On the maximum likelihood estimator statistics for unimodal elliptical distributions in the high signal-to-noise ratio regime

Steeve Zozor *Member IEEE* , Chengfang Ren *Member IEEE* and Alexandre Renaux

Abstract—In this paper, we study the behavior of the maximum likelihood estimator in the framework of low noise level (or high signal-to-noise ratio), when the data follow an unimodal elliptical distribution. The maximum likelihood estimator appears to be the same as in the Gaussian context, regardless the noise distribution. We also show that the asymptotic distribution of this estimator is unimodal elliptical, where the law is intimately linked to that of the noise distribution. Additionally, this estimator is shown to be not efficient, except in the Gaussian noise case. Finally, we validate our analytic results by some simulations.

Index Terms—Maximum likelihood estimator statistics, high signal-to-noise ratio regime, elliptically symmetric distribution.

I. INTRODUCTION

The statistics of the maximum likelihood estimator (MLE) is fundamental in order to fully characterize the quality of this estimator. The determination of its estimates distribution is not a tractable problem in a general setting. But such a study is generally performed in an asymptotic framework. By asymptotic, one understand in general that the number of data tends to infinity (*i.e.,* is large enough) [1].

However, one may face to the estimation problem dealing with a small number of data, and to be interested in the situation where the signal-to-noise ratio (SNR) is high, or equivalently the noise level is low. This framework is of interest for instance in array processing because the collected data are, in practice, limited by the small number of sensors, by the acquisition time slot and by the memory of the system. This estimation problem in high SNR context was recently studied in [2]–[4] for a very precise context, namely that of the estimation of a parameter observed through a nonlinear function and corrupted by an additive Gaussian noise.

Due to the central limit theorem, modeling an observation to be Gaussian is usual, since data are generally collected under numerous number of sources of perturbations. However, in many situations, this modeling seems not adequate. In the high resolution radar domain for instance, the noise can be consequence of the sum of a small number of echos, this number being different cell to cell, and thus modeled as random. In this domain, the noise is more likely modeled as

S. Zozor is with the Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France (steeve.zozor@gipsa-lab.grenoble-inp.fr).
* Institute of Engineering Univ. Grenoble Alpes
C. Ren is with SONDRA, CentraleSupélec, 91192 Gif-sur-Yvette, France, (chengfang.ren@centralesupelec.fr)
A. Renaux is with LSS, Université Paris-Sud, 91192 Gif-sur-Yvette, France, (alexandre.renaux@u-psud.fr)
This work is partially funded by the ANR ASTRID # ANR-17-ASTR-0015.

a Gaussian scale mixture $\mathbf{x} \stackrel{d}{=} \sqrt{\tau}\,\mathbf{g}$ where τ is a positive random variable and $\mathbf{g}$ a Gaussian vector, independent of τ [5]–[8]. If a such model is quite common in radar processing [9]–[11], its use is also frequent in image processing, where the random variable τ models the texture of an image [12], [13]. The Gaussian scale mixtures belong to the more general class of elliptically distributed random vectors, characterized by the ellipsoidal form of the level sets for the probability density function [14]. Such random vectors are intensively studied in signal processing domains [15]–[20]. The study of the MLE in such a context is thus of high interest. The purpose of this paper is precisely to extend the work of [2], [4] in the context of unimodal elliptically distributed noise.

The paper is organized as follows. Section II presents the basics on the elliptically distributed random vector that we need for our study. Then, section III, the core of the paper, is devoted to the maximum likelihood estimation (MLE) of a parameter $\boldsymbol{\theta}$, observed via n samples through a possibly nonlinear function, and additively corrupted by an unimodal elliptically distributed random vector . The goal of this section is to determine the statistics of the MLE of $\boldsymbol{\theta}$ in the high SNR context. To this end, we follow step by step the outline of [2] and show that, in this low noise level regime, the MLE belongs to the class of elliptically distributed random vectors with a distribution intimately linked to that of the noise. Moreover, we prove that the Cramér-Rao lower bound (CRLB) is not reached, except when the noise is Gaussian. These results differ from what happens in general in the asymptotic regime in terms of high number of samples [1, thm 3.10 p. 449]. Finally, in section IV, we experimentally validate our analytic study with several simulations and goodness-of-fit tests.

II. ELLIPTICALLY DISTRIBUTED RANDOM VECTORS

In the sequel, we place the study to the framework of random vectors $\mathbf{x}$ admitting a probability density function (pdf) $p_{\mathbf{x}}$ w.r.t. the Lebesgue measure.

Definition 1 (Fang *et al.* [14])**.** A random vector $\mathbf{x}$, defined over a set $\Omega \subseteq \mathbb{R}^n$, is said *elliptically distributed* around a deterministic vector $\boldsymbol{\mu} \in \Omega$ if there exist a symmetric positive definite matrix $\mathbf{R}$ such that for any orthogonal matrix $\mathbf{O}$, $\mathbf{O}\mathbf{R}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu}) \stackrel{d}{=} \mathbf{R}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu})$ where "$\stackrel{d}{=}$" means equality in distribution and where $\mathbf{R}^{\frac{1}{2}}$ is the unique symmetric positive definite matrix being square root of $\mathbf{R}$. When $\mathbf{x}$ admits a pdf as assumed all along this paper, its pdf is of the form

$$p_{\mathbf{x}}(\boldsymbol{x}) = |\mathbf{R}|^{-\frac{1}{2}}\, d_n\left((\boldsymbol{x} - \boldsymbol{\mu})^t\, \mathbf{R}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \qquad (1)$$

where $\cdot^t$ denotes the transposition. $\boldsymbol{\mu}$ is a *location parameter*, $\mathbf{R}$ is called *characteristic matrix*, and $d_n : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is called *density generator*. In the sequel, we write in the compact form $\mathbf{x} \sim \mathcal{ED}(\boldsymbol{\mu}, \mathbf{R}, d_n)$.

Under condition of existence, $\boldsymbol{\mu}$ is the mean of $\mathbf{x}$ and $\mathbf{R}$ is proportional to its covariance matrix. The definition could indeed extend to complex random vectors but one can restrict to the real context using the mapping in [14, Chap. 2] or [2].

The denomination "elliptically distributed" is justified by the form of the level set of the pdf that are ellipsoidal located around $\boldsymbol{\mu}$. Moreover, when $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{R} \propto \mathbf{I}$ the identity matrix, $\mathbf{x}$ is said *spherically distributed*, written in short, $\mathbf{x} \sim \mathcal{SD}(d_n)$.

The class of elliptical distributions contains various famous distributions [20], [21], such that the Gaussian law for $d_n(u) \propto \mathrm{e}^{-\frac{u}{2}}$, the Student-t distribution for $d_n(u) \propto \left(1 + \frac{u}{\nu}\right)^{-\frac{n+\nu}{2}}$, $\nu > 0$, the Student-r law where $d_n(u) \propto \left(1 - \frac{u}{\nu}\right)_+^{\frac{\nu-n}{2}}$, $\nu > n - 2$, the generalized Laplacian or K-distributions where $d_n(u) \propto u^{\frac{\nu-n}{4}} K_{\frac{\nu-n}{2}}(\sqrt{\nu u})$, $\nu > 0$), the sub-Gaussian α-stable laws, among many others [14], [22].

These distributions share numerous properties that can be found in [14], [20], [21]. Among them, a useful one is that an elliptically random vector can stochastically be written as a scale mixture of uniform over the n-dimensional unit sphere,

Property 1 (Fang *et al.* [14]). Let $\mathbf{x} \sim \mathcal{ED}(\boldsymbol{\mu}, \mathbf{R}, d_n)$. Then $r = \left\| \mathbf{R}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu}) \right\|$ and $\mathbf{u} = \frac{\mathbf{R}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu})}{\left\| \mathbf{R}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu}) \right\|}$ are independent, so that

$$\mathbf{x} = \boldsymbol{\mu} + r \mathbf{R}^{\frac{1}{2}} \mathbf{u} \qquad (2)$$

where $\mathbf{u}$ is uniformly distributed on the unit sphere $\mathbb{S}^{n-1} = \{s \in \mathbb{R}^n : \|s\| = 1\}$ and r is a positive random variable independent of $\mathbf{u}$ and of pdf $p_r(r) = \omega_n r^{n-1} d_n(r^2)$ with $\omega_n = 2\pi^{\frac{n}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)}$ the surface of $\mathbb{S}^{n-1}$. Moreover, $\mathbb{E}[\mathbf{u}\mathbf{u}^t] = n^{-1}\mathbf{I}$. Thus, if r admit a second order moment, the covariance matrix of $\mathbf{x}$ is given by $\mathbb{C}\mathrm{ov}[\mathbf{x}] = n^{-1}\mathbb{E}[r^2]\mathbf{R}$.

If the density generator d_n of $\mathbf{x}$ is completely monotonic, *i.e.*, for all $k, (-1)^k d_n^{(k)} \geq 0$ where $d_n^{(k)}$ stands for the kth derivative, then $\mathbf{x}$ also stochastically writes $\mathbf{x} \overset{d}{=} \boldsymbol{\mu} + \sqrt{\tau} \, \mathbf{R}^{\frac{1}{2}} \mathbf{g}$ where $\mathbf{g}$ is a standardized Gaussian vector [6]–[8], [23], *i.e.*, $\mathbf{x}$ is a Gaussian Scale Mixture (GSM). The mixture is said *consistent* [24] when τ is not parameterized by n. However, the GSM class is more limited than the elliptically distributed one, for instance the Student-r law is not a GSM (see also [20], [21] for various other examples).

Another quality is the conservation of elliptically symmetric pdf subject to a non degenerate affine transform:

Property 2 (Fang & Kotz [14]). If $\mathbf{x} \sim \mathcal{ED}(\boldsymbol{\mu}, \mathbf{R}, d_n)$, then, for any full rank $m \times n$ matrix $\mathbf{M}$ with $m \leq n$, and for any vector $\boldsymbol{\beta} \in \mathbb{R}^m$, $\boldsymbol{\beta} + \mathbf{Mx} \sim \mathcal{ED}(\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\mu}, \mathbf{MRM}^t, d_m)$, where, for $m < n$,

$$d_m(u) = \frac{\pi^{\frac{n-m}{2}}}{\Gamma\left(\frac{n-m}{2}\right)} \int_{\mathbb{R}_+} v^{\frac{n-m}{2}-1} d_n(u+v) \, \mathrm{d}v \qquad (3)$$

The assumption $m \leq n$ and $\mathbf{M}$ of full rank are necessary to insure that $\boldsymbol{\beta} + \mathbf{Mx}$ admits a pdf. Moreover, it is straightforward to see that for a consistent GSM, the texture τ being independent of the dimension n, the marginal density generator d_m has the same expression than d_n, simply replacing the dimension n by m. For instance, if $\mathbf{x}$ is Student-t with ν degrees of freedom, τ is Gamma distributed with scale and shape parameters both equals to $\frac{\nu}{2}$ [20], [25] and $\boldsymbol{\beta} + \mathbf{Mx}$ remains Student-t with the same degrees of freedom ν.

In the sequel, we slightly restrict the study to the framework of *unimodal laws*, *i.e.,* such that the density generator d_n is decreasing on $\mathbb{R}_+$. We denote such a distribution $\mathcal{ED}^+$ in the elliptical case, and $\mathcal{SD}_n^+$ the spherical restriction.

III. Behavior of the MLE in high SNR context and unimodal elliptical framework

We consider the estimation problem of a real parameter vector[1] $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ observed on $n \geq p$ samples, through a function $\mathbf{m}$, not necessarily linear, and corrupted by an additive noise. More specifically, the noise vector is assumed to be unimodal and elliptically distributed around $\mathbf{0}$ without loss of generality with characteristic matrix $\mathbf{R}$ and density generator d_n which implies that $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]^t \sim \mathcal{ED}^+(\mathbf{m}(\boldsymbol{\theta}), \mathbf{R}, d_n)$ with $\mathbf{m} : \Theta \mapsto \mathbb{R}^n$. Rewriting the characteristic matrix under the form $\mathbf{R} = \sigma^2 \mathbf{C}$ with matrix $\mathbf{C}$ normalized to have the trace $\mathrm{Tr}(\mathbf{C}) = n$, the observation rewrites as follows

$$\mathbf{x} = \mathbf{m}(\boldsymbol{\theta}) + \sigma\boldsymbol{\xi} \qquad (4)$$

with $\boldsymbol{\xi} \sim \mathcal{ED}^+(\mathbf{0}, \mathbf{C}, d_n)$ and where $\sigma \geq 0$ represents the noise level. Indeed, when $r \overset{d}{=} \left\| \mathbf{C}^{-\frac{1}{2}} \boldsymbol{\xi} \right\|$ admits a second order moment, $\sigma^2 \mathrm{Tr}(\mathbb{C}\mathrm{ov}[\boldsymbol{\xi}]) = \sigma^2 \mathbb{E}[r^2]$ represents the "total" power of the noise (see property 1) and thus σ tunes the SNR $\frac{\|\mathbf{m}(\boldsymbol{\theta})\|^2}{\sigma^2 \mathbb{E}[r^2]}$. Matrix $\mathbf{C}$ and function $\mathbf{m}$ are supposed to be known whereas σ is unknown. Moreover, $\mathbf{m}$ belongs to the class of C^2 and is an injective map. The last condition ensures the identifiability of the problem.

As introduced, our purpose is to extend the results of [2] on the ML estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ of $\boldsymbol{\theta}$ for the above described model (4), when $\sigma \to 0$. By definition, the MLE is defined by,

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} = \underset{\boldsymbol{\theta} \in \Theta}{\mathrm{argmax}} \; p_{\mathbf{x}}(\boldsymbol{x}; \boldsymbol{\theta}) \qquad (5)$$

where $p_{\mathbf{x}}(\boldsymbol{x}; \boldsymbol{\theta})$ is the pdf of $\mathbf{x}$, parameterized by $\boldsymbol{\theta}$. Since $p_{\mathbf{x}}(\boldsymbol{x}; \boldsymbol{\theta}) = \left| \sigma^2 \mathbf{C} \right|^{-\frac{1}{2}} d_n\left(\frac{1}{\sigma^2}(\boldsymbol{x} - \mathbf{m}(\boldsymbol{\theta}))^t \mathbf{C}^{-1}(\boldsymbol{x} - \mathbf{m}(\boldsymbol{\theta}))\right)$ where the density generator d_n is decreasing, we obtain

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} = \underset{\boldsymbol{\theta} \in \Theta}{\mathrm{argmin}} \; (\boldsymbol{x} - \mathbf{m}(\boldsymbol{\theta}))^t \mathbf{C}^{-1}(\boldsymbol{x} - \mathbf{m}(\boldsymbol{\theta})). \qquad (6)$$

In other words, the MLE is independent of the noise level σ and of the density generator d_n, leading to the same expression as in the Gaussian context. Thus, the study in the asymptotic regime $\sigma \to 0$ exactly follows the steps in [2]. We come back hereafter to the main steps, leading the asymptotic statistics of $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ in the elliptical context considered here.

---

[1]Dealing with complex parameters, one can write it as real by putting into two blocks its real and imaginary part respectively.

Proposition 1. Let $\boldsymbol{\theta}_0$ be the true parameter of underlying model (4) and $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ the MLE, solution of eq. (6). Then, in the asymptotic regime,

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} \xrightarrow[\sigma \to 0]{\text{a.s.}} \boldsymbol{\theta}_0. \tag{7}$$

where a.s. means that the convergence is almost sure.

Proof. Following [2], let us denote by g the quadratic form,

$$
\begin{aligned}
g(\boldsymbol{\theta}, \sigma) &= (\mathbf{x} - \mathbf{m}(\boldsymbol{\theta}))^t \mathbf{C}^{-1}(\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})) \\
&= (\mathbf{m}(\boldsymbol{\theta}_0) + \sigma \boldsymbol{\xi} - \mathbf{m}(\boldsymbol{\theta}))^t \mathbf{C}^{-1}(\mathbf{m}(\boldsymbol{\theta}_0) + \sigma \boldsymbol{\xi} - \mathbf{m}(\boldsymbol{\theta}))
\end{aligned}
$$

omitting the dependence of g on $\boldsymbol{\xi}$ to simplify the notations. $\mathbf{C}^{-1}$ being symmetric positive definite, $g(\boldsymbol{\theta}, \sigma) \geq 0$ with equality if and only if (iff)

$$\mathbf{m}(\boldsymbol{\theta}_0) + \sigma \boldsymbol{\xi} - \mathbf{m}(\boldsymbol{\theta}) = \mathbf{0} \tag{8}$$

In the limit $\sigma \to 0$, equality (8) is satisfied iff $\mathbf{m}(\boldsymbol{\theta}_0) = \mathbf{m}(\boldsymbol{\theta})$, *i.e.*, iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ due to the injectivity of $\mathbf{m}$ (identifiability assumption). Thus, for almost any realization of $\boldsymbol{\xi}$, $\boldsymbol{\theta}_0$ is the unique global minimum of $g(\boldsymbol{\theta}, 0)$, that finishes the proof. $\square$

The consistency of the estimator insure its convergence to the true parameter when $\sigma \to 0$. As in [2], one can go a step further, determining the asymptotic pdf of the estimator, or more precisely that of $\sigma^{-1}(\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} - \boldsymbol{\theta}_0)$, as follows.

Proposition 2. Let us denote by $\mathbf{J}_{\mathbf{m}}(\boldsymbol{\theta}) = \frac{\partial \mathbf{m}}{\partial \boldsymbol{\theta}^t}(\boldsymbol{\theta})$ the Jacobian matrix of $\mathbf{m}$, with components $\frac{\partial m_i}{\partial \theta_j}$. $\mathbf{m}$ being injective, this $n \times p$ (with $p \leq n$) matrix is of full rank [26, chap. 3] and thus $\mathbf{J}_{\mathbf{m}}^t(\boldsymbol{\theta}_0) \mathbf{C}^{-1} \mathbf{J}_{\mathbf{m}}(\theta)$ is a positive definite Hermitian matrix (and thus invertible). One thus shows that

$$\sigma^{-1}(\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} - \boldsymbol{\theta}_0) \xrightarrow[\sigma \to 0]{\text{d}} \mathcal{ED}^+(\mathbf{0}, \boldsymbol{\Psi}_{\mathbf{m}}, d_p) \tag{9}$$

where $\boldsymbol{\Psi}_{\mathbf{m}}^{-1} = \mathbf{J}_{\mathbf{m}}^t(\boldsymbol{\theta}_0) \mathbf{C}^{-1} \mathbf{J}_{\mathbf{m}}(\boldsymbol{\theta}_0)$ and where "$\xrightarrow{\text{d}}$" means that the convergence is in distribution.

Proof. From eq. (6), function $\mathbf{m}$ being of class C^2 and thus differentiable, the MLE $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ is solution of equation

$$\mathbf{h}(\boldsymbol{\theta}, \sigma) \triangleq \frac{\partial g}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}, \sigma) = \mathbf{0} \tag{10}$$

where $\frac{\partial g}{\partial \boldsymbol{\theta}}$ is a column vector of components $\frac{\partial g}{\partial \theta_i}$. We omit here also the dependence of $\mathbf{h}$ on $\boldsymbol{\xi}$. Eq. (10) has not necessarily a unique solution for a given σ and a given realization $\boldsymbol{\xi}$. However, when $\sigma \to 0$, the couple $(\boldsymbol{\theta}_0, 0)$ is the unique solution of eq. (10). Moreover, $\boldsymbol{\theta}_0$ is asymptotically achieved by $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ due to its consistency proved in proposition 1. Let us then consider $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ around $(\boldsymbol{\theta}_0, 0)$. Following the steps of [2], we use the implicit function theorem [26, chap. 3], that summarizes in our context as follows. (i) Let $\mathbf{f} : \mathbb{R}^p \times \mathbb{R} \mapsto \mathbb{R}^p$ be a function of class C^1, (ii) Let $(\boldsymbol{u}_0, v_0)$ be a point such that $\mathbf{f}(\boldsymbol{u}_0, v_0) = \mathbf{0}$, and (iii) Let assume that the Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \boldsymbol{u}^t}$ is invertible at point $(\boldsymbol{u}_0, v_0)$. Then, there exist a unique function $\boldsymbol{\varphi}$, of class C^1 in a neighbor V of v_0, and a neighbor W of $(\boldsymbol{u}_0, v_0)$ in $\mathbb{R}^p \times V$ such that

$$\forall (\boldsymbol{u}, v) \in W, \quad \mathbf{f}(\boldsymbol{u}, v) = \mathbf{0} \Leftrightarrow \boldsymbol{u} = \boldsymbol{\varphi}(v), \tag{11}$$

$$\text{and} \quad \frac{\partial \boldsymbol{\varphi}}{\partial v}(v_0) = -\left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{u}^t}(\boldsymbol{u}_0, v_0)\right)^{-1} \frac{\partial \mathbf{f}}{\partial v}(\boldsymbol{u}_0, v_0). \tag{12}$$

One thus applies this theorem to $\mathbf{h}$: (i) Function $\mathbf{m}$ being C^2, function $\mathbf{h}$ is C^1; (ii) $\mathbf{h}$ vanishes in $(\boldsymbol{\theta}_0, 0)$; (iii) the Jacobian of $\mathbf{h}$ vs $\boldsymbol{\theta}$ in point $(\boldsymbol{\theta}_0, 0)$ writes

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}^t}(\boldsymbol{\theta}_0, 0) = 2 \mathbf{J}_{\mathbf{m}}^t(\boldsymbol{\theta}_0) \mathbf{C}^{-1} \mathbf{J}_{\mathbf{m}}(\boldsymbol{\theta}_0) = 2 \boldsymbol{\Psi}_{\mathbf{m}}^{-1} \tag{13}$$

that is is invertible. From the implicit function theorem, there exists a unique function $\boldsymbol{\varphi}$ defined in the neighbor of $(\boldsymbol{\theta}_0, 0)$ such that for any $(\boldsymbol{\theta}, \sigma)$ in this neighbor satisfying $\mathbf{h}(\boldsymbol{\theta}, \sigma) = \mathbf{0}$, then $\boldsymbol{\theta} = \boldsymbol{\varphi}(\sigma)$. Because $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ is a solution of $\mathbf{h}(\boldsymbol{\theta}, \sigma) = \mathbf{0}$, and from the consistency property $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} \xrightarrow[\sigma \to 0]{a.s.} \boldsymbol{\theta}_0$, there exists a value $\varepsilon > 0$ such that for any $\sigma \leq \varepsilon$, $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ is solution of equation $\mathbf{h}(\boldsymbol{\theta}, \sigma) = \mathbf{0}$ in a neighbor of $(\boldsymbol{\theta}_0, 0)$, *i.e.*, from eq. (11), for $\sigma < \varepsilon$, $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} = \boldsymbol{\varphi}(\sigma)$. The first order expansion of $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ in the neighbor of $\boldsymbol{\theta}_0$ writes thus (almost surely) $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} = \boldsymbol{\varphi}(\sigma) = \boldsymbol{\varphi}(0) + \sigma \frac{\partial \boldsymbol{\varphi}}{\partial \sigma}(0) + \mathbf{o}(\sigma)$ where $\mathbf{o}(\sigma)$ is a small-o or negligible remainder with respect to σ when $\sigma \to 0$. Now, from $\boldsymbol{\varphi}(0) = \boldsymbol{\theta}_0$, eqs. (12)-(13) and noting that $\frac{\partial \mathbf{h}}{\partial \sigma}(\boldsymbol{\theta}_0, 0) = 2 \mathbf{J}_{\mathbf{m}}^t(\boldsymbol{\theta}_0) \mathbf{C}^{-1} \boldsymbol{\xi}$, we get (almost surely)

$$\sigma^{-1}(\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} - \boldsymbol{\theta}_0) = -\boldsymbol{\Psi}_{\mathbf{m}} \mathbf{J}_{\mathbf{m}}^t(\boldsymbol{\theta}_0) \mathbf{C}^{-1} \boldsymbol{\xi} + \sigma^{-1} \mathbf{o}(\sigma). \tag{14}$$

The result is thus a consequence of property 2. $\square$

One of the main consequence of proposition 2 is the asymptotic unbiasness and efficiency of the MLE as stated in the next proposition:

Proposition 3. Assume that $\sigma^{-2}|\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} - \boldsymbol{\theta}_0|^2$ is uniformly integrable and that $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$ admit a first and second order moment. Thus, the MLE is unbiased in the asymptotic regime $\sigma \to 0$. However, the CRLB is asymptotically achieved by the MLE iff the noise is Gaussian distributed.

Proof. Under uniformly integrability hypothesis, the convergence in distribution implies the convergence of first and second order moment [27, p.84, thm 6.2]. Therefore, one immediately has $\frac{1}{\sigma} \mathbb{E}[\widehat{\boldsymbol{\theta}}_{\mathrm{ml}} - \boldsymbol{\theta}_0] \xrightarrow[\sigma \to 0]{} 0$ which proves the asymptotic unbiasness of $\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}$.

Then, from eq. (14) and property 1, we have

$$\sigma^{-2} \mathbb{C}\mathrm{ov}[\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}] \xrightarrow[\sigma \to 0]{} n^{-1} \mathbb{E}[r^2] \boldsymbol{\Psi}_{\mathbf{m}} \tag{15}$$

where $r \overset{d}{=} \|\mathbf{C}^{-\frac{1}{2}} \boldsymbol{\xi}\|$ has pdf $p_r(r) = \omega_n r^{n-1} d_n(r^2)$. Moreover, from the expression of the pdf of $\mathbf{x}$ we have $\frac{\partial \log p_{\mathbf{x}}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} = -\frac{2}{\sigma} \mathbf{J}_{\mathbf{m}}^t(\boldsymbol{\theta}_0) C^{-1} \boldsymbol{\xi} \phi_n(\boldsymbol{\xi}^t \mathbf{C}^{-1} \boldsymbol{\xi})$ where $\phi_n(r) = \frac{d_n'(r)}{d_n(r)}$. Hence, from property 1 we immediately obtain the Fisher information $\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial \log p_{\mathbf{x}}}{\partial \boldsymbol{\theta}} \frac{\partial \log p_{\mathbf{x}}}{\partial \boldsymbol{\theta}^t}\right]$ in $\boldsymbol{\theta}_0$ as

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0) = 4 n^{-1} \sigma^{-2} \mathbb{E}[r^2 \phi_n^2(r^2)] \boldsymbol{\Psi}_{\mathbf{m}}^{-1} \tag{16}$$

(see Slepian-Bangs like formula given in [28]). Then, one has $\mathbb{E}^2[r^2 \phi_n(r^2)] \leq \mathbb{E}[r^2] \mathbb{E}[r^2 \phi_n^2(r^2)]$ using Cauchy-Schwarz inequality. From an integration by parts, $\mathbb{E}[r^2 \phi_n(r^2)] = \omega_n \int_{\mathbb{R}_+} r^{n+1} d_n'(r^2) dr = -2n \int_{\mathbb{R}_+} \omega_n r^{n-1} d_n(r^2) dr = -2n$ so that $4n^2 \leq \mathbb{E}[r^2] \mathbb{E}[r^2 \phi_n^2(r^2)]$ (the existence of the first moment insures that the integrated term vanishes). Together with eqs. (15) and (16), the Cramér-Rao inequality $\mathbb{C}\mathrm{ov}[\widehat{\boldsymbol{\theta}}_{\mathrm{ml}}] \geq \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)$ is obviously recovered in the asymptotic regime. But more important, the equality case is reached

iff $r \propto r\,\phi_n\left(r^2\right)$, that means $\phi_n = \frac{d'_n}{d_n}$ is constant leading to d_n being an exponential, *i.e.,* the noise being Gaussian. $\quad\square$

IV. Experimental validation

A. Scenario

Let us consider the problem of spectral analysis which is frequently encountered in signal processing applications for direction of arrival estimation [29]–[32]:

$$x_t = \sin(2\pi f_0 t) + \sigma \xi_t, \quad t = 1, \ldots, n \tag{17}$$

where $f_0 \in \left[-\frac{1}{4}; \frac{1}{4}\right)$ is an unknown normalized frequency and the noise ξ_t is assumed such that the concatenated vector $\boldsymbol{\xi} = [\xi_1 \ \cdots \ \xi_n]^t \sim \mathcal{ED}^+(\mathbf{0}, \mathbf{C}, d_n)$. Matrix $\mathbf{C}$ is assumed known, for instance precalculated with secondary data which only contains noise measurement. In our application, we suppose that $\boldsymbol{\xi}$ is Student-t distributed with $\nu > 0$ degree of freedom, *i.e.,* $d_n(u) \propto \left(1 + \frac{u}{\nu}\right)^{-\frac{n+\nu}{2}}$, which is often consider as a more robust model that the Gaussian one [33] and $\mathbf{m}(f_0) = \sin(2\pi f_0 \mathbf{t})$, $\mathbf{t} = [1 \ \cdots n]^t$ (function of a vector is to be understood term by term). The MLE is thus given by eq. (6) where $\boldsymbol{\theta} = f \in \left[-\frac{1}{4}; \frac{1}{4}\right)$ and thanks to proposition 2 its asymptotic behavior is given by

$$\sigma^{-1}\left(\widehat{f}_{\mathrm{ml}} - f_0\right) \xrightarrow[\sigma \to 0]{\mathrm{d}} \mathcal{ED}^+(0, \psi, d_1) \tag{18}$$

where $\psi = \left(4\pi^2[\mathbf{t} \odot \cos(2\pi f_0 \mathbf{t})]^t \mathbf{C}^{-1}[\mathbf{t} \odot \cos(2\pi f_0 \mathbf{t})]\right)^{-1}$, the operator $\odot$ denoting the Hadamard product (element-wise). The density generator d_1 can be obtained using eq. (3) but the Student-t random vector being a consistent Gaussian scale mixture [20], [24], d_1 is the Student-t density generator with ν degrees of freedom, $d_1(u) \propto \left(1 + \frac{u}{\nu}\right)^{-\frac{1+\nu}{2}}$.

B. Simulation results

In order to illustrate the validity of our proposed asymptotic result, the histogram of MLE estimates is statistically tested with the predicted distribution for the scenario proposed in previous section. Data are generated with the following setup: $f_0 = 0.1$, $n = 16$, $\boldsymbol{\xi}$ is randomly generated following a Student-t distribution with $\nu = 5$, $\mathbf{C}$ is chosen randomly among symmetric positive definite matrix (and fixed along the simulations) with $\mathrm{Tr}(\mathbf{C}) = n$ and thus $\mathrm{SNR} = \frac{(\nu-2)}{n\nu\sigma^2}\|\mathbf{m}(f_0)\|^2$. Since frequency estimation is a non-convex problem, the MLE $\widehat{f}_{\mathrm{ml}}$ is computed with a Gauss-Newton algorithm initialized searching the minimum of the quadratic form (6) on a grid $\left[-\frac{1}{4}; \frac{1}{4}\right)$ with a step $\delta_f = 10^{-4}$.

In figure 1, the normalized histogram of $\widehat{f}_{\mathrm{ml}}$ is compared with Gaussian and Student-t distributions. We note that empirical distribution of $\widehat{f}_{\mathrm{ml}}$ is very close to the asymptotic distribution. In figure 2, we performed the Kolmogorov-Smirnov (KS), the chi-square (χ^2) and the Anderson-Darling (AD) tests [34]. These tests show that the difference is not significant to reject the hypothesis of Student-t distribution when the SNR is upper than 5 dB, -1 dB and 11 dB respectively.

The last figure, fig. 3, shows the variance and the mean of the MLE and the CRLB wrt SNR. One can remark that both the mean and the variance of the MLE asymptotically achieves

the predicted mean and variance given by prop. 3 and eq. (15). We also remark that a threshold phenomena occurs due to the non-linearity of our model as under Gaussian noise case [2]. Finally, as predicted, the MLE appears to be not efficient.
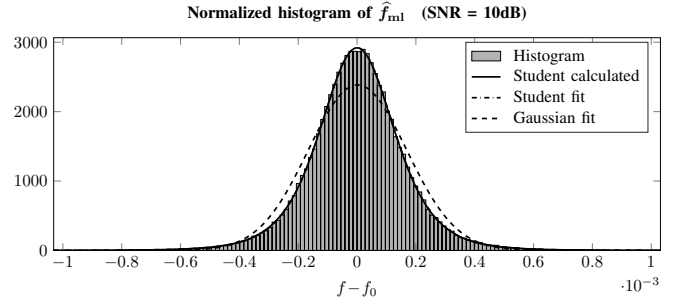

Fig. 1: Comparison between the empirical distribution of $\widehat{f}_{\mathrm{ml}}$ obtained from 10^5 trials of Monte-Carlo simulations, the predicted Student-t using eq. (18) and a fitted Gaussian and Student-t distribution computed by ML method.
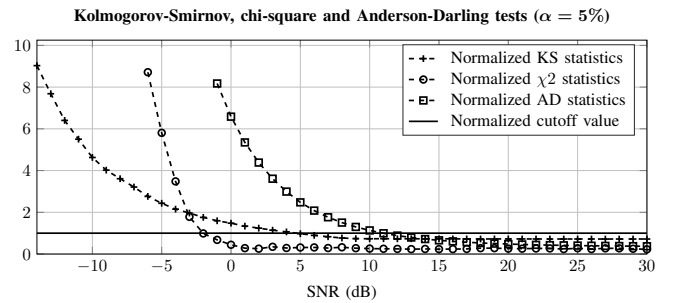

Fig. 2: Comparison of KS, χ^2 (with 2500 bins) and AD tests between the empirical distribution of $\widehat{f}_{\mathrm{ml}}$ and the asymptotic Student-t distribution versus SNR. The cutoff value is computed for a level of significance $\alpha = 0.05$. All the tests are normalized by their respective cutoff for comparisons purposes.
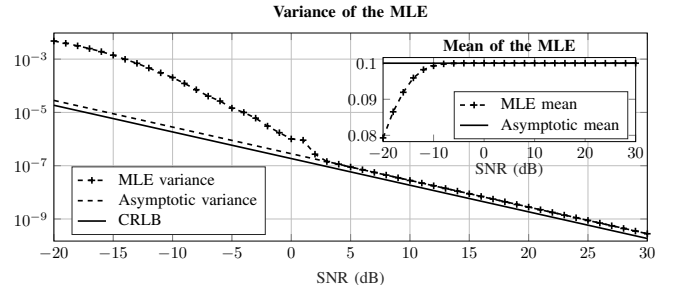

Fig. 3: Comparison between the empirical variance of $\widehat{f}_{\mathrm{ml}}$, the asymptotic variance and the CRLB wrt SNR.

V. Conclusion

In this paper, we have extended the works of Renaux *et al.* [2] and of Ding *et al.* [4] about the characterization of the MLE in high SNR regime when the additive noise is unimodal elliptically distributed but non necessarily Gaussian. In particular, the asymptotic distribution of the ML estimator is derived and shown to be intimately linked to the noise distribution. A close-form expression is provided for the asymptotic distribution of the MLE. It allows us to prove that, the MLE is unfortunately not efficient in the low noise level regime except in the case of Gaussian noise. Finally, the proposed theoretical results are applied to a frequency estimation that frequently arises in the DOA framework, showing a good agreement between the variance of the MLE, its asymptotic prediction and the discrepancy with the CRLB.

References

[1] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer-Verlag, New-York, 2nd edition, 1998.

[2] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal. On the high-SNR conditional maximum-likelihood estimator full statistical characterization. *IEEE Transactions on Signal Processing*, 54(12):4840 – 4843, December 2006.

[3] G. Ginolhac and P. Forster. A procedure for efficient estimation of multiple parameters in some nonlinear cases. *Signal Processing*, 89(4):680 – 684, 2009.

[4] Q. Ding and S. Kay. Maximum likelihood estimator under a misspecified model with high signal-to-noise ratio. *IEEE Transactions on Signal Processing*, 59(8):4012 – 4016, August 2011.

[5] H. Teicher. On the mixture of distributions. *Annals of Mathematical Statistics*, 31(1):55–73, March 1960.

[6] D. Kelker. Infinite divisibility and variance mixture of normal distributions. *The Annals of Mathematical Statistics*, 42(2):802–808, April 1971.

[7] K. Yao. A representation theorem and its applications to spherically-invariant random processes. *IEEE Transactions on Information Theory*, 19(5):600–608, September 1973.

[8] D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, 36(1):99–102, 1974.

[9] M. Rangaswamy, D. Weiner, and A. Öztürk. Non-Gaussian random vector identification using spherically invariant random processes. *IEEE Transactions on Aerospace and Electronics Systems*, 26(1):111–124, January 1993.

[10] M. Rangaswamy, D. Weiner, and A. Öztürk. Computer generation of correlated non-Gaussian radar clutter. *IEEE Transactions on Aerospace and Electronics Systems*, 31(1):106–1116, January 1995.

[11] L. Bombrun and J.-M. Beaulieu. Fisher distribution for texture modeling of polarimetric SAR data. *IEEE Geoscience and Remote Sensing Letters*, 5(3):512–516, July 2008.

[12] J. Portilla, V. Strela, J. Wainwright, and R. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, November 2003.

[13] F. Shi and I. W. Selesnick. An elliptically contoured exponential mixture model for wavelet based image denoising. *Applied and Computational Harmonic Analysis*, 23(1):131–151, July 2007.

[14] K. T. Fang, S. Kotz, and K. W. Ng. *Symmetric multivariate and related distributions*. Monographs on statistics and probability 36. Chapman & Hall, London, 1990.

[15] K.-C. Chu. Estimation and decision for linear systems with elliptical random processes. *IEEE Transactions on Automatic Control*, 18(5):499–505, October 1973.

[16] J. Goldman. Detection in the presence of spherically symmetric random vectors. *IEEE Transactions on Information Theory*, 22(1):52–59, January 1976.

[17] S. Bausson, F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal. First- and second-order moments of the normalized sample covariance matrix of spherically invariant random vectors. *IEEE Signal Processing Letters*, 14(6):425–428, June 2007.

[18] K. Todros and J. Tabrikian. Blind separation of independent sources using Gaussian mixture model. *IEEE Transactions on Signal Processing*, 55(7):3645–3658, July 2007.

[19] Y. Chitour and F. Pascal. Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(10):4563–4573, October 2008.

[20] S. Zozor and C. Vignat. Some results on the denoising problem in the elliptically distributed context. *IEEE Transactions on Signal Processing*, 58(1):134–150, January 2010.

[21] S. Zozor. *Bruit, Non-linéaire et Information : quelques résultats*. Habilitation à Diriger des Recherches, Institut National Polytechnique de Grenoble, Grenoble, France, June 2012.

[22] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes. Stochastic Models with infinite Variance*. Chapman & Hall, New-York, 1994.

[23] I. J. Schœnberg. Metric spaces and completely monotone functions. *Annals of Mathematics*, 39(4):811–841, October 1938.

[24] Y. Kano. Consistency property of elliptic probability density functions. *Journal of Multivariate Analysis*, 51(1):139–147, October 1994.

[25] N. Mukhopadhyay. *Probability and Statistical Inference*, volume 162 of *"Statistics: textbooks and monographs"*. Marcel Dekker, New-York, 5th edition, 2000.

[26] J. J. Duistermaat and J. A. C. Kolk. *Multidimensional Real Analysis I: Differentiation*. Cambridge University Press, New-York, 2004.

[27] A. DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer-Verlag, New-York, 2008.

[28] O. Besson and Y. I. Abramovich. On the Fisher information matrix for multivariate elliptically contoured distributions. *IEEE Signal Processing Letters*, 20(11):1130–1133, November 2013.

[29] D.C. Rife and R. R. Boorstyn. Multiple tone parameter estimation from discrete-time observations. *The Bell System Technical Journal*, 55(9):1389–1410, November 1976.

[30] S. Kay. A fast and accurate single frequency estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):1987–1990, December 1989.

[31] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10):1783–1795, October 1990.

[32] D. Starer and A. Nehorai. Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise. *IEEE Transactions on Signal Processing*, 40(6):1528–1534, June 1992.

[33] V. Ollier, M. N. E. Korso, R. Boyer, P. Larzabal, and M. Pesavento. Robust calibration of radio interferometers in non-gaussian environment. *IEEE Transactions on Signal Processing*, 65(21):5649–5660, Nov 2017.

[34] R.B. d'Agostino and M. A. Stephen, editors. *Goodness-of-Fit Techniques*. Marcel Dekker, Inc, New-York, 1986.