N-dimensional random vectors and let $p_X(\mathbf{x})$ the true pdf of $\mathbf{x}$. Let $F = \{f_X \,|\, f_X(\mathbf{x}; \boldsymbol{\theta}) \text{ is a p.d.f. } \forall \boldsymbol{\theta} \in \Theta \subset \square^d\}$ be a family of parametric pdfs 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 $\zeta(\boldsymbol{\theta}) \square E_p\{\ln f_X(\mathbf{x}_m; \boldsymbol{\theta})\} = \int \ln f_X(\mathbf{x}_m; \boldsymbol{\theta}) p_X(\mathbf{x}_m) d\mathbf{x}_m$ has a unique maximum on Θ at an interior point $\boldsymbol{\theta}_0$. (b) The matrix $\mathbf{A}_{\boldsymbol{\theta}_0}$ whose entries are

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

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 [17], [18], [19], and [20]) 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]:

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

where

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

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. Assumption A2 ensures the existence and the uniqueness of the so-called *pseudo-true parameters vector* $\boldsymbol{\theta}_0$. As seen later in the chapter, $\boldsymbol{\theta}_0$ plays a key role both in the definition of the MCRB and of the MML.

Definition 1 *(Regular models)* [5]: A parametric model F is regular with respect to (w.r.t.) the pdf $p_X(\mathbf{x})$ if Assumptions A1-A3 hold. It is regular w.r.t. a family P if it is regular w.r.t. every pdf in P. It is referred as "regular" if the regularity is w.r.t. every pdf in F.

The following lemma summarizes some useful properties of parametric models that are regular w.r.t. the pdf $p_X(\mathbf{x})$. For any $p_X(\mathbf{x})$ in P, we define the matrix $\mathbf{B}_{\boldsymbol{\theta}}$ as:

$$\left[\mathbf{B}_{\boldsymbol{\theta}}\right]_{ij} \triangleq \left[E_p\left\{\nabla_{\boldsymbol{\theta}} \ln f_X(\mathbf{x}_m;\boldsymbol{\theta})\nabla_{\boldsymbol{\theta}}^T \ln f_X(\mathbf{x}_m;\boldsymbol{\theta})\right\}\right]_{ij} = E_p\left\{\frac{\partial \ln f_X(\mathbf{x}_m;\boldsymbol{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln f_X(\mathbf{x}_m;\boldsymbol{\theta})}{\partial \theta_j}\right\}. \quad (4)$$

Lemma 1: Let $F = \left\{f_X \,|\, f_X(\mathbf{x};\boldsymbol{\theta}) \text{ is a p.d.f. } \forall \boldsymbol{\theta} \in \Theta \subset \square^d\right\}$ be a family of parametric pdfs which is regular w.r.t. the pdf $p_X(\mathbf{x})$. Then:

i.   The function $\zeta(\boldsymbol{\theta}) = \int \ln f_X(\mathbf{x}_m;\boldsymbol{\theta}) p_X(\mathbf{x}_m)d\mathbf{x}_m$ is finite and twice continuously differentiable on Θ, and for every $\boldsymbol{\theta} \in \Theta$:

$$\frac{\partial \zeta(\boldsymbol{\theta})}{\partial \theta_i} = \int \frac{\partial \ln f_X(\mathbf{x}_m;\boldsymbol{\theta})}{\partial \theta_i} p_X(\mathbf{x}_m)d\mathbf{x}_m < \infty \quad i = 1,\dots,p, \quad (5)$$

$$\frac{\partial^2 \zeta(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \left[\mathbf{A}_{\boldsymbol{\theta}}\right]_{ij} < \infty \quad i,j = 1,\dots,p \quad (6)$$

Moreover, $\left[\mathbf{B}_{\boldsymbol{\theta}}\right]_{ij} < \infty$ for $i,j = 1,\dots,p$, and $\mathbf{A}_{\boldsymbol{\theta}_0}$ is negative definite, where $\boldsymbol{\theta}_0$ is the pseudo-true parameters vector defined in eq. (2).

ii. If $p_X(\mathbf{x}) = f_X(\mathbf{x};\overline{\boldsymbol{\theta}})$ for some $\overline{\boldsymbol{\theta}} \in \Theta$, then $\boldsymbol{\theta}_0 = \overline{\boldsymbol{\theta}}$ and

$$\mathbf{A}_{\boldsymbol{\theta}_0} + \mathbf{B}_{\boldsymbol{\theta}_0} = \mathbf{0}. \quad (7)$$

The proof of this lemma can be found in [5]. In particular, eq. (7) represents the classical equivalence between the two expressions of the Fisher Information Matrix (FIM) under correct model specification.

4.2 MS-unbiased estimators and the MCRB

Upon 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 briefly introduced.

Definition 2 *(MS-unbiasedness)* [5]: Let P be a family of pdfs and assume that the (misspecified) parametric model F is regular w.r.t. P. Let $\mathbf{g}(\cdot)$ be a continuously differentiable mapping from Θ to $\Phi \subset \square^s$. Let $\boldsymbol{\varphi}(\mathbf{x})$ be a statistic from the iid observations $\mathbf{x} = \{\mathbf{x}_m\}_{m=1}^M$ that takes its values in Φ. Then, $\boldsymbol{\varphi}(\mathbf{x})$ is an MS-unbiased estimator of $\mathbf{g}(\boldsymbol{\theta}_0)$, derived under the misspecified model F, iff:

$$E_p\{\boldsymbol{\varphi}(\mathbf{x})\} = \int \boldsymbol{\varphi}(\mathbf{x}) p_X(\mathbf{x}) d\mathbf{x} = \mathbf{g}(\boldsymbol{\theta}_0), \quad \forall \ p_X(\mathbf{x}) \in P \ . \tag{8}$$

As in the classical estimation framework, the function $\mathbf{g}(\cdot)$ is introduced in order to take into account all cases in which one can be interested in subsets or, more generally, in a given invertible transformation of the (pseudo-true) parameter vector.

It is easy to show that the above definition is consistent with the classical definition of unbiasedness. Without lack of generality, assume that $\mathbf{g}$ is an identity mapping, i.e. $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$. Then, the statistic $\boldsymbol{\varphi}(\mathbf{x})$ is exactly an estimator of the pseudo-true parameter vector $\boldsymbol{\theta}_0$; so, it is reasonable to use the standard notation $\boldsymbol{\varphi}(\mathbf{x}) \square \hat{\boldsymbol{\theta}}(\mathbf{x})$. When the model F is correctly specified, there exists a

$\bar{\theta} \in \Theta$ such that $p_X(\mathbf{x}) = f_X(\mathbf{x};\bar{\theta})$ for every $\mathbf{x}$. Then, from Lemma 1, $\bar{\theta} = \theta_0$ and finally eq. (8) reduces to $E_{f_{\bar{\theta}}}\{\hat{\theta}(\mathbf{x})\} = \int \hat{\theta}(\mathbf{x}) f_X(\mathbf{x};\bar{\theta}) d\mathbf{x} = \bar{\theta}$ 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) [5], [10]): *Let F be a parametric model. Let $P(F)$ be the family of all pdfs w.r.t. which F is regular. Suppose that $P(F)$ is not empty. Let $\mathbf{g}(\cdot)$ be a continuously differentiable mapping from Θ to $\Phi \subset \square^s$. Let $\boldsymbol{\varphi}(\mathbf{x})$ be an MS-unbiased estimator derived under the misspecified model F from the iid observed vectors $\mathbf{x} = \{\mathbf{x}_m\}_{m=1}^M$. Then, for every $p_X(\mathbf{x})$ in $P(F)$:*

$$\mathbf{C}\big(\boldsymbol{\varphi}(\mathbf{x}), \mathbf{g}(\theta_0)\big) \geq \frac{1}{M} \mathbf{G}_{\theta_0} \mathbf{A}_{\theta_0}^{-1} \mathbf{B}_{\theta_0} \mathbf{A}_{\theta_0}^{-1} \mathbf{G}_{\theta_0}^T \square \; \mathrm{MCRB}(\theta_0), \tag{9}$$

where

$$\mathbf{C}_p\big(\boldsymbol{\varphi}(\mathbf{x}), \mathbf{g}(\theta_0)\big) \square \; E_p\left\{\big(\boldsymbol{\varphi}(\mathbf{x}) - \mathbf{g}(\theta_0)\big)\big(\boldsymbol{\varphi}(\mathbf{x}) - \mathbf{g}(\theta_0)\big)^T\right\} \tag{10}$$

is the error covariance matrix of $\boldsymbol{\varphi}(\mathbf{x})$, the matrices $\mathbf{A}_{\theta_0}$ and $\mathbf{B}_{\theta_0}$ are defined in eqs.(1) and (4) respectively and $\mathbf{G}_{\theta_0} = \nabla_{\theta_0}^T \mathbf{g}(\theta_0)$ is the Jacobian matrix of $\mathbf{g}$ evaluated at θ_0. Following [10], we refer to the right side of eq. (9) as the Misspecified Cramér-Rao Bound (MCRB).

The proof of this theorem can be found in [5]. It can be noted that the hypothesis that $P(F)$ is not empty is not so strong. In fact, it requires that there exists at least one pdf $p_X(\mathbf{x}_m)$ for which, from Assumption A2, the point θ_0 exists [5]. In the following, we provide many examples in which it is possible to evaluate θ_0 and so the MCRB applies. Other relevant signal processing problems in which the pseudo-true vector θ_0 can be evaluated are discussed in [10] and [11].

14

It is worth noting that the MCRB is consistent with the classical CRB. As for the unbiasedness, assuming for simplicity that $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$, then $\boldsymbol{\varphi}(\mathbf{x}) \square \hat{\boldsymbol{\theta}}(\mathbf{x})$, when the model F is correctly specified, $p_X(\mathbf{x}) = f_X(\mathbf{x}; \overline{\boldsymbol{\theta}})$ for some $\overline{\boldsymbol{\theta}} \in \Theta$. Then, from Lemma 1, the matrices $-\mathbf{A}_{\overline{\boldsymbol{\theta}}}$ and $\mathbf{B}_{\overline{\boldsymbol{\theta}}}$ are equal and correspond to the classical FIM, and finally:

$$
\begin{aligned}
\mathbf{C}_{f_{\overline{\boldsymbol{\theta}}}}\left(\hat{\boldsymbol{\theta}}(\mathbf{x}), \overline{\boldsymbol{\theta}}\right) &\geq \frac{1}{M} \mathbf{A}_{\overline{\boldsymbol{\theta}}}^{-1} \mathbf{B}_{\overline{\boldsymbol{\theta}}} \mathbf{A}_{\overline{\boldsymbol{\theta}}}^{-1} = -\frac{1}{M} \mathbf{A}_{\overline{\boldsymbol{\theta}}}^{-1} = \frac{1}{M} \mathbf{B}_{\overline{\boldsymbol{\theta}}} = -\frac{1}{M}\left(E_{f_{\overline{\boldsymbol{\theta}}}}\left\{\nabla_{\overline{\boldsymbol{\theta}}} \nabla_{\overline{\boldsymbol{\theta}}}^{T} \ln f_X\left(\mathbf{x}_m; \overline{\boldsymbol{\theta}}\right)\right\}\right)^{-1} \\
&= \frac{1}{M}\left(E_{f_{\overline{\boldsymbol{\theta}}}}\left\{\nabla_{\overline{\boldsymbol{\theta}}} \ln f_X\left(\mathbf{x}_m; \overline{\boldsymbol{\theta}}\right) \nabla_{\overline{\boldsymbol{\theta}}}^{T} \ln f_X\left(\mathbf{x}_m; \overline{\boldsymbol{\theta}}\right)\right\}\right)^{-1},
\end{aligned}
\tag{11}
$$

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

Remark 1: The statement and the proof of Theorem 1, given in [5], consider only the case of *real* parameter space, i.e., $\Theta \subset \square^d$. However, as shown in [10], the derivation can be easily extended to the complex case, i.e., when $\Theta \subset \square^d$. This is because all the pdfs are *real functions of complex variables* ($\mathbf{x}$ and $\boldsymbol{\theta}$), so we do not need sophisticated holomorphic calculus to generalize the derivatives w.r.t. a complex parameter vector $\boldsymbol{\theta}$. Insightful procedures, useful to generalize the Cramér-Rao inequality in the complex case, are discussed in [21], [22], and [23].

Remark 2: In order to evaluate the MCRB of (9), the knowledge of the true pdf $p_X(\mathbf{x})$ is required. However, this should not be seen as a limitation of its applicability. Think for example of 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. the rendering of the ML estimator is difficult or impossible to derive. In these cases, one typically assumes a simpler model, such as a Gaussian model, introducing a mismatch. The evaluation of the MCRB would show the potential performance loss due to the mismatch between the assumed and the true model. An example of this procedure is discussed in this chapter, in the context of the scatter matrix estimation problem for radar detection applications. Another useful application of the MCRB is the prediction of possible

15

weaknesses (i.e. breakdown of the estimation performance) of the system under uncommon 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 pdf is known but it 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. (4), the MCRB can be approximated by means of Monte Carlo simulations.

Remark 3: Before introducing the MML estimator, we briefly comment on the difference between the results obtained in [10] and the general derivation of the MCRB provided in [5]. Even if the final mismatched covariance inequality assumes exactly the same expression (at least when **g** is an identity mapping and then $\boldsymbol{\varphi}(\mathbf{x}) \square \, \hat{\boldsymbol{\theta}}(\mathbf{x})$), the proof in [5] is general, while the one provided in [10] relies on first order Taylor expansion of the estimation error (see eq. 41 in [10]). This derivation leads to define a restricted class of estimators for which the MCRB of (9) applies which is 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.,

$$\boldsymbol{\Xi}_{\boldsymbol{\theta}} = E_p\left\{\left(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\mu}\right)\boldsymbol{\eta}_{\boldsymbol{\theta}}(\mathbf{x})^T\right\} \tag{12}$$

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. The score function used in [10] and in [14] is $\boldsymbol{\eta}_{\boldsymbol{\theta}}(\mathbf{x}) = \nabla_{\boldsymbol{\theta}} \ln f_X\left(\mathbf{x};\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}} D\left(p \| f_{\boldsymbol{\theta}}\right)$. For this function, it turns out that the correlation matrix $\boldsymbol{\Xi}_{\boldsymbol{\theta}}$ must be equal to $\pm\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}$. This

particular choice of the score function has been motivated by its "tightness" properties discussed in [24].

As shown in [10], there is at least an estimator that *asymptotically* satisfies constraints 1) and 2). This estimator is exactly the MML estimator described in the next section. However, in general, it would be very difficult to characterize explicitly a class of estimators that satisfy these two constraints. The advantage of the proof in [5] is that it shows that inequality (9) holds for all the MS-unbiased estimators and not only for those satisfying 1) and 2).

4.3 The Mismatched Maximum Likelihood (MML) estimator

In the previous subsection, the MCRB has been introduced. In particular, we showed that it represents the counterpart of the classical CRB in the presence of model misspecification. At this point, a question that naturally arises is to whether there exists a mismatched estimator whose error covariance matrix is equal (at least asymptotically) to the MCRB. As we will see, the answer is "yes" and this "misspecified-efficient" estimator is a generalization of the classical Maximum Likelihood (ML) estimator, called the Mismatched Maximum Likelihood (MML) estimator or also the Quasi Maximum Likelihood (QML) estimator [2]. In the rest of this section, we assume that **g** is an identical mapping so that $\mathbf{G}_\theta \; \square \; \nabla_\theta^T \mathbf{g}(\theta) \equiv \mathbf{I}$, $\forall \theta \in \Theta$ where **I** is the identity matrix.

The MML estimator has been introduced in [1] and [2] as:

$$\hat{\boldsymbol{\theta}}_{MML}(\mathbf{x}) = \arg\max_{\theta \in \Theta} \ln f_X(\mathbf{x}; \theta) = \arg\max_{\theta \in \Theta} \sum_{m=1}^{M} \ln f_X(\mathbf{x}_m; \theta), \tag{13}$$

where $\mathbf{x}_m \; \square \; p_X(\mathbf{x}_m)$. It can be shown (see [1] and [1]) 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 KL-divergence between $p_X(\mathbf{x}_m)$ and $f_X(\mathbf{x}_m; \theta)$:

17

$$\hat{\boldsymbol{\theta}}_{MML}\left(\mathbf{x}\right) \underset{M\to\infty}{\overset{a.s.}{\to}} \boldsymbol{\theta}_0, \tag{14}$$

Under similar regularity conditions to those given in Sect. 4.1 (Assumptions A1, A2 and A3), the asymptotic normality of the MML estimator is proved in [1] and [2]. This result is summarized by the following theorem (see [1] or [2] for the proof):

Theorem 2 ([1], [2]): *Under suitable regularity conditions, it can be proved that*

$$\sqrt{M}\left(\hat{\boldsymbol{\theta}}_{MML}\left(\mathbf{x}\right)-\boldsymbol{\theta}_0\right) \underset{M\to\infty}{\overset{d.}{\square}} N\ \left(\mathbf{0}, \mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\right), \tag{15}$$

where $\underset{M\to\infty}{\overset{d.}{\square}}$ *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* (4), *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 important fact: the MML estimator is asymptotically MS-unbiased and its error covariance matrix asymptotically equates 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. However, it must be noted 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, in the mismatched case the convergence and the MS-unbiasedness of the MML estimator is always defined w.r.t. the pseudo-true parameter vector $\boldsymbol{\theta}_0$ of eq. (2). The next subsection provides some insights about this important fact.

18

4.4 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., $T \equiv \Theta \subset \Box^d$ [15]. As before, we assume that $\mathbf{g}$ is an identical mapping so that $\boldsymbol{\varphi}(\mathbf{x}) \Box \hat{\boldsymbol{\theta}}(\mathbf{x})$. More formally, assume that the true parametric pdf of the observations $p_X(\mathbf{x})$ and the assumed pdf $f_X(\mathbf{x}; \boldsymbol{\theta})$ belong to two (generally different) families of pdf's, P and F that can be parameterized by using the same parameter space Θ:

$$P = \left\{ p_{\boldsymbol{\theta}} \,\middle|\, p_X(\mathbf{x}; \boldsymbol{\theta}) \text{ is a p.d.f. } \forall \boldsymbol{\theta} \in \Theta \right\}, \quad F = \left\{ f_{\boldsymbol{\theta}} \,\middle|\, f_X(\mathbf{x}; \boldsymbol{\theta}) \text{ is a p.d.f. } \forall \boldsymbol{\theta} \in \Theta \right\}. \tag{16}$$

An application in which this particular case arises will be discussed in Section 6.1 of this chapter. Even if this is only a particular case of the theory developed in the previous sections, this type of mismatch allows us to deeply understand the nature of the MCRB and of the MML estimator. In particular, if condition (16) 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 parameter vector $\bar{\boldsymbol{\theta}}$, and as such the difference vector $\mathbf{r} \Box \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ is well-defined. In essence, the vector $\mathbf{r}$ is in general different from a zero-vector. In particular, vector $\mathbf{r}$ indicates the distance between the convergence point $\bar{\boldsymbol{\theta}}$ of the classical ML estimator when the true pdf $p_X(\mathbf{x}; \boldsymbol{\theta})$ is perfectly known and the convergence point $\boldsymbol{\theta}_0$ of the MML estimator when the mismatched pdf $f_X(\mathbf{x}; \boldsymbol{\theta})$ which satisfies the condition in (16) is adopted. Moreover, using $\mathbf{r}$, a bound on the *Mean Square Error* (MSE) of the estimate of $\bar{\boldsymbol{\theta}}$ in the presence of mismatched models can be easily established. To proceed, the error covariance matrix of eq. (10) can be rewritten as:

$$\text{MSE}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\bar{\boldsymbol{\theta}}\right) \square \, E_p\left\{\left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0+\boldsymbol{\theta}_0-\bar{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0+\boldsymbol{\theta}_0-\bar{\boldsymbol{\theta}}\right)^T\right\}$$

$$= \mathbf{C}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\boldsymbol{\theta}_0\right) - 2E_p\left\{\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0\right\}\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)^T + \left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)^T \qquad (17)$$

$$= \mathbf{C}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\boldsymbol{\theta}_0\right) + \mathbf{r}\mathbf{r}^T,$$

where $E_p\left\{\hat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}_0\right\}=0$ due to the MS-unbiasedness assumption. A similar expansion of the MSE

can be found in [10] (see eq. 70). Finally, by substituting the covariance inequality in (9) in eq. (17),

we can obtain a misspecified bound on the MSE of $\bar{\boldsymbol{\theta}}$ as:

$$\text{MSE}_p\left(\hat{\boldsymbol{\theta}}(\mathbf{x}),\bar{\boldsymbol{\theta}}\right) \geq \frac{1}{M}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1}\mathbf{B}_{\boldsymbol{\theta}_0}\mathbf{A}_{\boldsymbol{\theta}_0}^{-1} + \mathbf{r}\mathbf{r}^T. \qquad (18)$$

Moreover, if the condition in (16) 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 a.s. to the

true parameter vector $\bar{\boldsymbol{\theta}}$, i.e., $\hat{\boldsymbol{\theta}}\left(\mathbf{x}\right) \xrightarrow[M\to\infty]{a.s.} \boldsymbol{\theta}_0 = \bar{\boldsymbol{\theta}}$.

The mismatched MSE inequality in (18) and the concept of consistency for MS-unbiased

mismatched estimators can be useful to compare in a very intuitive and self-explicative manner the

nature of the MCRB and of the MML estimator, discussed in both Section 5 and Section 6.

4.5 The constrained MCRB: CMCRB

In some applications, one has to deal with additional constraints on the unknown parameter

vector that need to be satisfied by an estimation algorithm. To this end, a constrained version of the

classical (matched) CRB has been proposed in [25]. Successive generalizations can be found in

[26], [27] and [28]. The aim of this section is to show that a generalization of the results obtained

for the constrained CRB (CCRB) to the mismatched framework is also possible and a constrained

20

version of the MCRB, i.e., the CMCRB under a set of equality constraints on the parameter vector, can be derived. This was accomplished in a recent work [29] by the generalization of the procedure described in [28]. A local bijective transformation of the unknown parameter vector in a lower dimensional parameter space is exploited to build the constraints in the CRB derivation.

Let $\hat{\boldsymbol{\theta}}(\mathbf{x})$ be a d-dimensional *MS*-unbiased (in a proper sense discussed below) estimator derived under the misspecified model F. Suppose that $\hat{\boldsymbol{\theta}}(\mathbf{x})$ is required to satisfy k (with $k<d$) continuously differentiable constraints [27]:

$$\mathbf{f}\left(\hat{\boldsymbol{\theta}}(\mathbf{x})\right) = \mathbf{0}. \tag{19}$$

The $k \times d$ Jacobian matrix of the constraints, $\mathbf{F}_{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}}^T \mathbf{f}(\boldsymbol{\theta})$ is assumed to have full rank for any $\boldsymbol{\theta} \in \Theta$ satisfying (19). Then there exists a matrix $\mathbf{U} \in \square^{d \times (d-k)}$ whose columns form an orthonormal basis for the null space of $\mathbf{F}_{\boldsymbol{\theta}}$, that is:

$$\mathbf{F}_{\boldsymbol{\theta}}\mathbf{U} = \mathbf{0}, \ \mathbf{U}^T\mathbf{U} = \mathbf{I}. \tag{20}$$

The matrix $\mathbf{U}$ can be obtained numerically by calculating the d-k orthonormal eigenvectors associated with the zero eigenvalue of $\mathbf{F}_{\boldsymbol{\theta}}$.

Using this framework, Stoica and Ng [27] derived a constrained version of the classical CRB. A different, yet equivalent, approach was adopted in [28], where the authors exploited the Implicit Function Theorem (see, e.g., [30, Th. 5-2]) to obtain the same CCRB of [27], but from a different standpoint. The starting point of the proof in [28] is that the constraint function $\mathbf{f}$ restricts $\boldsymbol{\theta}$ to a manifold $\overline{\Theta} = \{\boldsymbol{\theta} | \mathbf{f}(\boldsymbol{\theta}) = \mathbf{0}\}$ of the original vector space $\square^d$ with dimension d-k (since $\mathrm{rank}(\mathbf{F}_{\boldsymbol{\theta}}) = k$ by assumption). Therefore, from the Implicit Function Theorem, there exist two open sets O and P

of $\bar{\Theta}$ and $\square^{d-k}$, respectively, and a local continuously differentiable bijection $\mathbf{h}: P \ni \xi \to O \ni \boldsymbol{\theta}$ such that

$$\boldsymbol{\theta} = \mathbf{h}(\xi). \tag{21}$$

We denote as $\mathbf{H}_{\xi} = \nabla_{\xi}^{T}\mathbf{h}(\xi)$ the $d \times (d-k)$ full rank Jacobian matrix of the transformation $\mathbf{h}$. The idea behind the proof in [29] is the same as the one in [28]: given the MCRB, derived for an *intrinsic* parameter vector ξ_0 belonging to the reduced parameter space $\square^{d-k}$, we can "project back" such bound on the manifold $\bar{\Theta}$ by exploiting the bijective transformation $\mathbf{h}$. As we will see, the final expression of the CMCRB does not involve either the *intrinsic* parameter vector ξ_0 or $\mathbf{h}$. In particular, the explicit knowledge of ξ_0 and $\mathbf{h}$ is not required, only their existence (guaranteed by the Implicit Function Theorem) is necessary. This is an important fact, since in general, to obtain an explicit expression for $\mathbf{h}$ is not an easy task.

It is worth noticing that the constraints in eq. (19) apply to $\boldsymbol{\theta} \in \Theta$, i.e., the parameter vector that parameterizes the assumed (and possibly misspecified) pdf $f_X(\mathbf{x}; \boldsymbol{\theta})$. They are not supposed to apply to $\boldsymbol{\tau} \in \mathbf{T}$, i.e., the true and inaccessible parameter vector that in general may have, as discussed before, a completely different structure.

4.5.a The MCRB for the intrinsic parameter vector

As sketched in the previous subsection, the first step to derive the CMCRB is to analyze the conditions that guarantee the existence of the intrinsic pseudo-true parameter vector ξ_0. Then an explicit expression for the intrinsic MCRB for ξ_0 is obtained by building upon the results discussed in Sect. 4.2.

Existence of ξ_0

The assumptions needed to establish the existence of a pseudo-true intrinsic parameter vector ξ_0 are formally the same as the assumptions A1-A3 given in Sect. 4.1. In order to improve the clarity of the presentation, we recall them here by specializing them to the constrained problem at hand. In particular, we assume that:

i. The function $\zeta(\boldsymbol{\theta}) = E_p\left\{\ln f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right)\middle|\boldsymbol{\theta}\in\overline{\Theta}\right\}$ has a unique maximum on $\overline{\Theta}$ at an interior

point $\boldsymbol{\theta}_0$, and hence $\overline{\zeta}(\xi) = E_p\left\{\ln f_X\left(\mathbf{x};\mathbf{h}(\xi)\right)\middle|\xi\in\square^{d-k}\right\}$ has a unique maximum on $\square^{d-k}$ at

a point ξ_0,

ii. The matrix $\mathbf{A}_{\xi_0}$ defined as:

$$\mathbf{A}_{\xi_0} = E_p\left\{\nabla_{\xi_0}\nabla_{\xi_0}^T\ln f_X\left(\mathbf{x};\mathbf{h}(\xi_0)\right)\right\} \tag{22}$$

is non-singular. This can be recognized also as the identifiability condition (see [17], [18], [19] and [20]) for ξ_0. As before, the interior point $\overline{\Theta}\ni\boldsymbol{\theta}_0 = \mathbf{h}(\xi_0)$ can be equivalently seen as the point that minimizes the Kullback-Leibler divergence between the true distribution $p_X(\mathbf{x})$ and the assumed distribution $f_X(\mathbf{x};\boldsymbol{\theta})$ with $\boldsymbol{\theta}\in\overline{\Theta}$, i.e. $\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}\in\overline{\Theta}}{\arg\min}\left\{D\left(p_X\|f_{\boldsymbol{\theta}}\right)\right\}$. This minimization problem can be rewritten as function of the intrinsic *pseudo-true parameter* vector ξ_0 as:

$$\xi_0 = \underset{\xi\in\square^{d-k}}{\arg\min}\left\{D\left(p_X\|f_{\mathbf{h}(\xi)}\right)\right\} = \underset{\xi\in\square^{d-k}}{\arg\min}\left\{-E_p\left\{\ln f_X\left(\mathbf{x};\mathbf{h}(\xi)\right)\right\}\right\}. \tag{23}$$

MS-unbiasedness and MCRB in ξ_0

Under the two assumptions i) and ii) stated above, the definition of *MS*-unbiasedness and the expression of the MCRB given in (8) and (9) respectively, can be easily exploited to derive the MCRB in ξ_0. In particular, let $\hat{\theta}(\mathbf{x})$ be a constrained estimator derived under the misspecified model F from the iid observations $\mathbf{x} = \{\mathbf{x}_m\}_{m=1}^M$ that satisfies the set of equality constraints $\mathbf{f}\left(\hat{\theta}(\mathbf{x})\right) = \mathbf{0}$. Let $\mathbf{h}(\cdot)$ be the continuously differentiable bijective transformation in (21). Then, $\hat{\theta}(\mathbf{x})$ is an *MS*-unbiased estimator of $\mathbf{h}(\xi_0)$ iff:

$$E_p\left\{\hat{\theta}(\mathbf{x})\right\} = \int \hat{\theta}(\mathbf{x}) p_X(\mathbf{x}) d\mathbf{x} = \theta_0 = \mathbf{h}(\xi_0), \quad \forall p_X(\mathbf{x}) \in P, \ \theta_0 \in \bar{\Theta}. \tag{24}$$

Moreover, the following inequality holds true:

$$\mathbf{C}_p\left(\hat{\theta}(\mathbf{x}), \theta_0\right) = \mathbf{C}_p\left(\hat{\theta}(\mathbf{x}), \mathbf{h}(\xi_0)\right) \geq \frac{1}{M} \mathbf{H}_{\xi_0} \mathbf{A}_{\xi_0}^{-1} \mathbf{B}_{\xi_0} \mathbf{A}_{\xi_0}^{-1} \mathbf{H}_{\xi_0}^T \ \square \ \text{MCRB}(\xi_0) \tag{25}$$

where $\mathbf{C}_p\left(\hat{\theta}(\mathbf{x}), \mathbf{h}(\xi_0)\right) \square E_p\left\{\left(\hat{\theta}(\mathbf{x}) - \mathbf{h}(\xi_0)\right)\left(\hat{\theta}(\mathbf{x}) - \mathbf{h}(\xi_0)\right)^T\right\}$ is the error covariance matrix of $\hat{\theta}(\mathbf{x})$, the matrix $\mathbf{A}_{\xi_0}$ is given in (22), $\mathbf{B}_{\xi_0}$ is defined as:

$$\mathbf{B}_{\xi_0} \square E_p\left\{\nabla_{\xi_0} \ln f_X\left(\mathbf{x}; \mathbf{h}(\xi_0)\right) \nabla_{\xi_0}^T \ln f_X\left(\mathbf{x}; \mathbf{h}(\xi_0)\right)\right\}, \tag{26}$$

and $\mathbf{H}_{\xi_0} = \nabla_{\xi_0}^T \mathbf{h}(\xi_0)$ is the Jacobian matrix of $\mathbf{h}(\cdot)$ evaluated at ξ_0. The proof of the inequality in (25) follows directly by the proof of Theorem 1 by substituting the invertible transformation $\mathbf{g}(\cdot)$ with $\mathbf{h}(\cdot)$.

4.5.b The constrained MCRB (CMCRB)

Finally, from all previous results, an explicit expression for the CMCRB on θ_0 is provided in the following Theorem:

Theorem 3 [29]: *The constrained MCRB (MCRB) is given by:*

$$\text{CMCRB}(\boldsymbol{\theta}_0) = \frac{1}{M} \mathbf{U}\left(\mathbf{U}^T \mathbf{A}_{\boldsymbol{\theta}_0} \mathbf{U}\right)^{-1} \mathbf{U}^T \mathbf{B}_{\boldsymbol{\theta}_0} \mathbf{U}\left(\mathbf{U}^T \mathbf{A}_{\boldsymbol{\theta}_0} \mathbf{U}\right)^{-1} \mathbf{U}^T, \quad \boldsymbol{\theta}_0 \in \bar{\Theta}, \tag{27}$$

where the (possibly singular) matrices $\mathbf{A}_{\boldsymbol{\theta}_0}$ *and* $\mathbf{B}_{\boldsymbol{\theta}_0}$ *are defined, as in the unconstrained case, in eqs. (1) and (4).*

The proof of this theorem is given in [29] and an example of a possible application of the CMCRB to the scatter matrix estimation problem is provided in the following.

Remark 4: It is easy to verify that the CMCRB is consistent with the CCRB in [27]. When the model F is correctly specified, $p_X(\mathbf{x}) = f_X(\mathbf{x}; \bar{\boldsymbol{\theta}})$ for some $\bar{\boldsymbol{\theta}} \in \bar{\Theta}$. Then, the matrices $-\mathbf{A}_{\bar{\boldsymbol{\theta}}}$ and $\mathbf{B}_{\bar{\boldsymbol{\theta}}}$ are equal and correspond to the classical FIM, and finally:

$$\begin{aligned}
\text{CMCRB}(\bar{\boldsymbol{\theta}}) &= \mathbf{U}\left(\mathbf{U}^T \mathbf{A}_{\bar{\boldsymbol{\theta}}} \mathbf{U}\right)^{-1} \mathbf{U}^T \mathbf{B}_{\bar{\boldsymbol{\theta}}} \mathbf{U}\left(\mathbf{U}^T \mathbf{A}_{\bar{\boldsymbol{\theta}}} \mathbf{U}\right)^{-1} \mathbf{U}^T \\
&= -\mathbf{U}\left(\mathbf{U}^T \mathbf{A}_{\bar{\boldsymbol{\theta}}} \mathbf{U}\right)^{-1} \mathbf{U}^T = \mathbf{U}\left(\mathbf{U}^T \mathbf{B}_{\bar{\boldsymbol{\theta}}} \mathbf{U}\right)^{-1} \mathbf{U}^T = \text{CCRB}(\bar{\boldsymbol{\theta}}),
\end{aligned} \tag{28}$$

which represents the constrained CRB obtained in [27].

To conclude this section, it is worth noting that the derivation of an efficient constrained estimator able to achieve (at least asymptotically) the CMCRB is still an open problem. Moreover, the effects of inequality constraints need to be investigated as well.

5 Two illustrative examples

In order to clarify the use of the MCRB and the MML estimator, two simple toy examples are described in the following. The problem is either to estimate the mean value (Example 1) [14] or the

variance (Example 2) of Gaussian data [15]. Assume a set of M iid scalar observations $\mathbf{x} = \{x_m\}_{m=1}^{M}$, distributed according to a Gaussian pdf with mean value μ_X and variance σ_X^2, i.e.

$$p_X(x_m) \equiv N\ (\mu_X, \sigma_X^2).$$

Example 1: Estimation of the mean value: $\bar{\theta} = \mu_X$

It is well known that, given the set of Gaussian observations $\mathbf{x}$, the ML estimator of the mean value is given by the *sample mean estimator*, i.e. $\hat{\theta}_{ML} = \dfrac{1}{M} \sum_{m=1}^{M} x_m$. Suppose now that there is a mismatch in the assumed data variance. In other words, we assume for the data a Gaussian pdf $f_X(x_m; \theta) = N\ (\theta, \sigma^2)$ with a variance σ^2 that is in general different from σ_X^2. 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 Sect. 4.4. Following eq. (13), a MML estimator for the mean value of the data is given by:

$$\hat{\theta}_{MML}(\mathbf{x}) = \arg\max_{\theta \in \Theta} \ln f_X(\mathbf{x}; \theta) = \arg\max_{\theta \in \Theta} \sum_{m=1}^{M} \ln f_X(x_m; \theta), \tag{29}$$

where

$$\ln f_X(x_m; \theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}(x_m - \theta)^2. \tag{30}$$

It is immediately clear that the MML estimator coincides with the ML estimator, i.e.,

$$\hat{\theta}_{MML}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^{M} x_m = \hat{\theta}_{ML}(\mathbf{x}). \tag{31}$$

Following the theory (see eq. (14)), we know that the MML estimator converges a.s. to that point that minimizes the KL divergence between the true pdf $p_X(x_m)$ and the assumed pdf $f_X(x_m;\theta)$. The KL divergence between $p_X(x_m)$ and $f_X(x_m;\theta)$ is given by [31]:

$$D(p_X \| f_\theta) = \frac{(\mu_X - \theta)^2}{2\sigma^2} + \frac{1}{2}\left(\frac{\sigma_X^2}{\sigma^2} - 1 - \ln\frac{\sigma_X^2}{\sigma^2}\right). \tag{32}$$

By taking the derivative with respect to θ and by setting the result equal to 0, we obtain:

$$\left.\frac{\partial D(p_X \| f_\theta)}{\partial \theta}\right|_{\theta=\theta_0} = \left.\frac{\mu_X - \theta}{\sigma^2}\right|_{\theta=\theta_0} = 0, \tag{33}$$

whose solution is $\theta_0 = \mu_X = \bar{\theta}$. Eq. (33) shows that the MML converges a.s. to the true mean value and then, according to the definition provided in Sect. 4.4, it is a *consistent* estimator. From the scalar version of eq. (8), the mean value of the MML estimator w.r.t. the true joint pdf $p_X(\mathbf{x})$ is:

$$E_p\left\{\hat{\theta}_{MML}(\mathbf{x})\right\} = \mu_X = \theta_0 = \bar{\theta}. \tag{34}$$

Hence, according to Definition 1 given in Sect. 4.2, the MML estimator is *MS-unbiased*. The MCRB can be evaluated as shown in eq. (18) by evaluating the first and the second derivative of the $\ln f_X(x_m;\theta)$ as:

$$\frac{\partial \ln f_X(x_m;\theta)}{\partial \theta} = -\frac{1}{\sigma^2}(x_m - \theta), \quad \frac{\partial^2 \ln f_X(x_m;\theta)}{\partial \theta^2} = \frac{1}{\sigma^2}. \tag{35}$$

In this case, matrices $\mathbf{A}_{\theta_0}$ of eq. (1) and $\mathbf{B}_{\theta_0}$ of eq. (4) are scalars and are easily derived to be:

$$A_{\theta_0} = E_p\left\{\frac{1}{\sigma^2}\right\} = \frac{1}{\sigma^2}, \tag{36}$$

$$B_{\theta_0} = E_p \left\{ \frac{1}{\sigma^4} (x_k - \theta_0)^2 \right\} = \frac{1}{\sigma^4} E_p \left\{ (x_k - \mu_X)^2 \right\} = \frac{\sigma_X^2}{\sigma^4}. \tag{37}$$

Finally, from (18), we have that:

$$\mathrm{MCRB}(\mu_X) = \sigma^2 \frac{\sigma_X^2}{M\sigma^4} \sigma^2 = \frac{\sigma_X^2}{M} = \mathrm{CRLB}(\mu_X). \tag{38}$$

Note that, for a consistent estimator **r=0**. The fact that the MCRB and the CRLB are equal is in accordance with the fact that the MML estimator is equal to the ML estimator and it does not depend on the misspecified variance σ^2.

Example 2: **Estimation of the variance:** $\bar{\theta} = \sigma_X^2$

In this example we consider the problem of estimating the variance of a Gaussian pdf in the presence of misspecified mean value, e.g., we erroneously assume that the mean value is zero when it is not. It is easy to show that, given the observation vector **x**, the ML estimator of the variance is given by $\hat{\theta}_{ML}(\mathbf{x}) = M^{-1} \sum_{m=1}^{M} (x_m - \mu_X)^2$, where, as before, $p_X(x_m) \equiv N\ (\mu_X, \sigma_X^2)$. Suppose now that the assumed Gaussian pdf is $f_X(x_m; \theta) \equiv N\ (\mu, \theta)$, so we misspecify the mean value. As in Example 1, the true unknown model $p_X(x_m)$ and the assumed model $f_X(x_m; \theta)$ admit the same parameterization. Hence, also in this case, we can apply the results of Sect. 4.4. From eq. (13), the MML estimator of the variance can be derived as:

$$\hat{\theta}_{MML}(\mathbf{x}) = \arg \max_{\theta \in \Theta} \ln f_X(\mathbf{x}; \theta) = \arg \max_{\theta \in \Theta} \sum_{m=1}^{M} \ln f_X(x_m; \theta), \tag{39}$$

where

$$\ln f_X(x_m; \theta) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \theta - \frac{1}{2\theta} (x_m - \mu)^2. \tag{40}$$

28

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

$$\hat{\theta}_{MML}(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^{M} (x_m - \mu)^2 . \tag{41}$$

In this case, the KL divergence between $p_X(x_m)$ and $f_X(x_m;\theta)$ can be expressed as [31]:

$$D(p_X \| f_\theta) = \frac{(\mu_X - \mu)^2}{2\theta} + \frac{1}{2}\left(\frac{\sigma_X^2}{\theta} - 1 - \ln \frac{\sigma_X^2}{\theta} \right). \tag{42}$$

By taking the derivative w.r.t. θ and by setting equal to zero the result, we get:

$$\left.\frac{\partial D(p\|f_\theta)}{\partial \theta}\right|_{\theta=\theta_0} = \left.\frac{-(\mu_X - \mu)^2 - \sigma_X^2}{2\theta^2} + \frac{1}{2\theta}\right|_{\theta=\theta_0} = 0 . \tag{43}$$

Hence, $\theta_0 = \sigma_X^2 + (\mu_X - \mu)^2 \neq \bar{\theta}$, i.e., 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 of this example is not consistent. From the scalar version of eq. (8), the mean value of the MML estimator with respect to the true distribution $p_X(\mathbf{x};\bar{\theta})$ is:

$$E_p\left\{ \hat{\theta}_{MML}(\mathbf{x}) \right\} = \sigma_X^2 + (\mu_X - \mu)^2 = \theta_0 . \tag{44}$$

Hence, the MML estimator is *MS-unbiased* and the MCRB can be evaluated as shown in (18). 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. (4) are obtained:

$$A_{\theta_0} = -\frac{1}{2\theta_0^2}, \quad B_{\theta_0} = \frac{3\sigma_X^4 + 6\sigma_X^2(\mu_X - \mu)^2 + (\mu_X - \mu)^4 - \theta_0^2}{4\theta_0^4}. \tag{45}$$

Finally, from (18), we have that:

$$\mathrm{MCRB}\left(\bar{\theta}\right) = \mathrm{MCRB}\left(\sigma_X^2\right) = \frac{2\sigma_X^4}{M} + \frac{4\sigma_X^2(\mu_X - \mu)^2}{M} + (\mu_X - \mu)^4. \tag{46}$$

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, we obtain $\mathrm{MCRB}(\sigma_X^2) \geq \mathrm{CRB}(\sigma_X^2)$, i.e., for this estimation problem, the MCRB is always

greater than or equal to the CRLB. This is an intuitive result, since if we assume a wrong data

model the best performance that we can achieve are worse than that in the case where we assume a

correct data model. However, a general proof of an inequality between MCRB and CRB in the

general case is still an open problem. When $\mu = \mu_X$, i.e. we correctly specify the mean value, then

$\theta_0 = \bar{\theta} = \sigma_X^2$ and $\mathrm{MCRB}(\sigma_X^2) = \mathrm{CRB}(\sigma_X^2)$.

In Figs. 1 and 2, we report the behavior of the square root of the MSE (RMSE) of the MML

estimator in eq. (41), the square root of the MCRB and of the CRLB, as function of the mismatched

parameter (in this case the mean value μ) and as function of the number M of available data,

respectively. In our simulation, the true mean value and the true variance are assumed to be $\mu_X=3$

and $\sigma_X^2 = 4$. As we can see from Fig. 1, the MCRB and the CRB are equal only when the true pdf

and the assumed pdf are equal, i.e., $\mu=\mu_X=3$ (the number of available data M is equal to 10). In Fig.

2, the RMSE, the MCRB and the CRB are plotted as a function of the available number of data M.

In this case, the assumed mean value is $\mu=0$, hence there is some model mismatch. It is evident that

the MCRB is a tight bound for the MML estimator, whereas the CRB is not.

Figure 1 – Comparison among the MSE of the MML estimator, the MCRB and the CRB as function of the assumed mean value.

Figure 2 – Comparison among the MSE of the MML estimator, the MCRB and the CRB as function of the available number of data

M.

30

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

In this section, we show the application of the MCRB theory to a realistic radar signal processing problem: the estimation of the disturbance covariance matrix from a set of acquired data vectors [15] (the so-called secondary data in the radar jargon). We recast this specific radar problem in the more general problem of estimating the $N \times N$ scatter matrix of Complex Elliptically Symmetric (CES) distributed data, given M iid realizations of the N-dimensional data vector $\mathbf{x}_m$, 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 [16]. The CES distributions, due to their flexibility, are widely applied in many areas, such as radar, sonar, and communications (see e.g. [16], [32], [33], [34], [35], [36] and [37]).

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

$$p_X\left(\mathbf{x}_m\right) = c_{N,h} \left|\boldsymbol{\Sigma}\right|^{-1} h\left(\left(\mathbf{x}_m - \boldsymbol{\gamma}\right)^H \boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_m - \boldsymbol{\gamma}\right)\right), \tag{47}$$

where h is the density generator, $c_{N,h}$ is a normalizing constant, $\boldsymbol{\gamma} \sim E_p\{\mathbf{x}_m\}$ and $\boldsymbol{\Sigma}$ is the normalized (or shape) covariance matrix, also called *scatter matrix*. It is important to note, that $\boldsymbol{\gamma}$, $\boldsymbol{\Sigma}$, and $h(\cdot)$ do not uniquely identify a CES distribution. In fact, given a CES distributed random vector $\mathbf{x}_m \sim CE_N(\boldsymbol{\gamma}, \boldsymbol{\Sigma}, h)$, for any $\alpha > 0$, we may define $\tilde{\boldsymbol{\Sigma}} = \alpha \boldsymbol{\Sigma}$ and $\tilde{h}(t) = h(t/\alpha)$ so that $CE_N(\boldsymbol{\gamma}, \boldsymbol{\Sigma}, h) \equiv CE_N(\boldsymbol{\gamma}, \tilde{\boldsymbol{\Sigma}}, \tilde{h})$ [16]. This ambiguity can be avoided by imposing a constraint on the scatter matrix $\boldsymbol{\Sigma}$, e.g. $\mathrm{tr}(\boldsymbol{\Sigma}) = N$, or by restricting the functional form of $h(\cdot)$ in a suitable way. The

difference between these two approaches is clarified in Section 6.1 and Section 6.2. Moreover, by imposing the constraint $\text{tr}(\boldsymbol{\Sigma}) = N$, if $\mathbf{M} \square E_p\{(\mathbf{x}_m - \boldsymbol{\gamma})(\mathbf{x}_m - \boldsymbol{\gamma})^H\}$ is the covariance matrix of the vector $\mathbf{x}_m$, then $\boldsymbol{\Sigma} = \left(N/\text{tr}(\mathbf{M})\right) \cdot \mathbf{M}$. It is important to observe that, for some CES distributions, the covariance matrix $\mathbf{M}$ does not exist, but the scatter matrix $\boldsymbol{\Sigma}$ is still well defined. Based upon the Stochastic Representation Theorem [16], any $\mathbf{x}_m \square CE_N(\boldsymbol{\gamma}, \boldsymbol{\Sigma}, h)$ with $rank(\boldsymbol{\Sigma}) = k \leq N$ admits the stochastic representation $\mathbf{x}_m =_d \boldsymbol{\gamma} + R\mathbf{T}\mathbf{u}$, where the notation $=_d$ means "has same distribution as". The non-negative random variable (r.v.) $R \square \sqrt{Q}$, the so-called *modular variate*, is a real, non-negative random variable, $\mathbf{u}$ 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 $\boldsymbol{\Sigma} = \mathbf{T}\mathbf{T}^H$ is a factorization of $\boldsymbol{\Sigma}$, where $\mathbf{T}$ is a Nxk matrix and $rank(\mathbf{T}) = k$. In the following derivations, we assume that $\boldsymbol{\Sigma}$ is full-rank, i.e., $rank(\mathbf{T}) = rank(\boldsymbol{\Sigma}) = N$, and that it is real. For the CES distributions, the term $\sigma_X^2 \square E\{Q\}/N$ can be interpreted as the statistical power of the random vector $\mathbf{x}_m$, i.e., the covariance matrix $\mathbf{M}$ and the scatter matrix $\boldsymbol{\Sigma}$ are related 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 when its mean vector $\boldsymbol{\gamma}$, scatter matrix $\boldsymbol{\Sigma}$, shape parameter λ, and scale parameter η are perfectly known [16]. Since in many scenarios (e.g., radar and sonar) the disturbance vectors are zero mean, we set $\boldsymbol{\gamma} = \mathbf{0}$ in all the following analyses.

The application of the mismatched estimation framework to the problem of estimating the scatter matrix of a CES distributed random vector is described in two steps. First, we assume that all the characteristic parameters of the particular CES distributions at hands are *a-priori* known, except for the elements of the scatter matrix $\boldsymbol{\Sigma}$. Since, by definition, the scatter matrix is symmetric and of $N \times N$ dimension, the parameters to be estimated are the $l = N(N+1)/2$ elements of the lower (or

upper) triangular sub-matrix of $\mathbf{\Sigma}$. Then, the parameter vector that parameterizes a zero-mean CES distribution can be defined as $\mathbf{\theta} = \text{vecs}(\mathbf{\Sigma})$, where the vecs-operator is the "symmetric" counterpart of the standard vec-operator that maps a symmetric $N \times N$ matrix $\mathbf{\Sigma}$ to a l-dimensional vector $\mathbf{\theta}$ whose entries are the elements of the lower (or upper) triangular sub-matrix of $\mathbf{\Sigma}$ [38]. In this case, the ambiguity between the scatter matrix and the density generator h is removed by the assumption of the *a-priori* knowledge of the extra-parameters, i.e., we assume to know exactly the density generator. Consequently, the constraint on the trace of $\mathbf{\Sigma}$ is not required. It is worth noting that this is an unrealistic case, since in practical situations the *a-priori* knowledge of the extra-parameters which characterize the CES distributions is generally not available. However, this knowledge provides better understanding of how to apply the results on the MCRB and the MML estimator to this estimation problem. The more realistic case of unknown extra-parameters is instead investigated in Section 6.2, where the problem of the joint estimation of the scatter matrix and of the extra-parameters in the presence of misspecification is addressed.

6.1 Misspecified estimation of the scatter matrix with perfectly known extra-parameters

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

$$p_X\left(\mathbf{x}_m\right) \Box\ p_X\left(\mathbf{x}_m;\overline{\mathbf{\Sigma}}\right) = c_{N,h}\left|\overline{\mathbf{\Sigma}}\right|^{-1} h\left(\mathbf{x}_m^H\overline{\mathbf{\Sigma}}^{-1}\mathbf{x}_m\right), \tag{48}$$

$$f_X\left(\mathbf{x}_m;\mathbf{\theta}\right)\Box\ f_X\left(\mathbf{x}_m;\mathbf{\Sigma}\right) = c_{N,g}\left|\mathbf{\Sigma}\right|^{-1} g\left(\mathbf{x}_m^H\mathbf{\Sigma}^{-1}\mathbf{x}_m\right), \tag{49}$$

33

where $\bar{\boldsymbol{\theta}} = \mathrm{vecs}(\bar{\boldsymbol{\Sigma}})$, $\boldsymbol{\theta} = \mathrm{vecs}(\boldsymbol{\Sigma})$, h is the density generator of the true pdf, and g is the density generator of the assumed pdf. We propose three different case studies:

- *Case Study 1*. Assumed pdf: complex Normal; true pdf: *t*-student.

- *Case Study 2*. Assumed pdf: complex Normal, true pdf: Generalized Gaussian.

- *Case Study 3*. Assumed pdf: Generalized Gaussian; true pdf: *t*-student.

It can be noted that the true unknown pdf $p_X\left(\mathbf{x}_m;\bar{\boldsymbol{\Sigma}}\right)$ and the assumed pdf $f_X\left(\mathbf{x}_m;\boldsymbol{\Sigma}\right)$ admit the same parameterization, so these examples fall in the particular case addressed in Sect. 4.4.

Case Study 1 Assumed pdf: complex Normal; true pdf: *t*-student.

We assume a complex Normal model for the data, i.e., each iid complex vector of the available dataset $\mathbf{x} = \left\{\mathbf{x}_m\right\}_{m=1}^{M}$ is distributed according to a complex Normal multivariate pdf, which belongs to the CES family:

$$f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right) \square \; f_X\left(\mathbf{x}_m;\boldsymbol{\Sigma}\right) = \frac{1}{\left(\pi\sigma^2\right)^N |\boldsymbol{\Sigma}|} \exp\left(-\frac{\mathbf{x}_m^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_m}{\sigma^2}\right). \tag{50}$$

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

$$p_X\left(\mathbf{x}_m;\bar{\boldsymbol{\theta}}\right) \square \; p_X(\mathbf{x}_m;\bar{\boldsymbol{\Sigma}}) = \frac{1}{\pi^N |\bar{\boldsymbol{\Sigma}}|} \frac{\Gamma\left(N+\lambda\right)}{\Gamma\left(\lambda\right)} \left(\frac{\lambda}{\eta}\right)^{\lambda} \left(\frac{\lambda}{\eta} + \mathbf{x}_m^H \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{x}_m\right)^{-(N+\lambda)}, \tag{51}$$

where λ is the shape parameter and η is the scale parameter characterizing the model ([16], [37]). The complex *t*-distribution has tails heavier than the Normal one for every $\lambda \in (0,\infty)$, while the limiting case $\lambda \to \infty$ yields the complex Normal distribution.

34

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),

$\hat{\mathbf{M}}_{MML} = M^{-1} \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^H$, so [39]:

$$\hat{\mathbf{\Sigma}}_{MML} = \frac{\hat{\mathbf{M}}_{MML}}{\sigma^2} = \frac{1}{M\sigma^2} \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^H , \tag{52}$$

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 $p_X(\mathbf{x}_m)$, considered here as a generic element of the CES family, and $f_X(\mathbf{x}_m; \mathbf{\Sigma})$ (the complex Normal pdf). This matrix is the convergence point of the MML estimator in eq. (52). The differential of the KL divergence with respect to $\mathbf{\Sigma}$ is given by [40]:

$$\partial D(p_X \| f_\Sigma) = -E_p\{\partial \ln f_X(\mathbf{x}_m; \mathbf{\Sigma})\} = -E_p\left\{\partial \ln |\mathbf{\Sigma}|^{-1} + \partial \ln g\left(\mathbf{x}_m^H \mathbf{\Sigma}^{-1} \mathbf{x}_m\right)\right\}$$
$$= \text{tr}\left(\mathbf{\Sigma}^{-1} \partial \mathbf{\Sigma}\right) + \text{tr}\left(E_p\left\{\frac{d \ln g(Q_\Sigma)}{dQ_\Sigma} \mathbf{\Sigma}^{-1} \mathbf{x}_m \mathbf{x}_m^H \mathbf{\Sigma}^{-1} \partial \mathbf{\Sigma}\right\}\right), \tag{53}$$

where

$$Q_\Sigma \ \Box \ \mathbf{x}_m^H \mathbf{\Sigma}^{-1} \mathbf{x}_m \tag{54}$$

The last equality in eq. (53) follows directly from the same calculus given in [35] and [41]. Since the assumed distribution $f_X(\mathbf{x}_m; \mathbf{\Sigma})$ is a complex Normal distribution, then $g(Q_\Sigma) = \exp(-Q_\Sigma/\sigma^2)$ and $d \ln g(Q_\Sigma)/dQ_\Sigma = -1/\sigma^2$. By substituting this result in eq. (53), we get:

$$\partial D(p_X \| f_\Sigma) = \text{tr}\left(\mathbf{\Sigma}^{-1} \partial \mathbf{\Sigma}\right) - \frac{1}{\sigma^2} \text{tr}\left(E_p\left\{\mathbf{\Sigma}^{-1} \mathbf{x}_m \mathbf{x}_m^H \mathbf{\Sigma}^{-1} \partial \mathbf{\Sigma}\right\}\right) = \text{tr}\left(\left[\mathbf{\Sigma}^{-1} - \frac{\sigma_X^2}{\sigma^2} \mathbf{\Sigma}^{-1} \bar{\mathbf{\Sigma}} \mathbf{\Sigma}^{-1}\right] \partial \mathbf{\Sigma}\right), \tag{55}$$

where we used the property $E_p\left\{\mathbf{x}_m\mathbf{x}_m^H\right\}=\sigma_X^2\overline{\mathbf{\Sigma}}$. Then, following the standard rules of matrix calculus [40], the derivative of the KL divergence w.r.t. $\mathbf{\Sigma}$ is:

$$\frac{\partial}{\partial\mathbf{\Sigma}}D\left(p_X\|f_\mathbf{\Sigma}\right)=\mathbf{\Sigma}^{-1}-\frac{\sigma_X^2}{\sigma^2}\mathbf{\Sigma}^{-1}\overline{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}. \tag{56}$$

Finally, by setting the derivative in eq. (56) equal to zero, we obtain matrix $\mathbf{\Sigma}_0$ that minimizes the KL divergence as:

$$\mathbf{\Sigma}_0=\frac{\sigma_X^2}{\sigma^2}\overline{\mathbf{\Sigma}}. \tag{57}$$

Eq. (57) shows that the MML estimator converges *a.s.* to a scaled version of the true scatter matrix, $\hat{\mathbf{\Sigma}}_{MML}(\mathbf{x})\underset{M\to\infty}{\overset{a.s.}{\to}}\mathbf{\Sigma}_0=\left(\sigma_X^2/\sigma^2\right)\overline{\mathbf{\Sigma}}$, so it is not consistent. It is consistent only when the two powers of the assumed and true pdf's are equal. The mean value of the MML estimator with respect to the true distribution is:

$$\mathbf{\mu}=E_p\left\{\hat{\mathbf{\Sigma}}_{MML}(\mathbf{x})\right\}=\frac{\sigma_X^2}{\sigma^2}\overline{\mathbf{\Sigma}}=\mathbf{\Sigma}_0. \tag{58}$$

Hence, the MML estimator is *MS-unbiased*. Given the MS-unbiasedness of the proposed MML estimator, we can evaluate the MCRB. In [41], the MCRB on the estimation of the scatter matrix was evaluated for two CES distributions, the complex-*t* and the Generalized Gaussian (GG), when the assumed misspecified distribution is a complex Normal pdf. In this case study, we assume that the true distribution is a complex-*t* distribution with pdf given in eq. (51). The GG case will be discussed in the next case study.

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. The power $\sigma_X^2\,\square\,E_p\left\{Q_{\overline{\mathbf{\Sigma}}}\right\}/N$ is

36

function of these two parameters. By applying the Stochastic Representation Theorem, we have that $Q_{\bar{\Sigma}}$ has an F-distribution [16] such that

$$p_{Q_{\bar{\Sigma}}}(q) = \frac{1}{B(N,\lambda)} q^{N-1} \left(\frac{\lambda}{\eta}\right)^{\lambda} \left(\frac{\lambda}{\eta}+q\right)^{-(N+\lambda)}, \qquad (59)$$

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

$$\sigma_X^2 = \frac{E_p\{Q_{\bar{\Sigma}}\}}{N} = \frac{\lambda}{\eta(\lambda-1)}, \qquad (60)$$

$$E_p\{Q_{\bar{\Sigma}}^2\} = \sigma_X^4 \frac{N(N+1)(\lambda-1)}{(\lambda-2)}, \quad \lambda > 2. \qquad (61)$$

In order to focus on the impact of the mismatch due to the difference between the density generators (or, in other words, in order to make the vector **r** in eq. (18) equals to zero), we assume that $\sigma_X^2 = \sigma^2$, so that $\boldsymbol{\Sigma}_0 = \bar{\boldsymbol{\Sigma}}$. This guarantees that the MML estimator is *consistent* and λ and η are chosen accordingly. In essence, we can set $\sigma_X^2 = \sigma^2 = 1$ and then, from eq. (60), we chose λ and η to satisfy $\eta = \lambda/(\lambda-1)$. It is worth noting that in practical situations, we have no control on the extra-parameters of the true distribution. However, this analysis is useful to better understand the nature of the MCRB and of the MML estimator. The more realistic case in which the power σ^2 is jointly estimated with $\boldsymbol{\Sigma}$ is discussed in Section 6.2.

A compact expression for the MCRB for two distributions in the CES family is given in Appendix C. Following the results in [41] and by applying eq. (C.10), the MCRB can be expressed as:

$$\mathrm{MCRB}\left(\overline{\boldsymbol{\theta}}\right) = \frac{1}{M} \mathbf{D}_N^\dagger \left[\frac{1}{(\lambda-2)} \mathrm{vec}\left(\overline{\boldsymbol{\Sigma}}\right) \mathrm{vec}\left(\overline{\boldsymbol{\Sigma}}\right)^T + \frac{(\lambda-1)}{(\lambda-2)} \overline{\boldsymbol{\Sigma}} \otimes \overline{\boldsymbol{\Sigma}} \right] \left(\mathbf{D}_N^\dagger\right)^T. \tag{62}$$

where $\mathbf{D}_N$ is the so-called Duplication matrix of order N ([38], [42], [43]). The duplication matrix is implicitly defined as the unique $N^2 \times l$ matrix (where $l = N(N+1)/2$) 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 FIM for t-distributed data evaluated in [35] and the properties of the vec and vecs operators, the duplication matrix $\mathbf{D}_N$ and of the Kronecker product $\otimes$ ([38], [42], [43], [44]), the CRB can be expressed as:

$$\mathrm{CRB}\left(\overline{\boldsymbol{\theta}}\right) = \frac{1}{M} \mathbf{D}_N^\dagger \left[\frac{N+\lambda+1}{\lambda(N+\lambda)} \mathrm{vec}\left(\overline{\boldsymbol{\Sigma}}\right) \mathrm{vec}\left(\overline{\boldsymbol{\Sigma}}\right)^T + \frac{N+\lambda+1}{N+\lambda} \overline{\boldsymbol{\Sigma}} \otimes \overline{\boldsymbol{\Sigma}} \right] \left(\mathbf{D}_N^\dagger\right)^T, \tag{63}$$

For the sake of comparison, in the following figures we report, along with the MSE of the MML, the MCRB and the CRLB, as well as the MSE of the robust (unconstrained) Tyler's estimator ([45], [46], [47], [48]). Tyler's estimator has been derived in the context of the CES distribution as the most robust estimator in min-max sense [48]. In particular, Tyler's estimator can be obtained as the recursive solution of the following (unconstrained) fixed-point (FP) matrix equation:

$$\boldsymbol{\Sigma} = \frac{N}{M} \sum_{m=1}^{M} \frac{\mathbf{x}_m \mathbf{x}_m^H}{\mathbf{x}_m^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_m}. \tag{64}$$

To solve eq. (64), we use the following iterative approach:

$$\begin{cases} \hat{\boldsymbol{\Sigma}}_T^{(0)} = \hat{\boldsymbol{\Sigma}}_{MML} \\ \hat{\boldsymbol{\Sigma}}_T^{(k+1)} = \dfrac{N}{M} \displaystyle\sum_{m=1}^{M} \dfrac{\mathbf{x}_m \mathbf{x}_m^H}{\mathbf{x}_m^H \left(\hat{\boldsymbol{\Sigma}}_T^{(k)}\right)^{-1} \mathbf{x}_m}, \, k = 0,\ldots,K \end{cases} \tag{65}$$

It can be noted that, unlike the recursive procedure proposed in [45], in (65) there is not a normalization constraint on the trace of $\hat{\boldsymbol{\Sigma}}_T^{(k)}$. The MCRB in (9) does not apply to the Tyler's estimator since $\hat{\boldsymbol{\Sigma}}_T^{(k)}$ 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:

$$\varepsilon \triangleq \left\| E_p \left\{ \left(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \bar{\boldsymbol{\theta}} \right) \left(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \bar{\boldsymbol{\theta}} \right)^T \right\} \right\|_F , \tag{66}$$

where $\hat{\boldsymbol{\theta}} = \mathrm{vecs}(\hat{\boldsymbol{\Sigma}})$, $\hat{\boldsymbol{\Sigma}}$ is an estimate of the true covariance matrix $\boldsymbol{\Sigma}$, $\bar{\boldsymbol{\theta}} = \mathrm{vecs}(\bar{\boldsymbol{\Sigma}})$ and $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}\left(\mathbf{A}^T \mathbf{A}\right)}$ is the Frobenius norm of matrix $\mathbf{A}$. Fig. 3 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:

$$\varepsilon_{\mathrm{MCRB}} \triangleq \left\| \mathrm{MCRB}\left(\bar{\boldsymbol{\theta}}\right) \right\|_F , \quad \varepsilon_{\mathrm{CRB}} \triangleq \left\| \mathrm{CRB}\left(\bar{\boldsymbol{\theta}}\right) \right\|_F \tag{67}$$

The true covariance matrix is assumed to be $[\boldsymbol{\Sigma}]_{i,j} = \rho^{|i-j|}$ [49], [50]. The value of the one-lag correlation coefficient is $\rho = 0.8$, the number of (secondary) data 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 CRB tend to be equal, since for $\lambda \to \infty$ the t-student pdf tends to a complex Gaussian pdf, and the matched and mismatched models tend to coincide. Moreover, as $\lambda \to \infty$, the MML estimator converges to the ML estimator, and then it attains the CRB. 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. 4, ε_{MML}, ε_{Tyler}, ε_{MCRB} and ε_{CRB} 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 ([46], [48]). For completeness, in Fig. 5 we investigate the performance of the MML and of Tyler's estimator as function of ρ for $\lambda = 3$, $N=8$ and $M=3N$. As expected, Tyler's estimator achieves better performance than the MML estimator 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 $\overline{\Sigma}$ (at least in its unconstrained version) its RMSE is always above the CRB.

Figure 3 – MSE indices, ε_{MML} and ε_{Tyler}, and bounds ε_{MCRB} and ε_{CRB} as function of the shape parameter of the *t*-distribution (ρ=0.8, *N*=8, *M*=3*N*).

Figure 4 – MSE indices, ε_{MML} and ε_{Tyler}, and bounds ε_{MCRB} and ε_{CRB} as function of the available data (ρ=0.8, *N*=8, λ=3).

Figure 5 – MSE indices, ε_{MML} and ε_{Tyler}, and bounds ε_{MCRB} and ε_{CRB} as function of the one-lag correlation coefficient ρ (λ=3, *N*=8, *M*=3*N*).

Case Study 2 Assumed pdf: complex Normal, true pdf: Generalized Gaussian.

As before, we assume a complex Normal model for the data, while the true data distribution is a GG distribution:

$$p_X\left(\mathbf{x}_m;\overline{\mathbf{\theta}}\right) \Box \ p_X\left(\mathbf{x}_m;\overline{\mathbf{\Sigma}}\right) = \frac{\beta\Gamma(N)b^{-N/\beta}}{\pi^N\Gamma(N/\beta)}\frac{1}{\left|\overline{\mathbf{\Sigma}}\right|}\exp\left(-\frac{\left(\mathbf{x}_m^H\overline{\mathbf{\Sigma}}^{-1}\mathbf{x}_m\right)^\beta}{b}\right), \tag{68}$$

where b is the scale parameter and β is the shape parameter characterizing the model [16], [51]. One advantage of the GG distribution with respect to the t-distribution is that it can be used to model both the heavy tailed ($\beta<1$) and the light tailed ($\beta<1$) distributions as compared to the Normal distribution ($\beta=1$).

Since we are assuming a complex Normal model for the data, the MML estimator can be derived exactly as discussed in the Case Study 1. In particular, the MML scatter matrix estimator for this mismatched scenario is still the SCM given in eq. (52). Moreover its convergence point, i.e. the point that minimizes the KL divergence between the true GG distribution and the assumed Normal distribution, is again the matrix $\mathbf{\Sigma}_0 = \left(\sigma_X^2/\sigma^2\right)\bar{\mathbf{\Sigma}}$ of eq. (57), where σ_X^2 is the power related to the GG distribution in eq. (68). As in the Case Study 1, in order to focus our analysis on the effects of the misspecification between the true and the assumed density generators (i.e. in order to make the vector $\mathbf{r}$ in eq. (18) equal to zero), we choose the shape β and scale b of the GG distribution such that $\sigma_X^2 = \sigma^2$, i.e. the MML estimator is consistent. The power $\sigma_X^2 \square E_p\left\{Q_{\bar{\Sigma}}\right\}/N$ of the GG distribution is function of β and b. In fact, by applying the Stochastic Representation Theorem, it can be shown that $Q_{\bar{\Sigma}}$ has pdf:

$$p_{Q_{\bar{\Sigma}}}(q) = \frac{\beta q^{N-1}}{b^{N/\beta}\Gamma\left(N\beta^{-1}\right)} e^{-\frac{q^\beta}{b}} . \tag{69}$$

Hence, we have:

$$\sigma_X^2 = E_p\left\{Q_{\bar{\Sigma}}\right\}/N = b^{1/\beta}\Gamma\left(\frac{N+1}{\beta}\right)\bigg/ N\Gamma\left(\frac{N}{\beta}\right), \tag{70}$$

$$E_p\left\{Q_{\bar{\Sigma}}^2\right\} = b^{2/\beta}\Gamma\left(\frac{N+2}{\beta}\right)\bigg/ \Gamma\left(\frac{N}{\beta}\right) = \sigma_X^4 N^2 \frac{\Gamma(N/\beta)\Gamma((N+2)/\beta)}{\Gamma((N+1)/\beta)^2} . \tag{71}$$

As in the Case Study 1, we set $\sigma_X^2 = \sigma^2 = 1$, and then, from eq. (70), β and b should be chosen to satisfy $b = \left[\Gamma(N + 1/\beta)/\Gamma(N/\beta)\right]^{\beta}$.

Following the results in [41] and by applying eq. (C.10), the MCRB can be expressed as:

$$\text{MCRB}(\bar{\boldsymbol{\theta}}) = \frac{1}{M} \mathbf{D}_N^{\dagger} \left[\upsilon \bar{\boldsymbol{\Sigma}} \otimes \bar{\boldsymbol{\Sigma}} + (\upsilon - 1) \text{vec}(\bar{\boldsymbol{\Sigma}}) \text{vec}(\bar{\boldsymbol{\Sigma}})^T \right] \left(\mathbf{D}_N^{\dagger}\right)^T , \tag{72}$$

where

$$\upsilon \square \frac{N\Gamma(N/\beta)\Gamma((N+2)/\beta)}{(N+1)\Gamma((N+1)/\beta)^2} . \tag{73}$$

The FIM for GG-distributed data has been evaluated for a single observation vector $\mathbf{x}_m$ in [35]. After some matrix manipulations, the CRB on $\bar{\boldsymbol{\Sigma}}$ is derived to be:

$$\text{CRB}(\bar{\boldsymbol{\theta}}) = \frac{1}{M} \mathbf{D}_N^{\dagger} \left[\frac{N+1}{N+\beta} \bar{\boldsymbol{\Sigma}} \otimes \bar{\boldsymbol{\Sigma}} + \frac{1-\beta}{\beta(N+\beta)} \text{vec}(\bar{\boldsymbol{\Sigma}}) \text{vec}(\bar{\boldsymbol{\Sigma}})^T \right] \left(\mathbf{D}_N^{\dagger}\right)^T . \tag{74}$$

As for the Case Study 1, in Fig. 6 we compare the MSE of the MML estimator and of the Tyler's estimator with the MCRB the CRB in terms of the indices ε_{MML}, ε_{Tyler}, ε_{MCRB} and ε_{CRB}, as a function of the shape parameter β. The value of the one-lag correlation coefficient is $\rho = 0.8$, the number of data vectors is $M=3N$. To calculate the MSE of the estimators we run 10^5 Monte Carlo trials. As expected, for $\beta=1$ the MCRB and the CRB are equal, since the Generalized Gaussian pdf becomes a complex Gaussian pdf, then matched and mismatched cases coincide. In heavy tail disturbance ($\beta<1$), thanks to its robustness, Tyler's estimator has better performance than the MML estimator. The reverse is true when $\beta>1$.

In Fig. 7 the indices ε_{MML}, ε_{Tyler}, ε_{MCRB} and ε_{CRB} are compared as function of the number of available data M. Finally, in Fig. 8 the performance of the MML and of the Tyler's estimator are plotted as function of ρ for $\beta = 0.2$ and N=8. Even in this case, Tyler's estimator has better performance than the MML estimator and is very close to the CRB.

Figure 6 – MSE indices, ε_{MML} and ε_{Tyler}, and bounds ε_{MCRB} and ε_{CRB} as function of the shape parameter of the *GG* distribution (ρ=0.8, *N*=8, *M*=3*N*).

Figure 7 – MSE indices, ε_{MML} and ε_{Tyler}, and bounds ε_{MCRB} and ε_{CRB} as function of the available data (ρ=0.8, *N*=8, β=0.2).

Figure 8 – MSE indices, ε_{MML} and ε_{Tyler}, and bounds ε_{MCRB} and ε_{CRB} as function of the one-lag correlation coefficient ρ (β=0.2, *N*=8, *K*=3*N*).

Case Study 3 Assumed pdf: Generalized Gaussian; true pdf: *t*-student.

In this study case, we want to investigate the scenario where we know that the data are not Gaussian, but we do not assume the correct "non-Gaussian" model for our data. As in the Case Study 1, we assume that the true distribution is a complex-*t* distribution, but unlike the previous case, we assume a complex GG distribution for modeling our data. The MML estimator, then, is the ML estimator for the GG data. Unlike the SCM (i.e. the ML estimator for Normal data), the ML estimator for GG data cannot be expressed with an explicit equation but has to be defined through a

43

fixed-point equation. In this subsection, 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 pdf given in eq. (51), while the assumed pdf is the GG distribution in eq. (68) recalled here for sake of clarity:

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

where β is the shape parameter and b is the scale parameter that are again assumed to be known. In this case, the MML estimator is the solution of the following fixed-point matrix equation [16], [35], [51]:

$$\hat{\boldsymbol{\Sigma}}_{MML}=\frac{1}{M}\sum_{m=1}^M\varphi\left(\mathbf{x}_m^H\hat{\boldsymbol{\Sigma}}_{MML}^{-1}\mathbf{x}_m\right)\mathbf{x}_m\mathbf{x}_m^H=H_M\left(\hat{\boldsymbol{\Sigma}}_{MML}\right),\qquad(75)$$

where the function φ is given by $\varphi\left(t\right)=\left(\beta/b\right)t^{\beta-1}$. Following Theorem 6 in [16], it can be shown that, for every (symmetric and positive-definite) starting matrix $\boldsymbol{\Sigma}^{(0)}$, the recursive version of eq. (75) converges to $\hat{\boldsymbol{\Sigma}}_{MML}$, i.e. $\hat{\boldsymbol{\Sigma}}^{k+1}=H_M\left(\hat{\boldsymbol{\Sigma}}^k\right)\underset{k\to\infty}{\to}\hat{\boldsymbol{\Sigma}}_{MML}$ iff $\beta\in(0,1)$. For $\beta>1$, i.e. when the tails of the GG distribution are lighter than those of the Normal distribution, the recursive estimator of the scatter matrix is no longer reliable. In fact, when $\beta>1$, the conditions on $\varphi(t)$ that guarantee the existence and the uniqueness of the estimator are not satisfied. Theorem 5 in [46] can be used to prove that for $\beta\in(0,1)$ we have $\hat{\boldsymbol{\Sigma}}_{MML}\underset{M\to\infty}{\overset{a.s.}{\to}}\boldsymbol{\Sigma}_0$, i.e. the MML estimator $\hat{\boldsymbol{\Sigma}}_{MML}$ converges with probability 1 to $\boldsymbol{\Sigma}_0$. From the theory previously discussed, the limiting value $\boldsymbol{\Sigma}_0$ must be the matrix that minimizes the KL divergence between $p_X\left(\mathbf{x}_m\right)$ and $f\left(\mathbf{x}_m;\boldsymbol{\Sigma}\right)$. In order to calculate $\boldsymbol{\Sigma}_0$,

44

we can apply eq. (53), where in this case the density generator is that of the GG distribution, i.e. $g(t) = \exp(-t^{\beta}/b)$. After some calculations, we get:

$$\partial D\left(p_X \| f_{\Sigma}\right) = \text{tr}\left(\Sigma^{-1}\partial\Sigma\right) - \frac{\beta}{b}\text{tr}\left(E_p\left\{\left(\mathbf{x}_m^H\Sigma^{-1}\mathbf{x}_m\right)^{\beta-1}\Sigma^{-1}\mathbf{x}_m\mathbf{x}_m^H\Sigma^{-1}\right\}\partial\Sigma\right). \tag{76}$$

By applying the Stochastic Representation Theorem, we have $\mathbf{x}_m =_d \sqrt{Q_{\bar{\Sigma}}}\mathbf{T}\mathbf{u}$, where $Q_{\bar{\Sigma}} \,\square\, \mathbf{z}^H\bar{\Sigma}^{-1}\mathbf{z}$, $\bar{\Sigma} = \mathbf{T}\mathbf{T}^H$ is a factorization of the shape $\bar{\Sigma}$, $\mathbf{u}$ 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 $E\{\mathbf{u}\mathbf{u}^H\} = N^{-1}\mathbf{I}$. Then, eq. (76) can be rewritten as:

$$\partial D\left(p_X \| f_{\Sigma}\right) = \text{tr}\left(\Sigma^{-1}\partial\Sigma\right) - \frac{\beta E\{Q_{\bar{\Sigma}}^{\beta}\}}{b}\text{tr}\left(E_p\left\{\left(\mathbf{u}^H\mathbf{T}^H\Sigma^{-1}\mathbf{T}\mathbf{u}\right)^{\beta-1}\Sigma^{-1}\mathbf{T}\mathbf{u}\mathbf{u}^H\mathbf{T}^H\Sigma^{-1}\right\}\partial\Sigma\right), \tag{77}$$

where $E\{Q_{\bar{\Sigma}}^{\beta}\}$ can be evaluated explicitly by using the integral in [52, p. 315, n. 194.3]:

$$E\{Q_{\bar{\Sigma}}^{\beta}\} = \left(\frac{\lambda}{\eta}\right)^{\beta}\frac{\Gamma(\beta+N)\Gamma(\lambda-\beta)}{\Gamma(N)\Gamma(\lambda)}. \tag{78}$$

From eq. (77), setting to zero the derivative of the KL divergence w.r.t. Σ leads to:

$$\Sigma^{-1} - \beta b^{-1}E\{Q_{\bar{\Sigma}}^{\beta}\}E\left\{\left(\mathbf{u}^H\mathbf{T}^H\Sigma^{-1}\mathbf{T}\mathbf{u}\right)^{\beta-1}\Sigma^{-1}\mathbf{T}\mathbf{u}\mathbf{u}^H\mathbf{T}^H\Sigma^{-1}\right\}\Big|_{\Sigma=\Sigma_0} = \mathbf{0}. \tag{79}$$

Through some standard matrix manipulation, we get:

$$\mathbf{T}^{-1}\Sigma_0\mathbf{T}^{-H} = \beta b^{-1}E\{Q_{\bar{\Sigma}}^{\beta}\}E\left\{\left(\mathbf{u}^H\mathbf{T}^H\Sigma_0^{-1}\mathbf{T}\mathbf{u}\right)^{\beta-1}\mathbf{u}\mathbf{u}^H\right\}. \tag{80}$$

Now, assuming that the solution of eq. (80) is a scaled version of the true shape matrix, i.e. $\Sigma_0 = \delta\bar{\Sigma}$, we have $\delta\mathbf{I} = \frac{\beta}{bN}E\{Q_{\bar{\Sigma}}^{\beta}\}\delta^{-\beta+1}\mathbf{I}$, so that

45

$$\delta = \frac{\lambda}{\eta} \left(\frac{\beta}{bN} \frac{\Gamma(\beta+N)\Gamma(\lambda-\beta)}{\Gamma(N)\Gamma(\lambda)} \right)^{1/\beta}. \tag{81}$$

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

$$\boldsymbol{\Sigma}_0 = \frac{\lambda}{\eta} \left(\frac{\beta}{bN} \frac{\Gamma(\beta+N)\Gamma(\lambda-\beta)}{\Gamma(N)\Gamma(\lambda)} \right)^{1/\beta} \bar{\boldsymbol{\Sigma}} \,\square\, \delta\bar{\boldsymbol{\Sigma}}. \tag{82}$$

Since $\boldsymbol{\Sigma}_0$ is a scaled version of the true scatter matrix, the MML estimator is not consistent in general. As shown in [16] and [46], for the estimator in eq. (75), the following asymptotic relation holds:

$$E_p \left\{ \varphi\left(\mathbf{x}_m^H \boldsymbol{\Sigma}_{f,MML}^{-1} \mathbf{x}_m \right) \mathbf{x}_m \mathbf{x}_m^H \right\} = \gamma \bar{\boldsymbol{\Sigma}}. \tag{83}$$

Eq. (83) 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 p_X is:

$$\boldsymbol{\mu} = E_p \left\{ \hat{\boldsymbol{\Sigma}}_{MML}(\mathbf{x}) \right\} \underset{M\to\infty}{=} \gamma\bar{\boldsymbol{\Sigma}}, \tag{84}$$

where the scalar term γ can be evaluated by solving the following integral equation [16]:

$$E\left\{ \varphi\left(\frac{Q_\Sigma}{\gamma} \right) \frac{Q_\Sigma}{\gamma} \right\} = N. \tag{85}$$

Given $\varphi(t) = \beta t^{\beta-1}/b$ and by using the integral in eq. (78), γ can be evaluated as:

$$\gamma = \frac{\lambda}{\eta} \left(\frac{\beta}{bN} \cdot \frac{\Gamma(\beta+N)\Gamma(\lambda-\beta)}{\Gamma(N)\Gamma(\lambda)} \right)^{1/\beta} = \delta. \tag{86}$$

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 and then the vector $\mathbf{r}$ in eq. (18) is equal to zero. In particular, 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:

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

Now, we can compare the estimation performance of the MML estimator directly with the MCRB in eq. (18). As before, the MCRB can be evaluated using the compact expression for $\mathbf{A}_{\bar{\boldsymbol{\theta}}}^{-1}\mathbf{B}_{\bar{\boldsymbol{\theta}}}\mathbf{A}_{\bar{\boldsymbol{\theta}}}^{-1}$ derived in Appendix C, eq. (C.10). The density generator for the GG distribution is $g(t) = \exp(-t^\beta/b)$, hence we have:

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

In order to evaluate the terms B_1, B_2, A_1 and A_2 in eqs. (C.2), (C.3), (C.6), and (C.7), respectively, the integral in eq. (78) is needed. In particular, we have:

$$E\left\{Q \frac{\partial \ln g(Q)}{\partial Q}\right\} = -\frac{\beta}{b}\left(\frac{\lambda}{\eta}\right)^\beta \frac{\Gamma(\beta + N)\Gamma(\lambda - \beta)}{\Gamma(N)\Gamma(\lambda)}, \tag{87}$$

$$E\left\{Q^2 \left(\frac{\partial \ln g(Q)}{\partial Q}\right)^2\right\} = \frac{\beta^2}{b^2}\left(\frac{\lambda}{\eta}\right)^{2\beta} \frac{\Gamma(2\beta + N)\Gamma(\lambda - 2\beta)}{\Gamma(N)\Gamma(\lambda)}, \tag{88}$$

$$E\left\{Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2}\right\} = -\frac{\beta(\beta-1)}{b}\left(\frac{\lambda}{\eta}\right)^\beta \frac{\Gamma(\beta + N)\Gamma(\lambda - \beta)}{\Gamma(N)\Gamma(\lambda)}, \tag{89}$$

with $\beta < \lambda/2$. Finally, the MCRB is evaluated using eq. (C.10), while the CRB is given in eq. (63).

In the following, we compare the MSE of the MML estimator and of Tyler's estimator given in eq. (64) with the MCRB and the CRB by calculating the indices ε_{MML}, ε_{Tyler}, ε_{MCRB}, and ε_{CRB} previously defined. Both the iterations to derive the MML and Tyler's estimators are initialized using the identity matrix **I**. As before, the value of the one-lag correlation coefficient is $\rho = 0.8$ and the number of data vectors is *M*=3*N*. To calculate the MSE of the estimators, we run 10^5 Monte Carlo trials.

The simulation results concern two different scenarios: (1) the super-Gaussian scenario, i.e. the true *t*-distribution has heavier tails than a Normal distribution and where *λ*=3, and (2) the quasi-Gaussian scenario, where *λ*=50 (*λ* is the shape parameter of the *t*-distribution).

As shown in Fig. 9, the MML estimator achieves better performance than Tyler's estimator when *β*<0.6, i.e. when the assumed GG distribution has extremely heavy tails. The MCRB gets close to the CRB when *β* decreases, i.e. the assumed pdf becomes spikier. As expected, in the quasi-Gaussian case of Fig. 10, 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 becomes better as *β* tends to 1, i.e. the MML estimator tends to the SCM. Also, the MCRB gets closer and closer to the CRB when *β* tends to 1. Clearly, the MSE of Tyler's estimator and the CRB do not depend on the value of the shape parameter *β* of the assumed pdf.

Figure 9 – MSE indices, ε_{MML} and ε_{Tyler}, and bounds ε_{MCRB} and ε_{CRB} as a function of the shape parameter of the *GG* distribution (ρ=0.8, *N*=8, *M*=3*N*, *λ*=3).

6.2 Misspecified joint estimation of the scatter matrix and of the extra-parameters

In the previous section, we showed how to apply the mismatched estimation framework to the problem of estimating the scatter matrix of a CES distributed random vector when all the extra parameters are *a-priori* known. Now, we investigate the more realistic scenario where all these parameters are unknown and should be jointly estimated. In this case, the constraint on the trace of the scatter matrix must be imposed in order to make the joint estimation problem well-defined [53].

We investigate the same scenario as in Case Study 1 of Section 6.2: the true data pdf is a complex *t*-distribution, while the joint MML estimator of the scatter matrix and of the data power is derived under a Normal model assumption. This is a recurring scenario in adaptive radar applications. In fact, the choice of the *t*-distribution as true data model has been motivated by experimental evidence (see e.g. [36] and [37]) that proved its reliability to model spiky clutter data. On the other hand, many radar systems exploit the Normal model for data inference due to its analytical tractability and the consequent real time implementation of the estimation algorithms based on it.

More formally, we assume that the *M* vectors of the available dataset $\mathbf{x} = \{\mathbf{x}_m\}_{m=1}^{M}$ are iid and each one has a complex Normal multivariate pdf given in eq. (50):

$$f_X\left(\mathbf{x}_m;\boldsymbol{\theta}\right) \square \ f_X\left(\mathbf{x}_m;\boldsymbol{\Sigma},\sigma^2\right) = \frac{1}{\left(\pi\sigma^2\right)^N |\boldsymbol{\Sigma}|}\exp\left(-\frac{\mathbf{x}_m^H\boldsymbol{\Sigma}^{-1}\mathbf{x}_m}{\sigma^2}\right), \quad \mathrm{tr}\left(\boldsymbol{\Sigma}\right) = N. \tag{90}$$

The covariance matrix is $\mathbf{M} = E\{\mathbf{x}_m \mathbf{x}_m^H\} = \sigma^2 \mathbf{\Sigma}$, where $\mathrm{tr}(\mathbf{\Sigma}) = N$ and σ^2 is the power. Hence, the parameter vector to be estimated can be expressed as $\boldsymbol{\theta} = \begin{bmatrix} \mathrm{vecs}(\mathbf{\Sigma})^T & \sigma^2 \end{bmatrix}^T \in \Theta$. However, the true data are distributed according to the complex t-distribution $p_X(\mathbf{x}_m; \boldsymbol{\tau}) \square\ p_X(\mathbf{x}_m; \bar{\mathbf{\Sigma}}, \lambda, \eta)$ of eq. (51):

$$p_X\left(\mathbf{x}_m; \boldsymbol{\tau}\right) \square\ p_X(\mathbf{x}_m; \bar{\mathbf{\Sigma}}, \lambda, \eta) = \frac{1}{\pi^N |\bar{\mathbf{\Sigma}}|} \frac{\Gamma(N+\lambda)}{\Gamma(\lambda)} \left(\frac{\lambda}{\eta}\right)^\lambda \left(\frac{\lambda}{\eta} + \mathbf{x}_m^H \bar{\mathbf{\Sigma}}^{-1} \mathbf{x}_m\right)^{-(N+\lambda)}, \ \mathrm{tr}\left(\bar{\mathbf{\Sigma}}\right) = N, \quad (91)$$

where $\boldsymbol{\tau} = \begin{bmatrix} \mathrm{vecs}(\bar{\mathbf{\Sigma}})^T & \lambda & \eta \end{bmatrix}^T \in \mathrm{T}$ is the true parameter vector and $\bar{\mathbf{\Sigma}}$ is the true scatter matrix that could be different to the scatter matrix $\mathbf{\Sigma}$ of the assumed Gaussian distribution.

It is worth observing that in the mismatched case the parameter space Θ that parameterizes the assumed distribution and the parameter space T that parameterizes the true distribution may be different. In the case at hand, for example $\mathrm{T} \subset \square^l \times (0, \infty) \times (0, \infty)$ while $\Theta \subset \square^l \times (0, \infty)$ where $\times$ indicates the Cartesian product and $l = N(N+1)/2$ as before. Moreover, the constraint on the trace of the scatter matrix limits both the true and assumed parameter vector to belong to two lower dimensional smooth manifolds $\bar{\mathrm{T}} = \{\boldsymbol{\tau} \in \mathrm{T} | \mathrm{tr}(\bar{\mathbf{\Sigma}}) = N\}$ and $\bar{\Theta} = \{\boldsymbol{\theta} \in \Theta | \mathrm{tr}(\mathbf{\Sigma}) = N\}$, respectively.

The main aspects which differentiate the mismatched estimation problem at hand form the one discussed in the Case Study 1 of Section 6.1 are:

- The assumed and the true parameter spaces, $\bar{\Theta}$ and $\bar{\mathrm{T}}$, are different. Consequently, the simplified approach discussed in Sect. 4.4 cannot be applied and a consistent MML estimator does not exist.

- No assumptions are made on the value of σ^2, i.e., the power of the assumed Normal distribution. It is considered an additional unknown parameter that needs to be jointly estimated with the scatter matrix $\boldsymbol{\Sigma}$.

- To guarantee the identifiability of σ^2 and $\boldsymbol{\Sigma}$, a constraint on $\boldsymbol{\Sigma}$, i.e., $\mathrm{tr}(\boldsymbol{\Sigma}) = N$, needs to be imposed. This means that we should compare the performance of a constrained MML estimator with the CMCRB derived in eq. (27).

In the following, the constrained MML estimator for the estimation of $\boldsymbol{\theta}$ is firstly derived and its convergence properties investigated. Then, the CMCRB for the joint estimation problem at hand is calculated.

6.2.a Derivation of the constrained MML (CMML) estimator

In order to obtain an estimation of $\boldsymbol{\theta}$, we apply the MML algorithm in eq. (13). In particular, under the assumption of complex Normal data model in eq. (90), the likelihood function can be expressed as:

$$L(\boldsymbol{\theta}) = \sum\nolimits_{m=1}^{M} \ln f_X\left(\mathbf{x}_m; \boldsymbol{\theta}\right) = -NM \ln \sigma^2 - M \ln |\boldsymbol{\Sigma}| - \sum\nolimits_{m=1}^{M} \mathbf{x}_m^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_m / \sigma^2 . \tag{92}$$

Then, the MML estimator can be obtained by maximizing $L(\boldsymbol{\theta})$ subject to the linear constraint $\mathrm{tr}(\boldsymbol{\Sigma}) = N$. To proceed, we do not rely on the Lagrange multiplier method, but rather we follow a different, yet equivalent (at least asymptotically), procedure [54]. We first derive the unconstrained MML estimator and then we project it on the lower dimensional manifold $\overline{\Theta}$ by imposing the constraint. Specifically, the MML estimator is the solution of the following problem:

$$\begin{cases} \dfrac{\partial L(\boldsymbol{\theta})}{\partial \sigma^2} = -\dfrac{NM}{\sigma^2} + \dfrac{1}{\sigma^4} \sum_{m=1}^{M} \mathbf{x}_m^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_m = 0 \\[2ex] \dfrac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}} = -M\boldsymbol{\Sigma}^{-1} + \dfrac{\boldsymbol{\Sigma}^{-1}}{\sigma^2} \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^H \boldsymbol{\Sigma}^{-1} = \mathbf{0} \\[2ex] \text{s.t.} \quad \mathrm{tr}(\boldsymbol{\Sigma}) = N \end{cases} \tag{93}$$

Then, we have:

$$\begin{cases} \hat{\sigma}_{MML}^2 = \dfrac{1}{NM} \sum_{m=1}^{M} \mathbf{x}_m^H \hat{\boldsymbol{\Sigma}}_{MML}^{-1} \mathbf{x}_m \\[2ex] \hat{\boldsymbol{\Sigma}}_{MML} = \dfrac{N}{\sum_{m=1}^{M} \mathbf{x}_m^H \hat{\boldsymbol{\Sigma}}_{MML}^{-1} \mathbf{x}_m} \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^H \\[2ex] \text{s.t.} \quad \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{MML}) = N \end{cases} \tag{94}$$

Hence, imposing the constraint we get the constrained MML (CMML) estimators of σ^2 and $\boldsymbol{\Sigma}$:

$$\begin{cases} \hat{\sigma}_{CMML}^2 = \dfrac{1}{NM} \sum_{m=1}^{M} \mathbf{x}_m^H \hat{\boldsymbol{\Sigma}}_{CMML}^{-1} \mathbf{x}_m \\[2ex] \hat{\boldsymbol{\Sigma}}_{CMML} = \dfrac{N}{\sum_{m=1}^{M} \mathbf{x}_m^H \mathbf{x}_m} \sum_{m=1}^{M} \mathbf{x}_m \mathbf{x}_m^H \end{cases} \tag{95}$$

Now we need to find the convergence point $\boldsymbol{\theta}_0$ of the CMML estimator (see eq. (14)). As discussed in Sect. 4.3, the convergence point $\boldsymbol{\theta}_0 = \begin{bmatrix} \mathrm{vecs}(\boldsymbol{\Sigma}_0)^T & \sigma_0^2 \end{bmatrix}^T$ is the vector that minimizes the KL divergence between the true t-distribution $p_X(\mathbf{x}_m; \boldsymbol{\tau})$ in eq. (91) and $f_X(\mathbf{x}_m; \boldsymbol{\theta})$ in eq. (90). To this end, we have to solve the following system:

$$\begin{cases} \dfrac{\partial D(p_X \| f_{\boldsymbol{\theta}})}{\partial \sigma^2} = -E_p \left\{ \dfrac{\partial \ln f_X(\mathbf{x}_m; \boldsymbol{\Sigma}, \sigma^2)}{\partial \sigma^2} \right\} = 0 \\[2ex] \dfrac{\partial D(p_X \| f_{\boldsymbol{\theta}})}{\partial \boldsymbol{\Sigma}} = -E_p \left\{ \dfrac{\partial \ln f_X(\mathbf{x}_m; \boldsymbol{\Sigma}, \sigma^2)}{\partial \boldsymbol{\Sigma}} \right\} = \mathbf{0} \end{cases} \tag{96}$$

The first equation immediately provides:

$$\frac{\partial D\left(p_X \| f_\theta\right)}{\partial \sigma^2} = E_p\left\{\frac{\partial}{\partial \sigma^2}\left(N\ln\sigma^2 + \frac{\mathbf{x}_m^H \boldsymbol{\Sigma}^{-1}\mathbf{x}_m}{\sigma^2}\right)\right\} = \frac{N}{\sigma^2} - \frac{E_p\{Q_\Sigma\}}{\sigma^4} = 0, \qquad (97)$$

where $Q_\Sigma \square \mathbf{x}_m^H \boldsymbol{\Sigma}^{-1}\mathbf{x}_m$. By solving eq. (97) we get: $\sigma_0^2 = \dfrac{E_p\{Q_\Sigma\}}{N}$.

The derivative of the KL divergence with respect to $\boldsymbol{\Sigma}$ is instead given by:

$$\frac{\partial D\left(p_X \| f_\theta\right)}{\partial \boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1} - \frac{E\{Q_\Sigma\}}{N\sigma^2}\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} = \mathbf{0},\ \mathrm{tr}\left(\boldsymbol{\Sigma}\right) = \mathrm{tr}\left(\bar{\boldsymbol{\Sigma}}\right) = N, \qquad (98)$$

whose solution is $\boldsymbol{\Sigma}_0 = \dfrac{E\{Q_{\Sigma_0}\}}{N\sigma^2}\bar{\boldsymbol{\Sigma}}$. Putting together the two solutions, we finally obtain:

$$\sigma_0^2 = \frac{E\{Q_{\Sigma_0}\}}{N}\ \text{ and }\ \boldsymbol{\Sigma}_0 = \frac{E\{Q_{\Sigma_0}\}}{N\sigma_0^2}\bar{\boldsymbol{\Sigma}} = \bar{\boldsymbol{\Sigma}},\ \text{where }\ \mathrm{tr}\left(\boldsymbol{\Sigma}_0\right) = \mathrm{tr}\left(\bar{\boldsymbol{\Sigma}}\right) = N, \qquad (99)$$

and

$$\sigma_0^2 = \frac{E\{Q_{\Sigma_0}\}}{N} = \frac{E\{Q_{\bar{\Sigma}}\}}{N} = \frac{E_p\{\mathbf{x}_m^H \bar{\boldsymbol{\Sigma}}^{-1}\mathbf{x}_m\}}{N} = \bar{\sigma}^2 = \frac{\lambda}{\eta\left(\lambda-1\right)}, \qquad (100)$$

where $\bar{\sigma}^2$ is the true statistical power of the data. Eqs. (99)-(100) show that the CMML estimator

converges *a.s.* to the parameter vector $\hat{\boldsymbol{\theta}}_{CMML}\left(\mathbf{x}\right) \underset{M\to\infty}{\overset{a.s.}{\to}} \boldsymbol{\theta}_0 = \left[\mathrm{vecs}(\bar{\boldsymbol{\Sigma}})^T\quad \bar{\sigma}^2\right]^T$, i.e.

$$\hat{\sigma}_{CMML}^2\left(\mathbf{x}\right) \underset{M\to\infty}{\overset{a.s.}{\to}} \bar{\sigma}^2 = \lambda/\eta\left(\lambda-1\right)\ \text{ and }\ \hat{\boldsymbol{\Sigma}}_{CMML}\left(\mathbf{x}\right) \underset{M\to\infty}{\overset{a.s.}{\to}} \bar{\boldsymbol{\Sigma}}. \qquad (101)$$

Hence, it provides "consistent" estimates for both the scatter matrix and the power of the true data

model [53]. From a practical point of view, this means that we can use the simple mismatched

estimator based on the Gaussian assumption to estimate the scatter matrix and the average power of the (complex t-distributed) data since it converges to the true required quantities. The analysis of the performance loss of the mismatched estimator in eq. (95) is reported in the next subsection (see [48] for more details).

6.2.b The CMCRB for the joint estimation of the scatter matrix and the power

In Section 6.1 (Case Study 1), the MCRB on the estimation of the scatter matrix has been evaluated for a complex-t distribution when the assumed misspecified distribution is a complex Normal pdf, under the assumption of a-priori known power. Here, we generalize the result for the case of unknown power, i.e., when the power σ^2 and the scatter matrix $\boldsymbol{\Sigma}$ are unknown and jointly estimated. In this case σ^2 and $\boldsymbol{\Sigma}$ are not identifiable unless a constraint on $\boldsymbol{\Sigma}$ is imposed, e.g., $\mathrm{tr}(\boldsymbol{\Sigma}) = N$. In order to incorporate this constraint in the MCRB, we calculate the constrained MCRB (CMCRB) of Theorem 3 in Sect 4.5. In the following, we specialize the general expression provided in eq. (27) for the case study at hand.

Evaluation of the matrix $\mathbf{A}_{\boldsymbol{\theta}_0}$. Matrix $\mathbf{A}_{\boldsymbol{\theta}_0}$ can be decomposed in the following blocks:

$$\mathbf{A}_{\boldsymbol{\theta}_0} = \mathbf{T}_1^T \begin{bmatrix} \mathbf{A}_{\bar{\boldsymbol{\Sigma}}} & \mathbf{A}_c \\ \mathbf{A}_c^T & A_{\bar{\sigma}^2} \end{bmatrix} \mathbf{T}_1, \tag{102}$$

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{D}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_i \end{bmatrix}, \tag{103}$$

where $\mathbf{I}_i$ is the identity matrix of dimension $i \times i$ and $\mathbf{D}_N$ is the so-called Duplication matrix of order N [38]. Following the procedure in [41], we have:

$$\mathbf{A}_{\bar{\boldsymbol{\Sigma}}} = -\bar{\boldsymbol{\Sigma}}^{-1} \otimes \bar{\boldsymbol{\Sigma}}^{-1}. \tag{104}$$

$$A_{\bar{\sigma}^2} = E_p \left\{ \frac{\partial^2}{\partial^2 \bar{\sigma}^2} \ln f_X \left(\mathbf{x}_m; \boldsymbol{\theta}_0 \right) \right\} = E_p \left\{ \frac{\partial^2}{\partial^2 \bar{\sigma}^2} \left(-N \ln \bar{\sigma}^2 - \frac{Q_{\bar{\Sigma}}}{\bar{\sigma}^2} \right) \right\}$$

$$= \frac{N}{\bar{\sigma}^4} - 2 \frac{E_p \{ Q_{\bar{\Sigma}} \}}{\bar{\sigma}^6} = -\frac{N}{\bar{\sigma}^4} , \tag{105}$$

$$[\mathbf{A}_c]_{i,1} = E_p \left\{ \frac{\partial^2 \ln f_X \left(\mathbf{x}_m; \boldsymbol{\theta}_0 \right)}{\partial \bar{\sigma}^2 \partial \mathrm{vec}(\bar{\boldsymbol{\Sigma}})_i} \right\} = -\frac{1}{\bar{\sigma}^4} E_p \left\{ \mathbf{x}_m^H \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{x}_m \right\}$$

$$= -\frac{1}{\bar{\sigma}^2} \mathrm{tr} \left(\bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \right) = -\frac{1}{\bar{\sigma}^2} \mathrm{vec} \left(\bar{\boldsymbol{\Sigma}}^{-1} \right)^T \mathrm{vec} \left(\mathbf{A}_i \right) \tag{106}$$

where $\mathbf{A}_i = \partial \boldsymbol{\Sigma} / \partial \theta_i$ is a symmetric 0-1 matrix. From (106) we get: $\mathbf{A}_c = -\frac{1}{\bar{\sigma}^2} \mathrm{vec} \left(\bar{\boldsymbol{\Sigma}}^{-1} \right)$.

Evaluation of the matrix $\mathbf{B}_{\boldsymbol{\theta}_0}$. Matrix $\mathbf{B}_{\boldsymbol{\theta}_0}$ can be decomposed in the following blocks:

$$\mathbf{B}_{\boldsymbol{\theta}_0} = \mathbf{T}_1^T \begin{bmatrix} \mathbf{B}_{\bar{\Sigma}} & \mathbf{B}_c \\ \mathbf{B}_c^T & B_{\bar{\sigma}^2} \end{bmatrix} \mathbf{T}_1 . \tag{107}$$

As before, following the procedure in [41], we get:

$$\mathbf{B}_{\bar{\Sigma}} = \frac{1}{\lambda - 2} \mathrm{vec} \left(\bar{\boldsymbol{\Sigma}}^{-1} \right) \mathrm{vec} \left(\bar{\boldsymbol{\Sigma}}^{-1} \right)^T + \frac{\lambda - 1}{\lambda - 2} \bar{\boldsymbol{\Sigma}}^{-1} \otimes \bar{\boldsymbol{\Sigma}}^{-1} . \tag{108}$$

$$B_{\bar{\sigma}^2} = E_p \left\{ \left[\frac{\partial \ln f_X \left(\mathbf{x}_m; \boldsymbol{\theta}_0 \right)}{\partial \bar{\sigma}^2} \right]^2 \right\} = E_p \left\{ \left[\frac{\partial}{\partial \bar{\sigma}^2} \left(-N \ln \bar{\sigma}^2 - \frac{Q_{\bar{\Sigma}}}{\bar{\sigma}^2} \right) \right]^2 \right\}$$

$$= \frac{N^2}{\bar{\sigma}^4} - 2 \frac{E_p \{ Q_{\bar{\Sigma}} \}}{\bar{\sigma}^6} + \frac{E_p \{ Q_{\bar{\Sigma}}^2 \}}{\bar{\sigma}^8} = \frac{N^2}{\bar{\sigma}^4} - 2 \frac{N^2}{\bar{\sigma}^4} + \frac{N(N+1)(\lambda - 1)}{\bar{\sigma}^4 (\lambda - 2)} \tag{109}$$

$$= \frac{N(N + \lambda - 1)}{\bar{\sigma}^4 (\lambda - 2)} .$$

$$
\begin{aligned}
[\mathbf{B}_c]_{i,1} &= E_p \left\{ \frac{\partial \ln f_X (\mathbf{x}_m; \boldsymbol{\theta}_0)}{\partial \bar{\sigma}^2} \cdot \frac{\partial \ln f_X (\mathbf{x}_m; \boldsymbol{\theta}_0)}{\partial \mathrm{vec}(\bar{\boldsymbol{\Sigma}})_i} \right\} \\
&= E_p \left\{ \left(\frac{E_p \{Q_{\bar{\Sigma}}\}}{\bar{\sigma}^4} - \frac{N}{\bar{\sigma}^2} \right) \left(\frac{1}{\bar{\sigma}^2} \mathbf{x}_m^H \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{x}_m - \mathrm{tr}\left(\bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \right) \right) \right\} \\
&= -\frac{N}{\bar{\sigma}^2} \mathrm{tr}\left(\bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \right) + \frac{1}{\bar{\sigma}^6} E_p \left\{ \mathbf{x}_m^H \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{x}_m \mathbf{x}_m^H \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \bar{\boldsymbol{\Sigma}}^{-1} \mathbf{x}_m \right\} \\
&= \left(-\frac{N}{\bar{\sigma}^2} + \frac{(N+1)(\lambda-1)}{\bar{\sigma}^2(\lambda-2)} \right) \mathrm{tr}\left(\bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \right) = \frac{N+\lambda-1}{\bar{\sigma}^2(\lambda-2)} \mathrm{tr}\left(\bar{\boldsymbol{\Sigma}}^{-1} \mathbf{A}_i \right).
\end{aligned}
\tag{110}
$$

Hence, we get $\mathbf{B}_c = \dfrac{N+\lambda-1}{\bar{\sigma}^2(\lambda-2)} \mathrm{vec}\left(\bar{\boldsymbol{\Sigma}}^{-1} \right)$.

Evaluation of the matrix $\mathbf{U}$

The continuously differentiable constraint $\mathrm{tr}(\boldsymbol{\Sigma}) = N$ can be rewritten as:

$$
f(\boldsymbol{\theta}) = \sum_{i \in I} \mathrm{vecs}(\boldsymbol{\Sigma})_i - N = 0,
\tag{111}
$$

where I is the set of the indices of the diagonal entries of $\boldsymbol{\Sigma}$ that can be explicitly described as:

$$
I = \left\{ i \mid i = 1 + N(j-1) - \frac{(j-1)(j-2)}{2}, j = 1, ..., N \right\}.
\tag{112}
$$

Following [29], we define the $(l+1)$-dimensional gradient vector as:

$$
\nabla f(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \left[\frac{\partial \sum_{i \in I} \mathrm{vecs}(\boldsymbol{\Sigma})_i}{\partial \mathrm{vecs}(\boldsymbol{\Sigma})^T} \quad 0 \right] = \left[\mathbf{1}_I^T \quad 0 \right],
\tag{113}
$$

where $\mathbf{1}_I$ is a l-dimensional column vector defined as:

$$
\left[\mathbf{1}_I \right]_i = \begin{cases} 1 & i \in I \\ 0 & \text{otherwise} \end{cases}.
\tag{114}
$$

The gradient $\nabla f(\boldsymbol{\theta})$ has clearly full row rank and hence there exists a matrix $\mathbf{U} \in \square^{(l+1)\times l}$ whose columns form an orthonormal basis for the null space of $\nabla f(\boldsymbol{\theta})$, that is $\nabla f(\boldsymbol{\theta})\mathbf{U} = \mathbf{0}$ where $\mathbf{U}^T\mathbf{U}=\mathbf{I}$. The matrix $\mathbf{U}$ can be obtained numerically by evaluating, e.g. using the Singular Value Decomposition (SVD), the l orthonormal eigenvectors associated to the zero eigenvalue of $\nabla f(\boldsymbol{\theta})$ in eq. (113).

6.2.c Performance analysis

We now compare the estimation performance of the CMML estimator of eq. (95) [55] with the CMCRB. For the sake of comparison, in the following figures, we also report the MSE of the constrained Tyler's estimator (C-Tyler) and the (matched) CCRB for the joint estimation of $\boldsymbol{\tau} = \begin{bmatrix} \text{vecs}(\bar{\boldsymbol{\Sigma}})^T & \lambda & \eta \end{bmatrix}^T$ derived in [56] and [57]. In particular, the C-Tyler estimator is the constrained version of the robust estimator discussed of eq. (64), that can be evaluated by using the following iterative approach proposed in [45]:

$$
\begin{cases}
\hat{\boldsymbol{\Sigma}}_T^{(0)} = \mathbf{I} \\
\mathbf{S}_T^{(k+1)} = \displaystyle\sum_{m=1}^{M} \frac{\mathbf{x}_m \mathbf{x}_m^H}{\mathbf{x}_m^H \left(\hat{\boldsymbol{\Sigma}}_T^{(k)} \right)^{-1} \mathbf{x}_m} \\
\hat{\boldsymbol{\Sigma}}_T^{(k+1)} = N\mathbf{S}_T^{(k+1)} \Big/ \text{tr}\left(\mathbf{S}_T^{(k+1)} \right)
\end{cases}
\tag{115}
$$

for $k=1,\ldots,K$, where K is the number of iterations. It can be noted that in (115) there is a normalization on the trace of $\hat{\boldsymbol{\Sigma}}_T^{(k)}$ at every step of the iterative procedure to impose the constraint on the trace. Asymptotic consistency and unbiasedness properties are discussed in [16] and [48]. It is worth noting that the performance of the C-Tyler estimator can be assessed by comparing its error covariance matrix on the estimation of $\boldsymbol{\Sigma}$ with the CCRB derived in [56] and [57].

57

As global MSE index for the scatter matrix estimators, we use the one defined in eq. (66) (e.g. ε_{CMML} and $\varepsilon_{C-Tyler}$), while as performance bounds, the following quantity are plotted:

$$\varepsilon_{\mathrm{CMCRB}} \square \left\| \mathrm{CMCRB}(\mathbf{\Sigma}) \right\|_F , \quad \varepsilon_{\mathrm{CCRB}} \square \left\| \mathrm{CCRB}(\mathbf{\Sigma}) \right\|_F . \tag{116}$$

The accuracy on the estimate of average power σ^2 in the mismatched case is measured through its MSE, which is compared with the CMCRB. To calculate the estimation accuracy, we run 10^5 Monte Carlo trials. The simulation results have been organized as follows:

1. Estimation accuracy as function of the number M of available data vectors (Figs. 11 and 12). Simulation parameters: ρ=0.8, N=16, λ=3, η=1, K=4.

2. Estimation accuracy as function of the shape parameter λ (Figs. 13 and 14). Simulation parameters: ρ=0.8, N=16, M=10N, η=1, K=4.

3. Estimation accuracy as function of the one-lag correlation coefficient ρ (Figs. 15 and 16). Simulation parameters: N=16, M=10N, λ=3, η=1, K=4.

Based on the numerical analysis, we observe that:

- Regarding the CMML estimator, it always achieves the CMCRB, both for the estimation of the scatter matrix and for the estimation of the average power. The CMML presents a small bias on the estimation of the scatter matrix. Hence, $\hat{\mathbf{\Sigma}}_{CMML}$ is not a *MS*-unbiased estimator (at least in the finite sample regime) [53]. For this reason, ε_{CMML} can be slightly below the CMCRB (we talk about *superefficiency* in this case). The loss in estimation accuracy due to the mismatch is particularly high for extremely heavy tailed data, i.e. when λ is small (see Fig. 13). When $\lambda \to 0$, the CMCRB rapidly increases while the CCRB is quite independent of λ. On the other hand, when $\lambda \to \infty$, the CMCRB and the CCRB tend to coincide, as expected.

58

- Regarding the scatter matrix estimation, the robust C-Tyler estimator is an "almost" efficient estimator, even if it is not the most efficient estimator for *t*-distributed data. The MSE index $\varepsilon_{C-Tyler}$ is close to the CCRB especially for small λ (see Figs. 11, 13, and 15). In particular, its performance is robust, i.e., it is not affected by the value of the shape parameter λ (see Fig. 13), even if it is not efficient for large λ.

Figure 11 – MSE indices, ε_{CMML} and $\varepsilon_{C-Tyler}$, and bounds ε_{CMCRB} and ε_{CCRB} as function of the number *M* of available data vectors (ρ=0.8, *N*=16, λ=3, η=1).

Figure 12 – MSE of the CMML estimator of σ^2 and CMCRB as function of the number *M* of available data vectors (ρ=0.8, *N*=16, λ=3, η=1).

Figure 13 – The MSE indices ε_{CMML} and $\varepsilon_{C-Tyler}$, and bounds ε_{CMCRB} and ε_{CCRB} as function of the shape parameter λ (ρ=0.8, *N*=16, *M*=10*N*, η=1).

Figure 14 – MSE of the CMML estimator of σ^2 and CMCRB as function of the shape parameter λ (ρ=0.8, *N*=16, *M*=10*N*, η=1).

Figure 15 – MSE indices ε_{CMML} and $\varepsilon_{C-Tyler}$, and bounds ε_{CMCRB} and ε_{CCRB} as function of the one-lag correlation coefficient ρ (*N*=16, *M*=10*N*, λ=3, η=1).

Figure 16 – MSE of the CMML estimator of σ^2 and CMCRB as function of the one-lag correlation coefficient ρ (*N*=16, *M*=10*N*, λ=3, η=1).

7 Hypothesis testing problem for target detection

The last section of this chapter focuses on the target detection problem. In particular, this section aims at comparing the detection performance of the *adaptive normalized matched filter* (ANMF) by exploiting the CMML and the C-Tyler estimators for the scatter matrix. The NMF has been derived

and analysed by many authors under different names (see e.g. [49], [50], [58], [59], [60], [61], [62], [63], [64], [65]) in its adaptive and non-adaptive (i.e. when the disturbance scatter matrix is assumed to be known) versions. One of the most remarkable property of the non-adaptive NMF is the fact that it is a distribution-free detector under CES distributed clutter, i.e. the pdf of the decision statistic is invariant w.r.t. the particular CES distribution followed by the clutter [16]. We now summarize briefly the classical radar detection problem.

The problem is to detect the possible presence of a complex signal vector **s** in the received data **z**=**s**+**c**, where **c** represents the additive unobserved complex disturbance (noise/clutter) random vector. The target signal **s** is modelled as **s**= α**p** where **p** (generally called target vector response or Doppler steering vector) is the transmitted known radar pulse vector and $\alpha = \gamma e^{j\phi} \in \Box$ is an unknown signal parameter accounting for both channel propagation effects and the target backscattering. α can be modelled as an unknown deterministic parameter or as a random variable depending on the application at hand. When modelled as a random quantity, α is assumed to be a circular Gaussian random variable $\alpha \Box CN\left(0, \sigma_\alpha^2\right)$ where the amplitude γ is Rayleigh distributed and the phase ϕ is uniformly distributed in $[0, 2\pi)$ and independent of γ. Regarding the complex noise vector **c**, it has been successfully modelled as a zero-mean CES distributed random vector with covariance matrix $\mathbf{M} = \sigma^2 \mathbf{\Sigma}$, where $\mathbf{\Sigma}$ and σ^2 represent the unknown scatter matrix and the unknown statistical noise power. In the following, **c** is modelled as a complex *t*-distributed random vector.

The target detection problem can be expressed as a composite binary hypothesis testing problem:

$$H_0 : |\alpha| = 0 \quad \text{v.s.} \quad H_1 : |\alpha| > 0, \tag{117}$$

or, more explicitly as:

$$\begin{cases} H_0 : \mathbf{z} = \mathbf{c}, & \mathbf{x}_m = \mathbf{c}_m, \quad m = 1, ..., M, \\ H_1 : \mathbf{z} = \alpha\mathbf{p} + \mathbf{c}, & \mathbf{x}_m = \mathbf{c}_m, \quad m = 1, ..., M, \end{cases} \tag{118}$$

where the secondary data $\{\mathbf{x}_m\}_{m=1}^M$ are assumed to be iid and can be used to estimate the scatter matrix and share the same distribution as the clutter $\mathbf{c}$ in the primary data vector under test.

7.1 The ANMF detector

The Normalized Matched Filter (NMF) has been proposed, e.g. in [49], [50], [58], [59], [60], [61], and can be expressed as:

$$\Lambda_{NMF} \equiv \Lambda_{NMF}(\mathbf{z}, \mathbf{\Sigma}) = \frac{\left|\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}\right|^2}{\left(\mathbf{p}^H \mathbf{\Sigma}^{-1}\mathbf{p}\right)\left(\mathbf{z}^H \mathbf{\Sigma}^{-1}\mathbf{z}\right)}, \tag{119}$$

where the scatter matrix $\mathbf{\Sigma}$ is assumed to be perfectly known.

An important feature of the detector in (119) is the invariance under scalar multiplies of $\mathbf{x}$. In particular, the distribution of the test statistic Λ_{NMF} under the hypothesis H_0 is independent of the unknown average noise power σ^2 or the functional form of the particular CES distribution of the noise, i.e. the NMF is a distribution-free detector under H_0. The proof of this property can be found in [16]. Moreover, it can be shown that $\Lambda_{NMF}|H_0$ follows a Beta distribution:

$$\Lambda_{NMF}|H_0 \square \operatorname{Beta}(1, N-1), \tag{120}$$

where $\operatorname{Beta}(x; \alpha, \beta) = \left(x^{\alpha-1}(1-x)^{\beta-1}\right)\big/\mathrm{B}(\alpha, \beta)$, N is the dimension of the data vector and $\mathrm{B}(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

It is clear that the NMF cannot be used in practical applications where the scatter matrix $\mathbf{\Sigma}$ of the data vectors is generally unknown. So, in the following simulations, we aim at investigating the

detection performance of the adaptive NMF (ANMF) when $\mathbf{\Sigma}$ is replaced by 1) $\hat{\mathbf{\Sigma}}_{CMML}$, i.e. the CMML estimator derived in eq. (95) or 2) $\hat{\mathbf{\Sigma}}_T$, i.e., its min-max (over the CES distributions) robust C-Tyler estimator. Hence, we get the two adaptive estimators:

$$\Lambda_{ANMF-CMML} \equiv \Lambda_{ANMF-CMML}(\mathbf{z}, \hat{\mathbf{\Sigma}}_{CMML}) = \frac{\left|\mathbf{p}^H \hat{\mathbf{\Sigma}}_{CMML}^{-1} \mathbf{z}\right|^2}{\left(\mathbf{p}^H \hat{\mathbf{\Sigma}}_{CMML}^{-1} \mathbf{p}\right)\left(\mathbf{z}^H \hat{\mathbf{\Sigma}}_{CMML}^{-1} \mathbf{z}\right)} \tag{121}$$

$$\Lambda_{ANMF-C-Tyler} \equiv \Lambda_{ANMF-C-Tyler}(\mathbf{z}, \hat{\mathbf{\Sigma}}_T) = \frac{\left|\mathbf{p}^H \hat{\mathbf{\Sigma}}_T^{-1} \mathbf{z}\right|^2}{\left(\mathbf{p}^H \hat{\mathbf{\Sigma}}_T^{-1} \mathbf{p}\right)\left(\mathbf{z}^H \hat{\mathbf{\Sigma}}_T^{-1} \mathbf{z}\right)}. \tag{122}$$

As a consequence of the consistency of both the CMML and Tyler's estimators, the resulting adaptive test statistic Λ_{ANMF} will have asymptotically (i.e. for large M) a Beta(1,N-1) distribution. Hence, for large M, Λ_{ANMF} is (approximately) CFAR w.r.t. $\mathbf{\Sigma}$, as desired [16]. Further discussions on the asymptotic properties of the Λ_{ANMF} can be found in [66] and [67].

In the following, the performance of the two Λ_{ANMF} detectors in (121) and (122) are compared with that of the clairvoyant *linear threshold detector* (LTD) [37], i.e., the GLRT (with respect to the unknown complex signal amplitude α) for *t*-distributed data under the assumption of known scatter matrix and known shape and scale parameters. In particular, the clairvoyant LTD has been derived in [37]:

$$\Lambda_{LTD} \equiv \Lambda_{LTD}(\mathbf{z}, \mathbf{\Sigma}, \lambda, \eta) = \frac{\left|\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{z}\right|^2}{\left(\mathbf{p}^H \mathbf{\Sigma}^{-1} \mathbf{p}\right)\left(\mathbf{z}^H \mathbf{\Sigma}^{-1} \mathbf{z} + \lambda/\eta\right)}. \tag{123}$$

The detection performance of the Λ_{LTD} provides a useful upper bound to the performance of any adaptive detection algorithms, and in particular to $\Lambda_{ANMF-CMML}$ and $\Lambda_{ANMF-C-Tyler}$. In our simulation, the detectors are compared in terms of *i*) Constant False alarm Rate (CFAR) property w.r.t. the

scatter matrix and the extra parameters, *ii*) Probability of Detection (P_D) as function of the Signal-to-Disturbance power Ratio (SDR), and *iii*) Receiver Operating Characteristic (ROC) curves.

7.2 Detection performance

The detection performance of the NMF detector, which exploits either the CMML estimator or the C-Tyler estimator, are investigated by deriving by monte Carlo simulation the following curves:

1. The probability of false alarm (P_{FA}) as function of the one-lag coefficient ρ (Fig. 17). This verifies the CFAR property of the $\Lambda_{ANMF-CMML}$ (eq. (121)) and the $\Lambda_{ANMF-C-Tyler}$ (eq. (122)) w.r.t. the correlation shape. Simulation parameters: N=16, M=3N, λ=3, η=1, K=4. The detection thresholds have been set to achieve a nominal P_{FA} of 10^{-3}. The number of Monte Carlo runs is 10^6.

2. The probability of false alarm (P_{FA}) as function of the shape parameter λ of the true complex *t*-distribution, i.e., for different spikiness levels (Fig. 18). This allows us to investigate the CFAR property of the two ANMF's w.r.t. the spikiness of the data. Simulation parameters: N=16, M=3N, ρ=0.8, η=1, K=4. The detection thresholds have been set to achieve a nominal P_{FA} of 10^{-3}. The number of Monte Carlo runs is 10^6.

3. The probability of detection (P_D) as function of the SDR (Fig. 19). We also plot the performance of the clairvoyant LTD Λ_{LTD} in eq. (123), that represent an upper bound to the performance achievable by $\Lambda_{ANMF-CMML}$ and $\Lambda_{ANMF-C-Tyler}$. Simulation parameters: N=16, M=3N, ρ=0.8, λ=1, η=1, K=4. The detection thresholds have been set to achieve a nominal P_{FA} of 10^{-3}. Moreover, $\alpha \sim CN(0, \sigma_\alpha^2)$ where σ_α^2 varies according to the desired value of the SDR. The number of Monte Carlo runs is 10^6.

4. The Receiver Operating Characteristic (ROC) curves (Fig. 20). The simulation parameters are: N=16, M=3N, ρ=0.8, λ=3, η=1, K=4. As before, $\alpha \sim CN(0, \sigma_\alpha^2)$ where σ_α^2 is set to have

an SDR equal to 3 dB. The number of Monte Carlo runs is 10^6. Also in this case, as upper bound, the performance of Λ_{LTD} is reported.

As we can see from Fig. 17, both the ANMF detectors are (approximately) CFAR with respect to the disturbance one-lag correlation coefficient ρ. Their P_{FA} curves are almost constant and very close to the nominal P_{FA} value of 10^{-3}. A similar behavior can be observed in Fig. 18, where the P_{FA} curves have been evaluated as function of λ. It can be noted that both $\Lambda_{ANMF-CMML}$ and $\Lambda_{ANMF-C-Tyler}$ are CFAR detector w.r.t. the data spikiness, except for very low values λ, where the CMML estimator has large estimation losses (see Fig. 13) and the P_{FA} of $\Lambda_{ANMF-CMML}$ rapidly increases. Finally, in Figs. 19 and 20, the P_D vs the SDR and the Receiver Operating Characteristic (ROC) curves of $\Lambda_{ANMF-CMML}$ and $\Lambda_{ANMF-C-Tyler}$ are shown. For the sake of comparison, we also plot the detection performance of the clairvoyant GLRT for *t*-distributed data, i.e. the Λ_{LTD} of eq. (123) where Σ, λ and η are assumed to be *a-priori* known. The performance of $\Lambda_{ANMF-CMML}$ and $\Lambda_{ANMF-C-Tyler}$ are pretty close to that of the clairvoyant detector Λ_{LTD}. However, the adaptation losses increase when P_{FA} gets lower.

Figure 17 – Probability of false alarm vs disturbance one-lag correlation coefficient ρ.

Figure 18 – Probability of false alarm vs λ.

Figure 19 – Probability of detection vs *SDR* for $P_{FA}=10^{-3}$.

Figure 20 – Receiver Operating Characteristic (ROC) curves.

8. Conclusions

In practical applications, a certain amount of mismatch between the true and the assumed statistical data model is inevitable. Several authors in the statistical literature have shown how the classical tools of the estimation theory can be generalized to a mismatched scenario. In the first part of this chapter, a comprehensive review of the main contributions to the mismatched maximum likelihood theory have been proposed and discussed. A CRB under mismatched condition, i.e., the MCRB, was described and the behavior of the mismatched maximum likelihood (MML) estimator was investigated. In particular, we showed that the MML estimator is asymptotically MS-unbiased and its error covariance matrix asymptotically equates the MCRB. Moreover, a constrained version of MCRB is also described. In the second part of the chapter, we showed how to apply these results to a well-known problem in radar signal processing, i.e., the problem of estimating the disturbance covariance matrix for adaptive radar target detection. We addressed this problem by putting it in the more general context of the scatter matrix estimation of Complex Elliptically Symmetric (CES) distributed random vectors under data mismodeling. Two relevant scenarios have been considered. In the first one, the extra-parameters of the particular CES distributions at hand are assumed *a-priori* known. This allowed us to investigate the performance losses in the scatter matrix estimation due to a wrong specification of the functional form of the density generator. In the second scenario, we investigated the more realistic case where all the parameters are unknown and should be jointly estimated. We finished the chapter with an analysis of an adaptive detection algorithm, i.e., the ANMF which exploits either the MML estimator or the robust Tyler's estimator of the disturbance scatter matrix. The respective detection performance was compared with that of the clairvoyant

GLRT detector that relies on the correct model assumption and knows *a-priori* the disturbance parameters.

The mismatched approach to signal processing problem is a relatively new research field and, even though it promises huge opportunities for applicability to a plethora of different signal processing areas, many aspects still remain unresolved. In [10] for example, a generalization of the misspecified approach to the Bhattacharyya bound, to the Barankin bound and to the Bobrovsky-Mayer-Wolf-Zakai bound has been proposed, but much work remains to be done. More importantly, in [12], Richmond posed the bases to apply the mismatched approach to the derivation of Bayesian bounds, but the path to reach a complete misspecified Bayesian estimation theory is still long.

Appendix A

A generalization of the Slepian formula under misspecification

In this appendix, we report the misspecified version of the Slepian formula [68] proposed by Richmond and Horowitz in their seminal paper [10]. For the sake of clarity, in the following we use the same notation as in [10]. In particular, $(\cdot)^H$ and $(\cdot)^*$ denote the Hermitian and the complex conjugate operators.

Let the complex data vector $\mathbf{x} \in \square^N$ have a true complex Gaussian distribution such that $\mathbf{x} \square \ p_X(\mathbf{x}) = CN\ (\mathbf{d}, \mathbf{B})$ where $\mathbf{d}$ and $\mathbf{B}$ denote the (possibly complex) true mean value vector and the true covariance matrix. Let the assumed distribution for the data vector be another complex Gaussian distribution such that $\mathbf{x} \square \ f_X(\mathbf{x}; \boldsymbol{\theta}) = CN\ (\mathbf{r}(\boldsymbol{\theta}), \mathbf{R})$ where the (possibly complex) assumed mean value $\mathbf{r}(\boldsymbol{\theta})$ is parameterized by a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \square^d$ and $\mathbf{R}$ denotes the (possibly

complex) assumed covariance matrix generally different from **B**. For the generalization to complex parameter vector we refer the reader to [10]. It can be noted that the assumed mean value $\mathbf{r}(\boldsymbol{\theta})$ may be different from the true one, d, for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Under these assumptions, the matrices $\mathbf{A}_{\boldsymbol{\theta}}$ and $\mathbf{B}_{\boldsymbol{\theta}}$ in eqs. (1) and (4) can be written explicitly through the so-called misspecified Slepian formulae [10]. In particular, we have:

$$
\begin{aligned}
\left[\mathbf{A}_{\boldsymbol{\theta}}\right]_{i,k} = & -\frac{\partial \mathbf{r}^{H}(\boldsymbol{\theta})}{\partial \theta_{i}} \mathbf{R}^{-1} \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_{k}} - \frac{\partial \mathbf{r}^{H}(\boldsymbol{\theta})}{\partial \theta_{k}} \mathbf{R}^{-1} \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_{i}} + \\
& + \frac{\partial^{2} \mathbf{r}^{H}(\boldsymbol{\theta})}{\partial \theta_{i} \partial \theta_{k}} \mathbf{R}^{-1} \left(\mathbf{d} - \mathbf{r}(\boldsymbol{\theta})\right) + \left(\mathbf{d} - \mathbf{r}(\boldsymbol{\theta})\right)^{H} \mathbf{R}^{-1} \frac{\partial^{2} \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_{i} \partial \theta_{k}},
\end{aligned} \tag{A.1}
$$

$$
\left[\mathbf{B}_{\boldsymbol{\theta}}\right]_{i,k} = \frac{\partial \mathbf{r}^{H}(\boldsymbol{\theta})}{\partial \theta_{i}} \mathbf{R}^{-1} \mathbf{B} \mathbf{R}^{-1} \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_{k}} + \frac{\partial \mathbf{r}^{T}(\boldsymbol{\theta})}{\partial \theta_{i}} \left(\mathbf{R}^{-1}\right)^{*} \mathbf{B}^{*} \left(\mathbf{R}^{-1}\right)^{*} \frac{\partial \mathbf{r}^{*}(\boldsymbol{\theta})}{\partial \theta_{k}}. \tag{A.2}
$$

These expressions have been used in [10] and in [11] to evaluate the MCRB in eqs. (9) and (18) for the DOA estimation problem with array position errors.

Appendix B

A generalization of the Bangs formula under misspecification

In this appendix, we provide a misspecifed version of the Bangs [69] formula derived, also in this case, by Richmond and Horowitz in [10]. As in Appendix A, let the complex data vector $\mathbf{x} \in \Box^{N}$ have a true complex Gaussian distribution such that $\mathbf{x} \Box \, p_{X}(\mathbf{x}) = C\!N \, (\mathbf{d}, \mathbf{B})$ where **d** and **B** denote the (possibly complex) true mean value vector and the true covariance matrix. Let the assumed distribution for the data vector be another complex Gaussian distribution such that $\mathbf{x} \Box \, f_{X}(\mathbf{x}; \boldsymbol{\theta}) = C\!N \, (\mathbf{r}, \mathbf{R}(\boldsymbol{\theta}))$ where **r** is the (possibly complex) assumed mean value, generally

different form $\mathbf{d}$, and $\mathbf{R(\theta)} \square \mathbf{R_\theta}$ denotes the (possibly complex) assumed covariance matrix parameterized by a parameter vector $\mathbf{\theta} \in \mathbf{\Theta} \subset \square^d$ and generally different from $\mathbf{B} \, \forall \mathbf{\theta}$. For the generalization to complex parameter vector we refer the reader to [10]. Under these assumptions, the matrices $\mathbf{A_\theta}$ and $\mathbf{B_\theta}$ in eqs. (1) and (4) can be written explicitly through the so-called misspecified Bangs formulae [10]. In particular, we have:

$$
\begin{aligned}
\left[\mathbf{A_\theta}\right]_{i,k} = & \operatorname{tr}\left(\mathbf{R_\theta}^{-1} \frac{\partial^2 \mathbf{R_\theta}}{\partial\theta_i \partial\theta_k} \left(\mathbf{R_\theta}^{-1}\left(\mathbf{B} + (\mathbf{d} - \mathbf{r(\theta)})(\mathbf{d} - \mathbf{r(\theta)})^H\right) - \mathbf{I}_N \right) \right) \\
& - \operatorname{tr}\left(\mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_k} \mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_i} \left(\mathbf{R_\theta}^{-1}\left(\mathbf{B} + (\mathbf{d} - \mathbf{r(\theta)})(\mathbf{d} - \mathbf{r(\theta)})^H\right) - \mathbf{I}_N \right) \right) \\
& - \operatorname{tr}\left(\mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_i} \mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_k} \mathbf{R_\theta}^{-1}\left(\mathbf{B} + (\mathbf{d} - \mathbf{r(\theta)})(\mathbf{d} - \mathbf{r(\theta)})^H\right) \right),
\end{aligned}
\tag{B.1}
$$

$$
\begin{aligned}
\left[\mathbf{B_\theta}\right]_{i,k} = & \operatorname{tr}\left(\mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_i} \mathbf{R_\theta}^{-1}\mathbf{B}\mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_k} \mathbf{R_\theta}^{-1}\left(\mathbf{B} + (\mathbf{d} - \mathbf{r(\theta)})(\mathbf{d} - \mathbf{r(\theta)})^H\right) \right) \\
& + (\mathbf{d} - \mathbf{r(\theta)})^H \mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_k} \mathbf{R_\theta}^{-1}\mathbf{B}\mathbf{R_\theta}^{-1} \frac{\partial \mathbf{R_\theta}}{\partial\theta_i} \mathbf{R_\theta}^{-1}(\mathbf{d} - \mathbf{r(\theta)}),
\end{aligned}
\tag{B.2}
$$

where $\mathbf{I}_N$ is the identity matrix of order N.

Appendix C

Compact expression for the MCRB in the CES family

In this appendix, we derive a compact expression useful to evaluate the MCRB for the scatter matrix estimation in the family of CES distribution. This expression follows directly from the results obtained in [41]. We assume that both the true distribution $p_X(\mathbf{x})$ (that implicitly depends on the true scatter matrix $\bar{\mathbf{\Sigma}}$, then according to the notation used before, $\bar{\mathbf{\theta}} = \operatorname{vecs}(\bar{\mathbf{\Sigma}})$) and the assumed

distribution $f_X(\mathbf{x}; \mathbf{\Sigma})$ belong to the zero-mean CES distribution class, as shown in eqs. (48) and (49). Moreover, we define $Q \triangleq \mathbf{x}^H \mathbf{\Sigma}^{-1} \mathbf{x}$ as in eq. (54).

Compact expression for the matrix $\mathbf{B}_{\bar{\theta}}$

In [41] the matrix $\mathbf{B}_{\bar{\theta}}$ has been obtained element-by-element as:

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

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

For notation simplicity, we define:

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

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

By using the properties of the vec operator, of the Duplication matrix $\mathbf{D}_N$ and of the Kronecker product [42] [43], we have:

$$
\mathbf{B}_{\bar{\theta}} = \mathbf{D}_N^T \left[B_1 \operatorname{vec}\left(\bar{\mathbf{\Sigma}}^{-1} \right) \operatorname{vec}\left(\bar{\mathbf{\Sigma}}^{-1} \right)^T + B_2 \bar{\mathbf{\Sigma}}^{-1} \otimes \bar{\mathbf{\Sigma}}^{-1} \right] \mathbf{D}_N
\tag{C.4}
$$

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

In [41] the matrix $\mathbf{A}_{\bar{\theta}}$ has been obtained element-by-element as:

$$
\begin{aligned}
\left[\mathbf{A}_{\bar{\theta}}\right]_{ij} &= E_p\left\{\frac{\partial^2 \ln f_X(\mathbf{x};\theta)}{\partial \theta_i \partial \theta_j}\right\} = \\
&= \left(1 + \frac{2}{N} E\left\{Q\frac{\partial \ln g(Q)}{\partial Q}\right\} + \frac{1}{N(N+1)} E\left\{Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2}\right\}\right)\mathrm{tr}\left(\bar{\Sigma}^{-1}\mathbf{A}_i\bar{\Sigma}^{-1}\mathbf{A}_j\right) + \quad (C.5) \\
&+ \frac{1}{N(N+1)} E\left\{Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2}\right\}\mathrm{tr}\left(\bar{\Sigma}^{-1}\mathbf{A}_i\right)\mathrm{tr}\left(\bar{\Sigma}^{-1}\mathbf{A}_j\right)
\end{aligned}
$$

For notation simplicity, we define:

$$
A_2 = 1 + \frac{2}{N} E\left\{Q\frac{\partial \ln g(Q)}{\partial Q}\right\} + \frac{1}{N(N+1)} E\left\{Q^2 \frac{\partial^2 \ln g(Q)}{\partial Q^2}\right\}, \quad (C.6)
$$

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

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

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

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

$$
\mathbf{A}_{\bar{\theta}}^{-1} = \mathbf{D}_N^\dagger\left[\frac{1}{A_2}\bar{\Sigma}\otimes\bar{\Sigma} - \frac{A_1}{A_2(A_2 + NA_1)}\mathrm{vec}\left(\bar{\Sigma}\right)\mathrm{vec}\left(\bar{\Sigma}\right)^T + \right]\left(\mathbf{D}_N^\dagger\right)^T. \quad (C.9)
$$

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

$$
\begin{aligned}
\mathrm{MCRB}(\bar{\theta}) &= \frac{1}{M}\mathbf{A}_{\bar{\theta}}^{-1}\mathbf{B}_{\bar{\theta}}\mathbf{A}_{\bar{\theta}}^{-1} \\
&= \frac{1}{M}\mathbf{D}_N^\dagger\left[\frac{B_2}{A_2^2}\bar{\Sigma}\otimes\bar{\Sigma} + \left(\frac{B_1}{A_2^2} - \frac{2A_1(B_2 + NB_1)}{A_2(A_2 + NA_1)} + \frac{2NA_1^2(B_2 + NB_1)}{A_2^2(A_2 + NA_1)^2}\right)\mathrm{vec}\left(\bar{\Sigma}\right)\mathrm{vec}\left(\bar{\Sigma}\right)^T\right]\left(\mathbf{D}_N^\dagger\right)^T. \quad (C.10)
\end{aligned}
$$

References

[1]   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.

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

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

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

[5]   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.

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

[7]   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.

[8]   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.

[9]   C. D. Richmond, L. L. Horowitz, "Parameter bounds under misspecified models," *Conference on Signals, Systems and Computers, 2013 Asilomar*, pp.176-180, 3-6 Nov. 2013.

[10]   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.

[11]   C. Ren, M. N. El Korso, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux, "Performances bounds under misspecification model for MIMO radar application," *in Proc. of Eur. Signal Process. Conf.* (EUSIPCO), Nice, France 2015, pp. 514-518.

[12]   C. D. Richmond, P. Basu, "Bayesian framework and radar: on misspecified bounds and radar-communication cooperation," *IEEE Workshop on Statistical Signal Processing 2016 (SSP)*, Palma de Mallorca, Spain, 26 – 29 June, 2016.

[13]   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.

[14]   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.

[15]   S. Fortunati, F. Gini, M. S. Greco, "The Misspecified Cramér-Rao Bound and its Application to the Scatter Matrix estimation in Complex Elliptically Symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2387-2399, May 2016.

[16] 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.

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

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

[19] 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.

[20] P. Stoica and T. Söderström, "On non singular information matrices and local identifiability", *Int. J. Control*, vol. 36, pp. 323-329, 1982.

[21] P.J. Schreier, L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge University Press 2010.

[22] A. K. Jagannatham, B. D. Rao, "Cramér-Rao lower bound for constrained complex parameters," *IEEE Signal Process. Lett.*, vol. 11, no. 11, Nov. 2004.

[23] 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.

[24] L. T. McWhorter, L. L. Scharf, "Properties of quadratic covariance bounds," *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993.* , pp.1176 – 1180, vol. 2, 1-3 Nov 1993.

[25] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints", *IEEE Trans. on Inf. Theory*, vol. 6, no. 6, pp.1285–1301, Nov. 1990.

[26] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramér-Rao bound," *IEEE Trans. Signal Processing*, vol. 41, pp. 2247–2249, June 1993.

[27] P. Stoica, and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," IEEE Signal Process. Lett., vol. 5, no. 7, pp. 177-179, Jul. 1998.

[28] T. J. Moore, R. J. Kozick, and B. M. Sadler, "The constrained Cramér–Rao bound from the perspective of fitting a model," IEEE Signal Process. Lett., vol.14, no.8, pp.564-567, Aug. 2007.

[29] S. Fortunati, F. Gini, M. S. Greco, "The Constrained Misspecified Cramér-Rao Bound," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 718-721, May 2016.

[30] M. Spivak, *Calculus on Manifolds*, MA: Addison-Wesley, 1965.

[31] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, NY, 1991.

[32] 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.

[33] A. Balleri, A. Nehorai, J. Wang, "Maximum likelihood estimation for compound-gaussian clutter with inverse gamma texture," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 43, No. 2, pp. 775-779, April 2007.

[34] J. Wang, A. Dogandzic and A. Nehorai, "Maximum Likelihood Estimation of Compound-Gaussian Clutter and Target Parameters," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3884-3898, Oct. 2006.

[35] 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.

[36] A.Younsi, M. Greco, F. Gini, A. Zoubir, "Performance of the adaptive generalised matched subspace constant false alarm rate detector in non-Gaussian noise: An experimental analysis," IET Radar, Sonar and Navigation, vol. 3, no. 3, pp. 195–202, 2009.

[37] 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.

[38] J. R. Magnus, H. Neudecker, "The commutation matrix: some properties and applications," *The Annals of Statistics*, vol. 7, pp. 381-394, 1979.

[39] F. Gini and J. H. Michels, "Performance Analysis of Two Covariance Matrix Estimators in Compound-Gaussian Clutter," *IEE Proceedings Part-F*, vol. 146, No. 3, pp. 133-140, June 1999.

[40] 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.

[41] 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.

[42] 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.

[43] J. R. Magnus, H. Neudecker, "Symmetry, 0-1 matrices and Jacobians: a review," *Econometric Theory*, vol. 2, pp. 157-190, 1986.

[44] K. B. Petersen, M. S. Pedersen, *The Matrix Cookbook*, 2012 [http://matrixcookbook.com].

[45] 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.

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

[47] 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.

[48] D. Tyler, "A distribution-free *M*-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234-251, January 1987.

[49]   F. Gini, M. Greco, and A. Farina, "Clairvoyant and Adaptive Signal Detection in Non-Gaussian Clutter: a Data-Dependent Threshold Interpretation," *IEEE Trans. on Signal Processing*, vol. 47, No. 6, pp. 1522-1531, June 1999.

[50]   F. Gini and M. Greco, "A Suboptimum Approach to Adaptive Coherent radar Detection in Compound-Gaussian Clutter," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 35, No. 3, pp. 1095-1104, July 1999.

[51]   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.

[52]   I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, Seventh Edition, 2007.

[53]   S. Fortunati, F. Gini, M. S. Greco, "On Scatter Matrix Estimation in the Presence of Unknown Extra Parameters: Mismatched Scenario," *EUSIPCO 2016*, Budapest, Hungary, 29 August - 2 September, 2016.

[54]   C.C. Heyde, R. Morton, "On constrained quasi-likelihood estimation," *Biometrika*, vol. 80, no. 4, pp. 755-61, 1993.

[55]   F. Gini, "Performance Analysis of Two Structured Covariance Matrix Estimators in Compound-Gaussian Clutter," *Signal Processing*, vol. 80, No. 2, pp. 365-371, February 2000.

[56] S. Fortunati, F. Gini, M. S. Greco, "Matched, mismatched and robust scatter matrix estimation and hypothesis testing in complex t-distributed data", *EURASIP Journal on Advances in Signal Processing (2016)* 2016:123.

[57] S. Fortunati, M. S. Greco, F. Gini, "The impact of unknown extra parameters on scatter matrix estimation and detection performance in complex t-distributed data," *IEEE Workshop on Statistical Signal Processing 2016 (SSP)*, Palma de Mallorca, Spain, 26 – 29 June, 2016.

[58] L. L. Scharf, B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol.42, no.8, pp.2146-2157, 1994.

[59] F. Gini, "A cumulant-based adaptive technique for coherent radar detection in a mixture of K-distributed clutter and Gaussian disturbance," *IEEE Transactions on Signal Processing*, vol. 45, no. 6, pp. 1507-1519, Jun 1997.

[60] 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 Transactions on Signal Processing*, vol. 50, no. 8, pp. 1908-1915, Aug 2002.

[61] F. Gini, "Sub-optimum Coherent Radar Detection in a Mixture of K-Distributed and Gaussian Clutter," *IEE Proceedings Part-F*, vol. 144, No.1, pp. 39-48, February 1997.

[62] E. Conte, A. De Maio, "Mitigation techniques for non-Gaussian sea clutter," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 2, pp. 284-302, April 2004.

[63] E. Conte, A. De Maio, G. Ricci, "Adaptive CFAR detection in compound-Gaussian clutter with circulant covariance matrix," *IEEE Signal Processing Letters*, vol. 7, no. 3, pp. 63-65, March 2000.

[64]   E. Conte, A. De Maio, G. Ricci, "Covariance matrix estimation for adaptive CFAR detection in compound-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 415-426, Apr 2002.

[65]   E. Conte, A. De Maio, "Exploiting persymmetry for CFAR detection in compound-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 2, pp. 719-724, April 2003.

[66]   F. Pascal, J. P. Ovarlez, "Asymptotic detection performance of the robust ANMF," *23rd European Signal Processing Conference (EUSIPCO)*, 2015, Nice, 2015, pp. 524-528.

[67]   J. P. Ovarlez, F. Pascal, A. Breloy, "Asymptotic detection performance analysis of the robust Adaptive Normalized Matched Filter," *IEEE CAMSAP 2015*, Cancun, 2015 , pp. 137-140.

[68]   D. Slepian, "Estimation of signal parameters in the presence of noise," *Trans. IRE Prof. Group Inf. Theory*, vol. IT-3, pp. 68-89, Mar. 1954.

[69]   W. J. Bangs, *Array processing with generalized beamforming*, PhD dissertation, Yale Univ., New Haven, CT, USA, 1971.

[70]   A. Gusi-Amigó, P. Closas, A. Mallat and L. Vandendorpe, "Ziv-Zakai lower bound for UWB based TOA estimation with unknown interference," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014, pp. 6504-6508.

[71]   A. Gusi-Amigó, P. Ciosas and L. Vandendorpe, "Mean square error performance of sample mean and sample median estimators," *IEEE Statistical Signal Processing Workshop (SSP)*, Palma de Mallorca, 2016, pp. 1-5.

[72] A.Gusi-Amigó, P. Closas, L. Vandendorpe "Mean Square Error bounds for parameter estimation under model misspecification," 2015, arXiv:1511.03982 [math.ST].

[73] J. M. Kantor, C. D. Richmond, B. Correll, D. W. Bliss, "Prior Mismatch in Bayesian Direction of Arrival for Sparse Arrays," *IEEE Radar Conference, Philadelphia*, PA, May 2015, pp. 0811—0816.

[74] P. A. Parker and C. D. Richmond, "Methods and Bounds for Waveform Parameter Estimation with a Misspecified Model," *49th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2015, pp. 1702—1706.