MEASURE-TRANSFORMED QUASI MAXIMUM LIKELIHOOD ESTIMATION WITH APPLICATION TO SOURCE LOCALIZATION

Koby Todros

Ben-Gurion University of the Negev

Alfred O. Hero

University of Michigan

ABSTRACT

In this paper, we consider the problem of estimating a deterministic vector parameter when the likelihood function is unknown or not expressible. We develop an estimator, called measure-transformed quasi maximum likelihood estimator (MT-QMLE), that minimizes the empirical Kullback-Leibler divergence between the transformed probability measure of the data and a hypothesized Gaussian probability distribution. By judicious choice of the transform we show that the proposed estimator can gain sensitivity to higher-order statistical information and resilience to outliers. Under some regularity conditions we show that the MT-QMLE is consistent, asymptotically normal and unbiased. Furthermore, we derive a necessary and sufficient condition for its asymptotic efficiency. The MT-QMLE is applied to source localization in a simulation example that illustrates its sensitivity to higher-order information and resilience to outliers.

Index Terms— Higher-order statistics, parameter estimation, probability measure transform, robust estimation, source localization.

1. INTRODUCTION

Classical multivariate estimation [1], [2] deals with the problem of estimating a deterministic vector parameter using a sequence of multivariate samples from an underlying probability distribution. When the probability distribution is known to lie within a specified parametric family of probability measures, parameter estimation techniques such as the method of maximum likelihood [3] can be implemented that utilize complete statistical information. In many practical scenarios this knowledge is unavailable, and therefore, one must resort to other methods that require partial statistical information.

One of the most popular techniques of this kind is the method of moments [3], [4] that is based on fitting multivariate cumulants of certain orders to their empirical estimates. Usually, first and second-order cumulants, i.e., the mean vector and covariance matrix, are used. Their popularity arises from the fact that they are easy to manipulate, their sample estimates have simple implementations, and the performance analysis of the resulting estimators is often traceable. In some circumstances, first and second-order cumulants may be non-informative, such as for certain types of non-Gaussian data. In order to overcome this limitation higher-order cumulants may be incorporated. However, unlike first and second-order cumulants, higher-order cumulants involve complicated tensor analysis [5], and their empirical estimates are highly non-robust to outliers and have increased computational and sample complexity.

Based on the observation that the mean vector and the covariance matrix are the gradient vector and Hessian matrix of the cumulant generating function evaluated at the origin, another framework that does not require knowledge of the likelihood function has been proposed in [6]. This method operates by fitting the gradient or Hessian of the cumulant generating function evaluated at some off-origin points to their empirical estimates. These points are called processing points and can be judiciously chosen to improve estimation performance. Similarly to the mean vector and the covariance matrix, the resultant cumulant-related quantities have appealing decomposability properties and their sample estimates have simple implementations. Additionally, they have the key advantage that they involve higher-order statistical information. This approach has been successfully applied to signal gain estimation [6] and to auto-regression parameter estimation [7], where a data-driven procedure for optimal selection of the processing points has been devised that minimizes an empirical estimate of the asymptotic mean-squared-error (MSE). Other applications can be found in [8]-[13].

Recently, we have shown that the off-origin gradient vector and Hessian matrix of the cumulant generating function may be viewed as the standard mean vector and covariance matrix under some transformed probability measure [14]-[16]. Hence, a wider class of estimators than the one proposed in [6] can be obtained by considering other measure-transformed mean vectors and covariance matrices. These estimators can gain sensitivity to higher-order statistical information, resilience to outliers, and yet have the computational advantages of first and second-order methods of moments.

In this paper, we use this measure transformation approach to develop a new estimator that operates by jointly fitting a measure-transformed mean vector and covariance matrix to their empirical estimates. The proposed transform is structured by a non-negative function, called the MT-function, and maps the probability distribution into a set of new probability measures on the observation space. By modifying the MT-function, classes of measure transformations can be obtained that have different useful properties. Under the proposed transform we define the measure-transformed (MT) mean vector and covariance matrix, derive their strongly consistent estimates, and study their sensitivity to higher-order statistical information and resilience to outliers.

Similarly to the quasi maximum likelihood estimator (QMLE) [17], [18], the proposed estimator, called MT-QMLE minimizes the empirical Kullback-Leibler divergence [19] between the transformed probability measure and a hypothesized probability distribution. In this paper, the hypothesized distribution is a Gaussian probability measure with the MT-mean vector and MT-covariance matrix parameterized by the unknown parameter to be estimated. Under some regularity conditions we show that the MT-QMLE is consistent, asymptotically normal and unbiased. Additionally, a necessary and sufficient condition for asymptotic efficiency is derived. Furthermore, a data-driven procedure for optimal selection of the MT-function within some parametric class of functions is developed that minimizes an empirical estimate of the asymptotic MSE.

We illustrate the MT-QMLE for the problem of source local-

This work was partially supported by the US National Science Foundation under grant CCF-1217880.

ization in heavy-tailed compound Gaussian noise [21] that produces outliers. By specifying the MT-function within the family of zero-centered Gaussian functions parameterized by a scale parameter, we show that unlike the sample covariance matrix (SCM) based estimator [22] the MT-QMLE is resilient to outliers. Moreover, we show the MT-QMLE performs similarly to the maximum-likelihood estimator (MLE) that, unlike the MT-QMLE, requires complete knowledge of the likelihood function.

The paper is organized as follows. In Section 2, we define the MT-mean and MT-covariance and derive their empirical estimates. In Section 3, we use these quantities to construct the MT-QMLE. The proposed estimator is applied to source localization in Section 4. In Section 5, the main points of this contribution are summarized. Proofs for the theorems and propositions stated throughout the paper will be provided in the full length journal version.

2. MEASURE-TRANSFORMED MEAN AND COVARIANCE

In this section, we develop a transform on the probability measure of a random vector whose probability distribution is parametrized by the vector parameter to be estimated. Under the proposed transform, we define the measure-transformed mean vector and covariance matrix, derive their strongly consistent estimates, and establish their sensitivity to higher-order statistical information and resilience to outliers. These quantities will be used in the following section to construct the measure-transformed quasi maximum likelihood estimator.

2.1. Probability measure transform

We define the measure space $(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X};\boldsymbol{\theta}})$, where $\mathcal{X}$ is the observation space of a random vector $\mathbf{X} \in \mathbb{C}^p$, $\mathcal{S}_{\mathcal{X}}$ is a σ-algebra over $\mathcal{X}$, $P_{\mathbf{X};\boldsymbol{\theta}}$ is a probability measure that belongs to some unknown parametric family of probability measures $\{P_{\mathbf{X};\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \boldsymbol{\Theta}\}$ on $\mathcal{S}_{\mathcal{X}}$, and $\boldsymbol{\Theta} \subseteq \mathbb{R}^m$ denotes the parameter space.

Definition 1. *Given a non-negative function $u : \mathbb{C}^p \to \mathbb{R}_+$ satisfying*

$$0 < \mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] < \infty, \tag{1}$$

where $\mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] \triangleq \int_{\mathcal{X}} u\left(\mathbf{x}\right) dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right)$ and $\mathbf{x} \in \mathcal{X}$, a transform on $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined via the relation:

$$Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(A\right) \triangleq \mathrm{T}_u\left[P_{\mathbf{X};\boldsymbol{\theta}}\right]\left(A\right) = \int_A \varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right) dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right), \tag{2}$$

where $A \in \mathcal{S}_{\mathcal{X}}$ and $\varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right) \triangleq u\left(\mathbf{x}\right)/\mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right]$. The function $u\left(\cdot\right)$ is called the MT-function.

Proposition 1 (Properties of the transform)**.** *Let $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ be defined by relation (2). Then 1) $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$. 2) $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ is absolutely continuous w.r.t. $P_{\mathbf{X};\boldsymbol{\theta}}$, with Radon-Nikodym derivative [20]:*

$$dQ^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right)/dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right) = \varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right). \tag{3}$$

The probability measure $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ is said to be generated by the MT-function $u\left(\cdot\right)$. By modifying $u\left(\cdot\right)$, such that the condition (1) is satisfied, virtually any probability measure on $\mathcal{S}_{\mathcal{X}}$ can be obtained.

2.2. The MT-mean and MT-covariance

According to (3) the mean vector and covariance matrix of $\mathbf{X}$ under $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ are given by:

$$\boldsymbol{\mu}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right) \triangleq \mathrm{E}\left[\mathbf{X}\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] \tag{4}$$

and

$$\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right) \triangleq \mathrm{E}\left[\mathbf{X}\mathbf{X}^H\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] - \boldsymbol{\mu}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)\boldsymbol{\mu}^{(u)H}_{\mathbf{X}}\left(\boldsymbol{\theta}\right), \tag{5}$$

respectively. Equations (4) and (5) imply that $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ are weighted mean and covariance of $\mathbf{X}$ under $P_{\mathbf{X};\boldsymbol{\theta}}$, with the weighting function $\varphi_u\left(\cdot;\cdot\right)$ defined below (2). Hence, they can be estimated using only samples from the distribution $P_{\mathbf{X};\boldsymbol{\theta}}$. By modifying the MT-function $u\left(\cdot\right)$, such that the condition (1) is satisfied, the MT-mean and MT-covariance under $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ are modified. In particular, by choosing $u\left(\cdot\right)$ to be any non-zero constant valued function we have $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}} = P_{\mathbf{X};\boldsymbol{\theta}}$, for which the standard mean vector $\boldsymbol{\mu}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ are obtained.

Given a sequence of N i.i.d. samples from $P_{\mathbf{X};\boldsymbol{\theta}}$ the estimators of $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ are defined as:

$$\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} \triangleq \sum_{n=1}^{N} \mathbf{X}_n \hat{\varphi}_u\left(\mathbf{X}_n\right) \tag{6}$$

and

$$\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}} \triangleq \sum_{n=1}^{N} \mathbf{X}_n\mathbf{X}_n^H \hat{\varphi}_u\left(\mathbf{X}_n\right) - \hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}}\hat{\boldsymbol{\mu}}^{(u)H}_{\mathbf{X}}, \tag{7}$$

respectively, where $\hat{\varphi}_u\left(\mathbf{X}_n\right) \triangleq u\left(\mathbf{X}_n\right)/\sum_{n=1}^{N} u\left(\mathbf{X}_n\right)$. According to Proposition 2 in [23], if $\mathrm{E}\left[\|\mathbf{X}\|_2^2\, u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] < \infty$ then $\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} \xrightarrow{\text{w.p. 1}} \boldsymbol{\mu}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ and $\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}} \xrightarrow{\text{w.p. 1}} \boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}\left(\boldsymbol{\theta}\right)$ as $N \to \infty$, where "$\xrightarrow{\text{w.p. 1}}$" denotes convergence with probability (w.p.) 1 [24].

2.3. Robustness to outliers

Robustness of the empirical MT-covariance (7) to outliers was studied in [23] using its influence function [25] which describes the effect on the estimator of an infinitesimal contamination at some point $\mathbf{y} \in \mathbb{C}^p$. An estimator is said to be B-robust if its influence function is bounded [25]. In [23] we have shown that if the MT-function $u(\mathbf{y})$ and the product $u(\mathbf{y})\|\mathbf{y}\|_2^2$ are bounded over $\mathbb{C}^p$ then the influence function of the empirical MT-covariance is bounded. Similarly, it can be shown that under the same conditions the influence function of the empirical MT-mean (6) is bounded.

2.4. Sensitivity to higher-order statistical information

Notice that for any non-constant analytic MT-function $u\left(\cdot\right)$, which has a convergent Taylor series expansion, the MT-mean (4) and MT-covariance (5) involve higher-order statistical moments of $P_{\mathbf{X};\boldsymbol{\theta}}$. In particular, by choosing $u\left(\mathbf{x};\boldsymbol{t}\right) \triangleq \exp\left(\mathrm{Re}\left\{\boldsymbol{t}^H\mathbf{x}\right\}\right)$, $\boldsymbol{t} \in \mathbb{C}^p$, the resulting exponential MT-mean and MT-covariance are the gradient and Hessian of the cumulant generating function (up to some scaling factors) that have been used for parameter estimation in [6], [7]. Moreover, by choosing $u\left(\mathbf{x};\boldsymbol{t},\tau\right) \triangleq \exp\left(-\|\mathbf{x}-\boldsymbol{t}\|^2/\tau^2\right)$, $\tau \in \mathbb{R}_{++}$, we obtain the Gaussian MT-mean and MT-covariance that have been used for non-linear correlation analysis [14] and robust MUSIC [23].

3. THE MEASURE-TRANSFORMED QUASI MAXIMUM LIKELIHOOD ESTIMATOR

In this section we develop an estimator for $\boldsymbol{\theta}$ that minimizes the empirical Kulback-Leibler divergence between the transformed probability measure $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and a complex circular Gaussian probability distribution [26] $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\vartheta}}$, $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ with MT-mean $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ and MT-covariance $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$. Regularity conditions for consistency, asymptotically normality and unbiasedness are derived. Additionally, we provide a closed-form expression for the asymptotic MSE and obtain a necessary and sufficient condition for asymptotic efficiency. Optimal selection of the MT-function out of some parametric class of functions is also discussed.

3.1. The MT-QMLE

The Kullback-Leibler divergence (KLD) between $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\vartheta}}$ is defined as [19] :

$$D\left(Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\vartheta}}\right) \triangleq \mathrm{E}\left[\log\frac{q^{(u)}(\mathbf{X};\boldsymbol{\theta})}{\phi^{(u)}(\mathbf{X};\boldsymbol{\vartheta})};Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\right], \tag{8}$$

where $q^{(u)}(\mathbf{x};\boldsymbol{\theta})$ and $\phi^{(u)}(\mathbf{x};\boldsymbol{\vartheta})$ are the density functions of $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\vartheta}}$, respectively. According to (3), $D\left(Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\vartheta}}\right)$ can be estimated using only samples from $P_{\mathbf{X};\boldsymbol{\theta}}$. Therefore, similarly to (6) and (7), an empirical estimate of (8) given a sequence of samples $\mathbf{X}_n$, $n=1,\ldots,N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined as:

$$\hat{D}\left(Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\vartheta}}\right) \triangleq \sum_{n=1}^{N}\hat{\varphi}_u(\mathbf{X}_n)\log\frac{q^{(u)}(\mathbf{X}_n;\boldsymbol{\theta})}{\phi^{(u)}(\mathbf{X}_n;\boldsymbol{\vartheta})}, \tag{9}$$

where $\hat{\varphi}_u(\cdot)$ is defined below (7). The proposed estimator of $\boldsymbol{\theta}$ is obtained by minimization of (9) w.r.t. $\boldsymbol{\vartheta}$, which by (6) and (7) amounts to maximization of the following objective function:

$$\begin{aligned} J_u(\boldsymbol{\vartheta}) \triangleq & -\log\det[\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})] - \mathrm{tr}[(\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta}))^{-1}\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}}] \\ & -(\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} - \boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta}))^H(\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta}))^{-1}(\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} - \boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})). \end{aligned}$$

Hence, the proposed MT-QMLE is given by:

$$\hat{\boldsymbol{\theta}}_u = \arg\max_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} J_u(\boldsymbol{\vartheta}). \tag{10}$$

3.2. Asymptotic performance analysis

Here, we study the asymptotic performance of the estimator (10). For simplicity, we assume a sequence of i.i.d. samples $\mathbf{X}_n$, $n=1,\ldots,N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$.

Theorem 1 (Consistency of $\hat{\boldsymbol{\theta}}_u$)**.** *Assume that the following conditions are satisfied: 1) The parameter space $\boldsymbol{\Theta}$ is compact. 2) $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta}) \neq \boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ or $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta}) \neq \boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ $\forall\boldsymbol{\theta}\neq\boldsymbol{\vartheta}$. 3) $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ is non-singular $\forall\boldsymbol{\vartheta}\in\boldsymbol{\Theta}$ 4) $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ are continuous in $\boldsymbol{\Theta}$. 5) $\mathrm{E}\left[\|\mathbf{X}\|_2^2 u(\mathbf{X});P_{\mathbf{X};\boldsymbol{\theta}}\right]<\infty$. Then,*

$$\hat{\boldsymbol{\theta}}_u \xrightarrow{P} \boldsymbol{\theta} \quad \text{as} \quad N\to\infty,$$

where "$\xrightarrow{P}$" denotes convergence in probability [24].

Theorem 2 (Asymptotic normality and unbiasedness of $\hat{\boldsymbol{\theta}}_u$)**.** *Assume that the following conditions are satisfied: 1) $\hat{\boldsymbol{\theta}}_u$ is consistent. 2) $\boldsymbol{\theta}$ lies in the interior of $\boldsymbol{\Theta}$ which is assumed to be compact. 3) $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\vartheta})$ are twice continuously differentiable in $\boldsymbol{\Theta}$. 4) $\mathrm{E}\left[u^2(\mathbf{X});P_{\mathbf{X};\boldsymbol{\theta}}\right]<\infty$ and $\mathrm{E}\left[\|\mathbf{X}\|_2^4 u^2(\mathbf{X});P_{\mathbf{X};\boldsymbol{\theta}}\right]<\infty$. Then,*

$$\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta} \xrightarrow{D} \mathcal{N}(\mathbf{0},\mathbf{C}_u(\boldsymbol{\theta})) \quad as \quad N\to\infty,$$

where "$\xrightarrow{D}$" denotes convergence in distribution [24]. The asymptotic MSE takes the form:

$$\mathbf{C}_u(\boldsymbol{\theta}) = N^{-1}\mathbf{F}_u^{-1}(\boldsymbol{\theta})\mathbf{G}_u(\boldsymbol{\theta})\mathbf{F}_u^{-1}(\boldsymbol{\theta}), \tag{11}$$

where $\mathbf{F}_u(\boldsymbol{\theta}) \triangleq -\mathrm{E}[u(\mathbf{X})\boldsymbol{\Gamma}_u(\mathbf{X};\boldsymbol{\theta});P_{\mathbf{X};\boldsymbol{\theta}}]$, $\boldsymbol{\Gamma}_u(\mathbf{X};\boldsymbol{\vartheta}) \triangleq \frac{\partial^2\log\phi^{(u)}(\mathbf{X};\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^T}$, $\mathbf{G}_u(\boldsymbol{\theta}) \triangleq \mathrm{E}\left[u^2(\mathbf{X})\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta})\boldsymbol{\psi}_u^T(\mathbf{X};\boldsymbol{\theta});P_{\mathbf{X};\boldsymbol{\theta}}\right]$, $\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\vartheta}) \triangleq (\frac{\partial\log\phi^{(u)}(\mathbf{X};\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}})^T$, and it is assumed that $\mathbf{F}_u(\boldsymbol{\theta})$ is non-singular.

The following proposition which relates the asymptotic MSE of $\hat{\boldsymbol{\theta}}_u$ and the Cramér-Rao lower bound (CRLB) [27], [28], follows directly from (11), the covariance semi-inequality [3], and the identity $\mathbf{F}_u(\boldsymbol{\theta}) = \mathrm{E}[u(\mathbf{X})\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta})\boldsymbol{\eta}^T(\mathbf{X};\boldsymbol{\theta});P_{\mathbf{X};\boldsymbol{\theta}}]$, where the vector $\boldsymbol{\eta}(\mathbf{X};\boldsymbol{\vartheta}) \triangleq (\frac{\partial\log f(\mathbf{X};\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}})^T$ and $f(\mathbf{X};\boldsymbol{\vartheta})$ is the density of $P_{\mathbf{X};\boldsymbol{\vartheta}}$.

Proposition 2 (Relation to the CRLB)**.** *Assume that the Fisher information matrix $\mathbf{I}_{\mathrm{FIM}}(\boldsymbol{\theta}) \triangleq \mathrm{E}\left[\boldsymbol{\eta}(\mathbf{X};\boldsymbol{\theta})\boldsymbol{\eta}^T(\mathbf{X};\boldsymbol{\theta});P_{\mathbf{X};\boldsymbol{\theta}}\right]$ is non-singular. Then,*

$$\mathbf{C}_u(\boldsymbol{\theta}) \succeq N^{-1}\mathbf{I}^{-1}_{\mathrm{FIM}}(\boldsymbol{\theta}),$$

where equality holds if and only if

$$\boldsymbol{\eta}(\mathbf{X};\boldsymbol{\theta}) = \mathbf{I}_{\mathrm{FIM}}(\boldsymbol{\theta})\mathbf{F}_u^{-1}(\boldsymbol{\theta})\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta})u(\mathbf{X}) \quad \textit{w.p. 1.} \tag{12}$$

One can verify that when $P_{\mathbf{X};\boldsymbol{\theta}}$ is a Gaussian measure, the condition (12) is satisfied only for non-zero constant valued MT-functions for which $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}} = P_{\mathbf{X};\boldsymbol{\theta}}$ and the resulting MT-mean and MT-covariance (4), (5) only involve first and second-order moments. This implies that in the Gaussian case, non-constant MT-functions will always lead to asymptotic performance degradation. In the non-Gaussian case, however, there may be cases where the MT-function should deviate from a constant value in order to decrease the asymptotic MSE. This deviation results in weighted mean and covariance that involve higher-order moments.

3.3. Optimal choice of the MT-function

We propose to specify the MT-function within some parametric family $\{u(\mathbf{X};\boldsymbol{\omega}),\boldsymbol{\omega}\in\boldsymbol{\Omega}\subseteq\mathbb{C}^r\}$ that satisfies the conditions stated in Definition 1 and Theorem 2. An optimal choice of the MT-function parameter $\boldsymbol{\omega}$ would be this that minimizes an empirical estimate of the asymptotic MSE (11). The proposed empirical estimate is constructed by the same sequence of samples used for obtaining the MT-QMLE (10) and takes the following form:

$$\begin{aligned} \hat{\mathbf{C}}_u(\hat{\boldsymbol{\theta}}_u(\boldsymbol{\omega}),\boldsymbol{\omega}) \triangleq & N^{-1}\hat{\mathbf{F}}_u^{-1}(\hat{\boldsymbol{\theta}}_u(\boldsymbol{\omega}),\boldsymbol{\omega})\hat{\mathbf{G}}_u(\hat{\boldsymbol{\theta}}_u(\boldsymbol{\omega}),\boldsymbol{\omega}) \\ & \times\hat{\mathbf{F}}_u^{-1}(\hat{\boldsymbol{\theta}}_u(\boldsymbol{\omega}),\boldsymbol{\omega}), \end{aligned} \tag{13}$$

where $\hat{\mathbf{F}}_u(\boldsymbol{\vartheta},\boldsymbol{\omega}) \triangleq -\frac{1}{N}\sum_{n=1}^{N}u(\mathbf{X}_n;\boldsymbol{\omega})\boldsymbol{\Gamma}_u(\mathbf{X}_n;\boldsymbol{\vartheta},\boldsymbol{\omega})$, and $\hat{\mathbf{G}}_u(\boldsymbol{\vartheta},\boldsymbol{\omega}) \triangleq \frac{1}{N}\sum_{n=1}^{N}u^2(\mathbf{X}_n;\boldsymbol{\omega})\boldsymbol{\psi}_u(\mathbf{X}_n;\boldsymbol{\vartheta},\boldsymbol{\omega})\boldsymbol{\psi}_u^T(\mathbf{X}_n;\boldsymbol{\vartheta},\boldsymbol{\omega})$. It can be shown that if the conditions in Theorem 2 are satisfied (13) is a consistent estimator of the asymptotic MSE.

4. APPLICATION: SOURCE LOCALIZATION

In this section, we illustrate the use of the proposed MT-QMLE (10) for robust source localization in heavy-tailed compound Gaussian noise. We consider an array of p sensors that receive a signal generated by a narrowband incoherent far-field point source with azimuthal direction-of-arrival (DOA) θ. Under this model the array output satisfies [29]:

$$\mathbf{X}_n = S_n\mathbf{a}(\theta) + \mathbf{W}_n, \tag{14}$$

where $n \in \mathbb{N}$ is a discrete time index, $\mathbf{X}_n \in \mathbb{C}^p$ is the vector of received signals, $S_n \in \mathbb{C}$ is the emitted signal, $\mathbf{a}(\theta) \in \mathbb{C}^p$ is the steering vector of the array toward direction θ and $\mathbf{W}_n \in \mathbb{C}^p$ is an additive noise. We assume that the following conditions are satisfied: 1) θ lies in the interior of a closed interval $\Theta \subset \mathbb{R}$, 2) the emitted signal is symmetrically distributed about the origin, 3) $\mathbf{a}(\vartheta)$ is twice continuously differentiable with $\|\mathbf{a}(\vartheta)\|_2 = \sqrt{p}$, 4) $\mathbf{a}(\theta) \neq \mathbf{a}(\vartheta)$ $\forall \theta \neq \vartheta \in \Theta$ 5) S_n and $\mathbf{W}_n$ are statistically independent and first-order stationary, and 6) the noise component is compound Gaussian with stochastic representation [21]:

$$\mathbf{W}_n = \nu_n \mathbf{Z}_n, \tag{15}$$

where $\nu_n \in \mathbb{R}_{++}$ is a first-order stationary process, called the texture component, and $\mathbf{Z}_n \in \mathbb{C}^p$ is a proper-complex wide-sense stationary Gaussian process with zero-mean and scaled unit covariance $\sigma^2_{\mathbf{Z}}\mathbf{I}$, which is statistically independent of ν_n.

In order to gain robustness against outliers, as well as sensitivity to higher-order moments, we specify the MT-function in the zero-centred Gaussian family of functions parametrized by a width parameter ω, i.e.,

$$u_{\mathrm{G}}(\mathbf{x}, \omega) = (\pi\omega^2)^{-p} \exp\left(-\|\mathbf{x}\|^2/\omega^2\right), \quad \omega \in \mathbb{R}_{++}. \tag{16}$$

Notice that the MT-function (16) satisfies the B-robustness conditions stated in Subsection 2.3. Using (4), (5) and (14)-(16) it can be shown that the MT-mean and MT-covariance under $Q^{(u_{\mathrm{G}})}_{\mathbf{X};\vartheta}$ are:

$$\boldsymbol{\mu}^{(u_{\mathrm{G}})}_{\mathbf{X}}(\vartheta, \omega) = \mathbf{0} \tag{17}$$

and

$$\boldsymbol{\Sigma}^{(u_{\mathrm{G}})}_{\mathbf{X}}(\vartheta, \omega) = r_S(\omega)\,\mathbf{a}(\vartheta)\,\mathbf{a}^H(\vartheta) + r_W(\omega)\,\mathbf{I}, \tag{18}$$

respectively, where $r_S(\omega)$ and $r_W(\omega)$ are some strictly positive functions of ω. Hence, by substituting (16)-(18) into (10) we obtain:

$$\hat{\theta}_{u_{\mathrm{G}}}(\omega) = \arg\max_{\vartheta \in \Theta} \mathbf{a}^H(\vartheta)\,\hat{\mathbf{C}}^{(u_{\mathrm{G}})}_{\mathbf{X}}(\omega)\,\mathbf{a}(\vartheta),$$

where $\hat{\mathbf{C}}^{(u_{\mathrm{G}})}_{\mathbf{X}}(\omega) \triangleq \hat{\boldsymbol{\Sigma}}^{(u_{\mathrm{G}})}_{\mathbf{X}}(\omega) + \hat{\boldsymbol{\mu}}^{(u_{\mathrm{G}})}_{\mathbf{X}}(\omega)\,\hat{\boldsymbol{\mu}}^{(u_{\mathrm{G}})H}_{\mathbf{X}}(\omega)$.

Under the considered settings, it can be shown that the conditions stated in Theorems 1 and 2 are satisfied. The resulting asymptotic MSE (11) takes the form:

$$C_{u_{\mathrm{G}}}(\theta, \omega) = \frac{\mathrm{E}\left[\left(\bar{\nu}^4 + \frac{\bar{\nu}^2\omega^2 p|S|^2}{2\bar{\nu}^2+\omega^2}\right) h\left(\sqrt{2p}S, \sqrt{2}\bar{\nu}, \omega\right); P_{S,\nu}\right]}{\mathrm{E}^2\left[p|S|^2 h\left(\sqrt{p}S, \bar{\nu}, \omega\right); P_{S,\nu}\right]} \times \frac{6}{\pi^2\cos^2(\theta)(p^2-1)N}, \tag{19}$$

where $h(S, \nu, \omega) \triangleq \left((\nu^2+\omega^2)/\omega^2\right)^{-p-2} \exp\left(-|S|^2/(\nu^2+\omega^2)\right)$ and $\bar{\nu} \triangleq \nu\sigma^2_{\mathbf{Z}}$. Furthermore, its empirical estimate (13) is given by:

$$\hat{C}_{u_{\mathrm{G}}}(\hat{\theta}_{u_{\mathrm{G}}}(\omega), \omega) = \frac{\sum_{n=1}^{N} \alpha^2(\mathbf{X}_n, \hat{\theta}_{u_{\mathrm{G}}}(\omega))\,u^2_{\mathrm{G}}(\mathbf{X}_n, \omega)}{\left(\sum_{n=1}^{N} \beta(\mathbf{X}_n, \hat{\theta}_{u_{\mathrm{G}}}(\omega))\,u_{\mathrm{G}}(\mathbf{X}_n, \omega)\right)^2}, \tag{20}$$

where $\alpha(\mathbf{X}, \vartheta) \triangleq 2\mathrm{Re}\left\{\dot{\mathbf{a}}^H(\vartheta)\,\mathbf{X}\mathbf{X}^H\mathbf{a}(\vartheta)\right\}$, $\dot{\mathbf{a}}(\vartheta) \triangleq d\mathbf{a}(\vartheta)/d\vartheta$, $\beta(\mathbf{X}, \vartheta) \triangleq 2\mathrm{Re}\left\{\ddot{\mathbf{a}}^H(\vartheta)\,\mathbf{X}\mathbf{X}^H\mathbf{a}(\vartheta) + |\dot{\mathbf{a}}^H(\vartheta)\,\mathbf{X}|^2\right\}$ and $\ddot{\mathbf{a}}(\vartheta) \triangleq d^2\mathbf{a}(\vartheta)/d^2\vartheta$.

In the following simulation examples, we consider a BPSK signal impinging on a 4-element uniform linear array [29] with half wavelength spacing from DOA $\theta = 30°$. We consider an ϵ-contaminated Gaussian noise model [21] under which the texture component of the compound Gaussian noise (15) is a binary random variable with $\nu = 1$ w.p. $1-\epsilon$ and $\nu = c$ w.p. ϵ. The parameters ϵ and c that control the heaviness of the noise tails were set to 0.2 and 100, respectively. We define the signal-to-noise-ratio (SNR) as $\mathrm{SNR} \triangleq 10\log_{10}\sigma^2_S/\left[\sigma^2_{\mathbf{Z}}\left((1-\epsilon)+\epsilon c^2\right)\right]$.

In the first example, we compared the asymptotic MSE (19) to its empirical estimate (20) as a function of ω obtained from a single realization of $N = 10000$ i.i.d. snapshots with $\mathrm{SNR} = -25$ [dB]. Observing Fig. 1, one sees that due to the consistency of (20) the compared quantities are very close.

In the second example, we compared the empirical, asymptotic (19) and empirical asymptotic (20) MSEs of the MT-QMLE to the empirical MSEs obtained by the SCM-based estimator [22] and the MLE. The optimal Gaussian MT-function parameter ω_{opt} was obtained by minimizing (20) over $\Omega = [1, 30]$. Here, $N = 1000$ snapshots were used with averaging over 10^3 Monte-Carlo simulations. The SNR is used to index the performances as depicted in Fig. 2. One sees that the MT-QMLE outperforms the non-robust SCM-based estimator and performs similarly to the MLE that, unlike the MT-QMLE, requires complete knowledge of the likelihood function. Additionally, one can notice that the empirical, asymptotic and empirical asymptotic MSEs are nearly identical.

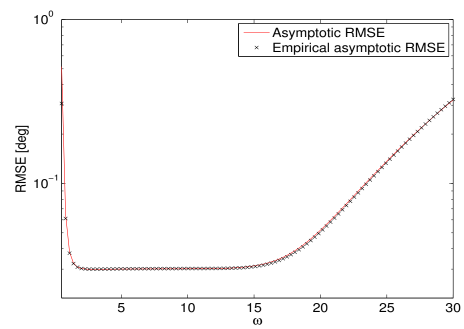

Fig. 1. Asymptotic RMSE (19) and its empirical estimate (20) versus the width parameter ω (16) ($\theta = 30°$, $N = 10^4$, $\mathrm{SNR} = -25$ [dB]).

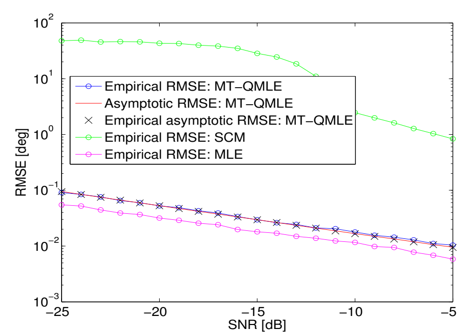

Fig. 2. The empirical, asymptotic (19) and empirical asymptotic (20) RMSEs of the MT-QMLE as compared to the empirical RMSEs of the SCM-based estimator and the MLE ($\theta = 30°$, $N = 10^3$).

5. CONCLUSION

In this paper a new multivariate estimator, called MT-QMLE, was derived by applying a transform to the probability distribution of the data. By specifying the MT-function in the Gaussian family, the proposed estimator was applied to robust source localization in compound Gaussian noise. Exploration of other MT-functions may result in additional estimators in this class that have different useful properties.

6. REFERENCES

[1] R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London,* Series A, vol. 222, pp. 309-368, 1922.

[2] R. A. Fisher, "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society,* vol. 22, pp. 700-725, 1925.

[3] E. L. Lehmann, and G. Casella, *Theory of Point Estimation,* Springer, 1998.

[4] K. Pearson, "Contributions to the mathematical theory of evolution," *Philosophical Transactions of the Royal Society of London,* vol. A, pp. 71-110, 1894.

[5] B. Porat, *Digital processing of random signals: theory and methods,* pp. 294-327, Prentice-Hall, 1994.

[6] A. Slapak and A. Yeredor, "Charrelation and Charm: generic statistics incorporating higher-order information," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5089-5106, 2012.

[7] A. Slapak and A. Yeredor, "Charrelation-based estimation of the parameters of non-Gaussian autoregressive processes," *Proceedings of SSP 2012*, pp. 448-451.

[8] A. Yeredor, "blind source separation via the second characteristic function," *Signal Processing,* vol. 80, pp. 897-902, 2000.

[9] A. Slapak and A. Yeredor, ""Weighting for more": enhancing characteristic-function based ICA with asymptotically optimal weighting," *Signal Processing,* vol. 91, no. 8, pp. 2016-2027, 2011.

[10] A. Yeredor, "Blind channel estimation using first and second derivatives of the characteristic function," *IEEE Signal Processing Letters,* vol. 9, pp. 100-103, 2002.

[11] A. Yeredor, "Substituting the cumulants in the super-exponential blind equalization algorithm," *Proceedings of the ICASSP 2008*.

[12] E. Edinger and A. Yeredor, "Blind MIMO identification using the second characteristic function," *IEEE Transactions on Signal Processing,* vol. 53, pp. 4067-4079, 2005.

[13] A. Yeredor, "MUSIC using off-origin Hessians of the second characteristic function," *Proceedings of SAM 2006*, pp. 176-180.

[14] K. Todros and A. O. Hero, "On measure-transformed canonical correlation analysis," *IEEE Transactions on Signal Processing,* vol. 60, no. 9, pp. 4570-4585, 2012.

[15] K. Todros and A. O. Hero, "Measure transformed canonical correlation analysis with application to financial data," *Proceedings of SAM 2012,* pp. 361-364, 2012.

[16] K. Todros and A. O. Hero, "Measure independent component analysis," available online as arXiv:1302.0730.

[17] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the Econometric Society,* pp. 1-25, 1982.

[18] Q. Ding and S. Kay, "Maximum likelihood estimator under a misspecified model with high signal-to-noise ratio," *IEEE Transactions on Signal Processing,* vol. 59, no. 8, pp. 4012-4016, 2011.

[19] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics,* vol. 22, pp. 79-86, 1951.

[20] G. B. Folland, *Real Analysis*. John Wiley and Sons, 1984.

[21] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions on Signal Processing,* vol. 60, no. 1, pp. 5597-5625, 2012.

[22] H. L. Van Trees, *Optimum Array Processing,* p. 922, John Wiley & Sons, 2002.

[23] K. Todros and A. O. Hero, "Robust measure-transformed MUSIC for DOA estimation," *Proceedings of ICASSP 2014,* pp. 4190-4194.

[24] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.

[25] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

[26] P. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals.* p. 39, Cambridge University Press, 2010.

[27] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Akt. Tidskr.,* vol. 29, pp. 85-94, 1946.

[28] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.,* vol. 37, pp. 81-91, 1945.

[29] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine,* vol. 13, no. 4, pp. 67-94, 1996.