ROBUST SEMIPARAMETRIC JOINT ESTIMATORS OF LOCATION AND SCATTER IN ELLIPTICAL DISTRIBUTIONS

Stefano Fortunati, Alexandre Renaux, Frédéric Pascal

Université Paris-Saclay, CNRS, CentraleSupeléc, Laboratoire des signaux et systèmes,
91190, Gif-sur-Yvette, France.
e-mails: {stefano.fortunati, alexandre.renaux, frederic.pascal}@centralesupelec.fr

ABSTRACT

This paper focuses on the joint estimation of the location vector and the shape matrix of a set of Complex Elliptically Symmetric (CES) distributed observations. This well-known estimation problem is framed in the original context of semiparametric models allowing us to handle the (generally unknown) density generator as an *infinite-dimensional* nuisance parameter. A joint estimator, relying on the Tyler's M-estimator of location and on a new R-estimator of shape matrix, is proposed and its Mean Squared Error (MSE) performance compared with the Semiparametric Cramér-Rao Bound (CSCRB).

Index Terms— Elliptical distribution, robust estimator, covariance estimation, semiparametric model, efficiency.

1. INTRODUCTION

Among the wide range of statistical tools exploited in Machine Learning (ML), covariance matrix estimators are certainly of primary importance [1]. In fact, popular ML inference methods, e.g. image segmentation, clustering, dimension reduction and distance learning, rely on covariance matrix estimation as a first preliminary step. The concept of covariance matrix implies the existence of a statistical model able to describe the data behaviour and consequently, the possibility to derive inference algorithms based on it. A common approach in many ML inference problems is the *model-based* one that consists in considering the available data as sampled from a known probability density function (pdf) with possibly unknown fixed parameters. More formally, let $\{\mathbf{z}_l\}_{l=1}^L$ be a set of L independent and identically distributed (i.i.d.) N-dimensional observations sharing the same pdf, i.e. $\mathbb{C}^N \ni \mathbf{z}_l \sim p_Z, \; \forall l$. In model-based inference, p_Z is assumed to belong to a parametric model $\mathcal{P}_{\boldsymbol{\theta}} \triangleq \{p_Z | p_Z(\mathbf{z}_l; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ that represents a family of pdfs parametrized by a finite dimensional vector $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{C}^q$. In multivariate Gaussian-based inference, for example, $\boldsymbol{\theta}$ is set up by the mean vector $\boldsymbol{\mu}$ and by the vectorized version of the covariance/scatter matrix $\text{vec}(\boldsymbol{\Sigma})$,[1] i.e. $\boldsymbol{\theta} \triangleq (\boldsymbol{\mu}^T, \boldsymbol{\mu}^H, \text{vec}(\boldsymbol{\Sigma})^T)^T$. Note that, since we are considering the general case of complex parameter vectors, the use of *Wirtinger calculus* is required [2]. This is the reason why, in $\boldsymbol{\theta}$ the entries of the mean vector and of the (Hermitian) covariance/scatter matrix have to be listed together with theirs complex conjugates.

Due to its simplicity, the Gaussian assumption has been of mainstream use for decades. However, extensive statistical analysis, performed in a large number of applications, have shown the ubiquitous non-Gaussian nature of the data.

Different statistical models, aiming at combining the "conciseness" of the Gaussian one with the heavy-tailed behaviour of the observed data, have been developed in statistic and ML literature. Among others, a non-Gaussian data model, that has been recognized to provide a reliable characterization of the data behaviour in many applications [3, 4], is the set of Complex Elliptically Symmetric (CES) distributions [5]. Each pdf p_Z belonging to the CES model is parametrized by a *finite-dimensional* parameter vector of interest $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{C}^q$ and by an *infinite-dimensional* parameter, usually called density generator, $\mathcal{G} \ni h : \mathbb{R}^+ \to \mathbb{R}^+$ that characterizes the data non-Gaussianity and that is generally unknown. In particular, as in the Gaussian case, $\boldsymbol{\theta}$ accounts for a location parameter $\boldsymbol{\mu} \in \mathbb{C}^N$ and the covariance structure of the data. Due to the well-known scale ambiguity affecting the CES model, the covariance/scatter matrix $\boldsymbol{\Sigma}$ is not identifiable and only scaled versions, usually called *shape matrices*, $\mathbf{V} \triangleq \boldsymbol{\Sigma}/s(\boldsymbol{\Sigma})$ can be estimated [5]. The rigorous definition of the assumptions that a scale functional $s(\cdot)$ has to satisfy can be found in [6–8]. Here, following [8], among all the possible $s(\cdot)$, we choose the one that forces the shape matrix to have the first top-left entry equal to 1, i.e. $s(\boldsymbol{\Sigma}) = [\boldsymbol{\Sigma}]_{1,1}$. According to the notation introduced in [8], the resulting (Hermitian and positive-definite) shape matrix will be indicated as $\mathbf{V}_1 \triangleq \boldsymbol{\Sigma}/[\boldsymbol{\Sigma}]_{1,1}$. By relying again on the Wirtinger calculus, the finite dimensional parameter vector $\boldsymbol{\theta}$

---

[1]The standard vectorization operator vec maps column-wise the entry of an $N \times N$ matrix $\mathbf{A}$ in an N^2-dimensional column vector $\text{vec}(\mathbf{A})$.

that uniquely parametrizes the CES model, is given by: [2]

$$\boldsymbol{\theta} \triangleq (\boldsymbol{\mu}^T, \boldsymbol{\mu}^H, \underline{\text{vec}}(\mathbf{V}_1)^T)^T \in \Theta \subseteq \mathbb{C}^q, \qquad (1)$$

where $q = N(N+2) - 1 \; (= 2N + N^2 - 1)$. Note that the term "$2N$" is due to the location $\boldsymbol{\mu} \in \mathbb{C}^N$ and to its conjugate, the term "N^2" represents the total number of entries of the shape matrix $\mathbf{V}_1$, while the "-1" is due to the constraint on its first top-left element. The infinite-dimensional parameter h, characterizing the CES pdf p_Z, belongs to the set $\mathcal{G} = \left\{ h : \mathbb{R}^+ \to \mathbb{R}^+ \,\middle|\, \int_0^\infty t^{N-1} h(t) dt < \infty, \int p_Z = 1 \right\}$ [5].

Due to the "mixed" (both finite and infinite) dimensionality of its parameter space, the CES model can then be framed in the context of *semiparametric models* [9]. In particular, as already shown in [6, 7] and in [10, 11], the CES model is the set of pdf $p_Z \equiv CES_N(\boldsymbol{\mu}, \mathbf{V}_1, h)$ such that:

$$\mathcal{P}_{\boldsymbol{\theta}, h} = \left\{ p_Z | p_Z(\mathbf{z}|\boldsymbol{\theta}, h) = |\mathbf{V}_1|^{-1} \times \right. \\ \left. h\left((\mathbf{z} - \boldsymbol{\mu})^H \mathbf{V}_1^{-1}(\mathbf{z} - \boldsymbol{\mu})\right); \boldsymbol{\theta} \in \Theta, h \in \mathcal{G} \right\}, \qquad (2)$$

Inference problems in CES models then involve the estimation of $\boldsymbol{\theta} \in \Theta$ in the presence of a functional *nuisance* parameter represented by the density generator $h \in \mathcal{G}$.

Previous papers from both statistic [7] and signal processing communities [8], have already investigated the possibility to derive *robust* and *semiparametric efficient* estimators for the shape matrix $\mathbf{V}_1$. Specifically, in the recent work [8], a theoretical and simulative analysis of the "finite-sample" performance of a new R-estimator of $\mathbf{V}_1$ is provided. Our aim here is then to complete the analysis already developed in [8] by investigating the joint estimation problem of the location parameter $\boldsymbol{\mu}$ and the shape matrix $\mathbf{V}_1$ in the presence of an unknown density generator h.

Algebraic notation: For the sake of clarity, in the rest of this paper, we adopt the same notation already introduced in [8]. In addition to the list of symbols detailed in [8], we will make extensive use of some specific matrices whose definitions are collected below. In particular:

$$\mathbf{L}_{\mathbf{V}_1} \triangleq \mathbf{P} \left(\mathbf{V}_1^{-T/2} \otimes \mathbf{V}_1^{-1/2} \right) \Pi_{\text{vec}(\mathbf{I}_N)}^\perp, \qquad (3)$$

where $\mathbf{V}_1$ is the shape matrix previously introduced, $\mathbf{P} \triangleq [\mathbf{e}_2 | \mathbf{e}_3 | \cdots | \mathbf{e}_{N^2}]^T$ where $\mathbf{e}_i$ is the i-th vector of the canonical basis of $\mathbb{R}^{N^2}$ and $\Pi_{\text{vec}(\mathbf{I}_N)}^\perp \triangleq \mathbf{I}_{N^2} - N^{-1}\text{vec}(\mathbf{I}_N)\text{vec}(\mathbf{I}_N)^T$.

CES-related notation: CES distributions have been the subject of extensive theoretical and applicative-oriented research papers and tutorial presentations. Here, we will collected the very basic facts and notation that will be used in the rest of this work. Let $\boldsymbol{\theta}_0 \triangleq (\boldsymbol{\mu}_0^T, \boldsymbol{\mu}_0^H, \text{vec}(\mathbf{V}_{1,0})^T)^T$ be the "true" parameter vector to be estimated and let h_0 be the actual (and unknown) density generator. Let $\mathbb{C}^N \ni \mathbf{z} \sim p_0(\mathbf{z}) \equiv p_Z(\mathbf{z}; \boldsymbol{\theta}_0, h_0) \equiv CES_N(\boldsymbol{\mu}_0, \mathbf{V}_{1,0}, h_0)$ a CES-distributed random vector characterized by a location vector

$\boldsymbol{\mu}_0$, a shape matrix $\mathbf{V}_{1,0} \triangleq \boldsymbol{\Sigma}_0/[\boldsymbol{\Sigma}_0]_{1,1}$ where $\boldsymbol{\Sigma}_0$ represents the relevant scatter matrix and a density generator $h_0 \in \mathcal{G}$. Then, $\mathbf{z}$ satisfies the following stochastic representation:

$$\mathbf{z} =_d \boldsymbol{\mu}_0 + \sqrt{\mathcal{Q}} \boldsymbol{\Sigma}_0^{1/2} \mathbf{u}, \qquad (4)$$

where $\mathbf{u} \sim \mathcal{U}(\mathbb{C}S^N)$ is a complex random vector uniformly distributed on the unit N-sphere and $=_d$ stands for "has the same distribution as". The *2nd-order modular variate* $\mathcal{Q}$ is independent from $\mathbf{u}$ and such that:

$$\mathcal{Q} =_d (\mathbf{z} - \boldsymbol{\mu}_0)^H \boldsymbol{\Sigma}_0^{-1} (\mathbf{z} - \boldsymbol{\mu}_0) \triangleq Q, \qquad (5)$$

Moreover, $\mathcal{Q}$ is distributed according to the following pdf:

$$p_{\mathcal{Q},0}(q) = \pi^N \Gamma(N)^{-1} q^{N-1} h_0(q), \qquad (6)$$

where $\Gamma(\cdot)$ is the Gamma function. For any $h \in \mathcal{G}$, the function ψ is defined as

$$\psi(t) \triangleq d \ln h(t)/dt. \qquad (7)$$

Finally, the expectation operator of any *measurable* function f with respect to $p_0(\mathbf{z})$ is indicated as $E_0\{f(\mathbf{z})\} \triangleq \int f(\mathbf{z}) p_Z(\mathbf{z}; \boldsymbol{\theta}_0, h_0) d\mathbf{z}$.

2. ON THE SEMIPARAMETRIC JOINT ESTIMATION OF LOCATION AND SHAPE

Let $\{\mathbf{z}_l\}_{l=1}^L$ be a set of CES distributed vectors such that $\mathbb{C}^N \ni \mathbf{z}_l \sim p_0 \equiv CES_N(\boldsymbol{\mu}_0, \mathbf{V}_{1,0}, h_0), \forall l$. Our goal then is to investigate the problem of jointly estimating $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$ in the presence of an unknown density generator h_0.

The two fundamental questions that we are going to address in this section are the following:

1. What is the impact of not knowing h_0 on the joint estimation of $(\boldsymbol{\mu}_0, \mathbf{V}_{1,0})$?

2. Which is the (asymptotic) impact that the lack of knowledge of $\boldsymbol{\mu}_0$ has on the estimation of $\mathbf{V}_{1,0}$ and vice versa?

To answer these two questions, we need to introduce the *semiparametric efficient score vector* $\bar{\mathbf{s}}_{\boldsymbol{\theta}_0}$ and the *semiparametric Fisher Information Martix* (SFIM) $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)$. As discussed in the relevant statistical literature for a generic semiparametric model [9] and recently investigated for the specific CES model [10, 11], the semiparametric efficient score vector for the joint estimation of location and shape matrix of a set of CES distributed data is given by:

$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_0} = [\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}^T, \bar{\mathbf{s}}_{\boldsymbol{\mu}_0^*}^T, \bar{\mathbf{s}}_{\text{vec}(\mathbf{V}_{1,0})}^T]^T = \mathbf{s}_{\boldsymbol{\theta}_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{h_0}), \qquad (8)$$

where $\mathbf{s}_{\boldsymbol{\theta}_0}$ is the "classical" score vector defined, by means of the Wirtinger derivatives, as [11]:

$$[\mathbf{s}_{\boldsymbol{\theta}_0}]_i \triangleq \partial \ln p_Z(\mathbf{z}; \boldsymbol{\theta}, h_0)/\partial \theta_i^* \big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0}, \; i = 1, \ldots, q \qquad (9)$$

---
[2]The operator $\underline{\text{vec}}(\mathbf{A})$ defines the $N^2 - 1$-dimensional vector obtained from vec $(\mathbf{A})$ by deleting its first element, i.e. vec $(\mathbf{A}) \triangleq [a_{11}, \underline{\text{vec}}(\mathbf{A})^T]^T$.

and $\boldsymbol{\theta}$ is given in (1) and $q = N(N+2) - 1$. The term $\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{h_0})$ indicates the orthogonal projection of $\mathbf{s}_{\boldsymbol{\theta}_0}$ on the nuisance tangent space $\mathcal{T}_{h_0}$ of the CES model $\mathcal{P}_{\boldsymbol{\theta},h}$ in (2) evaluated at the true density generator h_0. Specifically, $\Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{h_0})$ tells us the loss of information on the estimation of $\boldsymbol{\theta}_0$ due to the lack of knowledge of h_0. As shown in [11, eq. (23)], the orthogonal projection operator of $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}^T$ and $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0^*}^T$ onto $\mathcal{T}_{h_0}$ is equal to zero and consequently, the lack of knowledge of h_0 does not have any impact on the (asymptotic) estimation of the location parameter $\boldsymbol{\mu}_0$. On the other hand, not knowing h_0 does have an impact on the estimation of the shape matrix $\mathbf{V}_{1,0}$, since as proved in [11, eq. (24)], $\Pi(\mathbf{s}_{\underline{\mathrm{vec}}(\mathbf{V}_{1,0})}|\mathcal{T}_{h_0}) \neq \mathbf{0}$. This answers the first question.

To address the second point about the (asymptotic) cross-information between $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$, we need to check the structure of the SFIM $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)$. The SFIM for the joint estimation of $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$ in the CES semiparametric model $\mathcal{P}_{\boldsymbol{\theta},h}$ in (2) has been evaluated in [10, 11] as:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0) \triangleq E_0\{\bar{\mathbf{s}}_{\boldsymbol{\theta}_0}(\mathbf{z}_l)\bar{\mathbf{s}}_{\boldsymbol{\theta}_0}^H(\mathbf{z}_l)\}$$
$$= \begin{pmatrix} \bar{\mathbf{I}}(\boldsymbol{\mu}_0|h_0) & \mathbf{0}_{2N \times (N^2-1)} \\ \mathbf{0}_{(N^2-1) \times 2N} & \bar{\mathbf{I}}(\mathbf{V}_{1,0}|h_0) \end{pmatrix}, \quad (10)$$

$$\bar{\mathbf{I}}(\boldsymbol{\mu}_0|h_0) = \frac{E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\}}{N} \begin{pmatrix} \boldsymbol{\Sigma}_0^{-1} & \mathbf{0}_{N \times N} \\ \mathbf{0}_{N \times N} & \boldsymbol{\Sigma}_0^{-*} \end{pmatrix}, \quad (11)$$

$$\bar{\mathbf{I}}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0) = \frac{E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}}{N(N+1)} \mathbf{L}_{\mathbf{V}_{1,0}}\mathbf{L}_{\mathbf{V}_{1,0}}^H, \quad (12)$$

where $\mathbf{L}_{\mathbf{V}_{1,0}}$ is defined in (3) and the function ψ_0 is given in (7). As we can clearly see from (10), the efficient SFIM $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)$ is a block-diagonal matrix. This implies that the cross-information terms between the location $\boldsymbol{\mu}_0$ and the shape matrix $\mathbf{V}_{1,0}$ are equal to zero. Consequently, the relevant estimation problems are (asymptotically) decorrelated and can be considered as two separate estimation problems. In particular, in estimating the shape matrix $\mathbf{V}_{1,0}$, the true (and generally unknown) location vector $\boldsymbol{\mu}_0$ can be substituted by any of its $\sqrt{L}$-consistent estimators without any impact on the (asymptotic) performance of the estimator of $\mathbf{V}_{1,0}$. Of course, the vice versa holds true as well, i.e. any $\sqrt{L}$-consistent estimator of $\mathbf{V}_{1,0}$ can be used in place of the true shape matrix without any (asymptotic) impact on the estimation of $\boldsymbol{\mu}_0$. In the next Section, we exploit this result to implement a *robust, semiparametric efficient* joint estimator for the location and shape matrix in CES distributed data.

3. A ROBUST SEMIPARAMETRIC EFFICIENT JOINT ESTIMATOR OF LOCATION AND SHAPE

Robust estimation of location and shape in elliptical distributions is a well-known topic in statistics and signal processing since the seminal paper of Maronna [12]. Let us first recall the framework we are interested in. As before, let $\{\mathbf{z}_l\}_{l=1}^L$ be a set of CES-distributed vectors such that $\mathbb{C}^N \ni \mathbf{z}_l \sim p_0 \equiv$

$CES_N(\boldsymbol{\mu}_0, \mathbf{V}_{1,0}, h_0), \forall l$. In [12], a general class of joint M-estimators of $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$ (in the presence of an unknown density generator h_0) has been introduced as the "fixed-point" solution of the following system of equations:

$$\sum_{l=1}^L u_1(\hat{Q}_l^{1/2})(\mathbf{z}_l - \hat{\boldsymbol{\mu}}) = \mathbf{0}, \quad (13)$$

$$L^{-1}\sum_{l=1}^L u_2(\hat{Q}_l)(\mathbf{z}_l - \hat{\boldsymbol{\mu}})(\mathbf{z}_l - \hat{\boldsymbol{\mu}})^H = \widehat{\mathbf{V}}_1, \quad (14)$$

where, according to (5), $\hat{Q}_l = (\mathbf{z}_l - \hat{\boldsymbol{\mu}})^H \widehat{\mathbf{V}}_1^{-1}(\mathbf{z}_l - \hat{\boldsymbol{\mu}})$, while the functions u_1 and u_2 have to satisfy a given set of assumptions that guarantees the existence and the uniqueness of the solution of (13) and (14) (see [12] for the real case and [5] for the extension to the complex one).

3.1. Tyler's joint M-estimator of $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$

Among different possible choices for u_1 and u_2, Tyler in [13] (see also [3] and [4]) showed that the functions $u_1(Q) = Q^{-1/2}$ and $u_2(Q) = NQ^{-1}$ lead to the "minimax robust" M-estimator of the location and shape. Specifically, by defining

$$\hat{Q}_l^{(k)} = (\mathbf{z}_l - \hat{\boldsymbol{\mu}}^{(k)})^H[\widehat{\mathbf{V}}_1^{(k)}]^{-1}(\mathbf{z}_l - \hat{\boldsymbol{\mu}}^{(k)}), \quad (15)$$

where k indicates the iteration number, we have that the Tyler's joint M-estimator of location and shape, i.e. $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,Ty})$, can be obtained as the convergence points $(k \to \infty)$ of the following iterations:

$$\hat{\boldsymbol{\mu}}_{Ty}^{(k+1)} = \left[\sum_{l=1}^L [\hat{Q}_l^{(k)}]^{-1/2}\right]^{-1} \sum_{l=1}^L \left(\hat{Q}_l^{(k)}\right)^{-1/2} \mathbf{z}_l, \quad (16)$$

$$\begin{cases} \widehat{\mathbf{V}}_{Ty}^{(k+1)} = \frac{N}{L}\sum_{l=1}^L [\hat{Q}_l^{(k)}]^{-1}(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}^{(k)})(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}^{(k)})^H. \\ \widehat{\mathbf{V}}_{1,Ty}^{(k+1)} \triangleq \widehat{\mathbf{V}}_{Ty}^{(k+1)}/[\widehat{\mathbf{V}}_{Ty}^{(k+1)}]_{1,1}. \end{cases}$$
$$(17)$$

Note that, even if a formal proof of the joint convergence of (16) and (17) is still an open problem, these iterative algorithm has been shown to provide reliable estimates in most of the scenarios of possible interest in practical applications. We refer to [3] where joint M-estimators of the form (13)-(14) have been exploited in hyperspectral anomaly detection problems and to [4] where joint M-estimators have been derived as part of a general Expectation-Maximization (EM) algorithm for clustering applications.

The estimators $\hat{\boldsymbol{\mu}}_{Ty}$ and $\widehat{\mathbf{V}}_{1,Ty}$ have the remarkable property of being $\sqrt{L}$-consistent under any (unknown) density generator $h \in \mathcal{G}$ [13]. Consistency, however, is only one of the properties that good robust estimators should have. Another important property is the (semiparametric) efficiency.

3.2. The Semiparametric Cramér-Rao Bound (SCRB)

With semiparametric efficiency we indicate the ability of a robust estimator to achieve the Semiparametric Cramér-Rao

Bound (SCRB) [9] as the number of available observations L goes to infinity. The SCRB for the specific estimation problem at hand has been recently derived in [10,11] as the inverse of the SFIM in (10). As discussed before, since the efficient score vectors for the location, i.e. $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}^T$ and $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0^*}^T$, are orthogonal to the nuisance tangent space $\mathcal{T}_{h_0}$, the SCRB on the estimation of $\boldsymbol{\mu}_0$ is equal to the "classical" CRB and it is given by

$$
\begin{aligned}
\mathrm{CRB}(\boldsymbol{\mu}_0|h_0) &\triangleq \bar{\mathbf{I}}(\boldsymbol{\mu}_0|h_0)^{-1} \\
&= \frac{N}{E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\}} \begin{pmatrix} \boldsymbol{\Sigma}_0 & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \boldsymbol{\Sigma}_0^* \end{pmatrix}.
\end{aligned} \quad (18)
$$

On the other hand, since $\Pi(\mathbf{s}_{\underline{\mathrm{vec}}(\mathbf{V}_{1,0})}|\mathcal{T}_{h_0}) \neq \mathbf{0}$ [11, eq. (24)], the SCRB on the estimation of the shape matrix $\mathbf{V}_{1,0}$ is tighter than the "classical" CRB (that is obtained for a perfectly known h_0) and is given by:

$$
\begin{aligned}
\mathrm{SCRB}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0) &\triangleq \bar{\mathbf{I}}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0)^{-1} \\
&= \frac{N(N+1)}{E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}} \left[\mathbf{L}_{\mathbf{V}_{1,0}}\mathbf{L}_{\mathbf{V}_{1,0}}^H\right]^{-1}.
\end{aligned} \quad (19)
$$

In [10,11], it has been shown that robust M-estimators of the shape matrix are not semiparametric efficient. This efficiency issue can be overcome by exploiting the results proposed by Hallin, Oja and Paindaveine in [7] for the real case and extended to the complex case in [8]. By combining the Le Cam's theory on "one-step" estimators and the robustness property of rank-based inference procedure, in [7] a robust semiparametric efficient R-estimator of $\mathbf{V}_{1,0}$ has been derived and its "finite-sample" performance investigated in [8] for the case of *zero-mean* CES observations. Here, an extension of the previously derived R-estimator of shape matrix is provided to CES data with unknown mean, while Sec. 4 will focus on a simulative analysis of its estimation performance. Note that, due to the orthogonality between the efficient score vectors of $\boldsymbol{\mu}_0$ and the nuisance tangent space $\mathcal{T}_{h_0}$, the estimation of the location parameter does not require any "one-step" correction *à la* Le Cam.

3.3. An R-estimator of $\mathbf{V}_{1,0}$

In their seminal works [7] (see [8] for a tutorial discussion and for the complex extension of these results), Hallin, Oja and Paindaveine showed that a robust and semiparametric efficient estimator of the shape matrix can be obtained by applying a linear, "one-step", correction to any $\sqrt{L}$-consistent preliminary estimator $\widehat{\mathbf{V}}_1^\star$ of $\mathbf{V}_{1,0}$, obtained from the centred CES dataset $\{\mathbf{z}_l - \hat{\boldsymbol{\mu}}^\star\}_{l=1}^L$, where $\hat{\boldsymbol{\mu}}^\star$ is any $\sqrt{L}$-consistent preliminary estimator of the location $\boldsymbol{\mu}_0$. Due to their properties of minimax robustness and $\sqrt{L}$-consistency under any density generator $h \in \mathcal{G}$, the Tyler's estimators previously introduced may be good candidates for this role, i.e. $\hat{\boldsymbol{\mu}}^\star \equiv \hat{\boldsymbol{\mu}}_{Ty}$ and $\widehat{\mathbf{V}}_1^\star \equiv \widehat{\mathbf{V}}_{1,Ty}$. Consequently, following the results obtained in [8], an R-estimator of $\mathbf{V}_{1,0}$ is given by: [3]

$$
\begin{aligned}
\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) &= \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,Ty}) + \frac{1}{L\hat{\alpha}} \left[\mathbf{L}_{\widehat{\mathbf{V}}_{1,Ty}}\mathbf{L}_{\widehat{\mathbf{V}}_{1,Ty}}^H\right]^{-1} \\
&\times \mathbf{L}_{\widehat{\mathbf{V}}_{1,Ty}} \sum_{l=1}^L K_{\mathrm{vdW}}\left(\frac{r_l^\star}{L+1}\right) \mathrm{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H),
\end{aligned} \quad (20)
$$

where $\{r_l^\star\}_{l=1}^L$ are the ranks[4] of the continuous random variables $\{\hat{Q}_l^\star\}_{l=1}^L$ such that:

$$
\hat{Q}_l^\star \triangleq (\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty})^H[\widehat{\mathbf{V}}_{1,Ty}]^{-1}(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}), \quad (21)
$$

while $\hat{\mathbf{u}}_l^\star$ are random vectors obtained as:

$$
\hat{\mathbf{u}}_l^\star \triangleq (\hat{Q}_l^\star)^{-1/2}[\widehat{\mathbf{V}}_{1,Ty}]^{-1/2}(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}). \quad (22)
$$

Moreover, the data-dependent term $\hat{\alpha}$ is defined in [8, Sec. IV.B], while the *van der Waerden* score $K_{\mathrm{vdW}} : (0,1) \to \mathbb{R}^+$ is a function defined as $K_{\mathrm{vdW}}(u) \triangleq \Phi_G^{-1}(u)$ where Φ_G^{-1} indicates the inverse function of the cdf of a Gamma-distributed random variable with parameters $(N,1)$. It is important to note that the R-estimator in (20) depends only on the ranks $r_l^\star$ and on the vectors $\mathbf{u}_l^\star$ whose statistics are invariant with respect to the true (and generally unknown) data pdf [7,8]. This is not the case for the Maronna's and Tyler's M-estimators that directly rely on the collected data (and consequently on their unknown pdf). Interestingly, the R-estimator in (20) is in that sense more robust than M-estimators. To conclude, the pseudocode for the implementation of the R-estimator in (20) is provided in the following.

4. NUMERICAL RESULTS

In this last Section we finally assess, through numerical simulations, the semiparametric efficiency of the joint estimator $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,R})$, where $\hat{\boldsymbol{\mu}}_{Ty}$ is the Tyler's estimator in (16) of the location vector $\boldsymbol{\mu}_0$, while $\widehat{\mathbf{V}}_{1,R}$ is the R-estimator in (20) of the shape matrix $\mathbf{V}_{1,0}$ exploiting the Tyler's joint estimator $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,Ty})$ as preliminary $\sqrt{L}$-consistent estimator. As basis of comparison, we also report the performance of the joint "sample" estimator $(\hat{\boldsymbol{\mu}}_{SM}, \widehat{\mathbf{V}}_{1,SCM})$, defined as:

$$
\hat{\boldsymbol{\mu}}_{SM} \triangleq L^{-1} \sum_{l=1}^L \mathbf{z}_l, \quad (23)
$$

$$
\begin{cases} \widehat{\boldsymbol{\Sigma}}_{SCM} \triangleq L^{-1} \sum_{l=1}^L (\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{SM})(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{SM})^H \\ \widehat{\mathbf{V}}_{1,SCM} \triangleq \widehat{\boldsymbol{\Sigma}}_{SCM}/[\widehat{\boldsymbol{\Sigma}}_{SCM}]_{1,1}. \end{cases} \quad (24)
$$

We generate the set of non-zero mean data $\{\mathbf{z}_l\}_{l=1}^L$ according to a Generalized Gaussian (GG) distribution [14], such that $\mathbb{C}^N \ni \mathbf{z}_l \sim p_0, \forall l$ where:

$$
p_0(\mathbf{z}_l) = |\boldsymbol{\Sigma}_0|^{-1} h_0\left((\mathbf{z}_l - \boldsymbol{\mu}_0)^H \boldsymbol{\Sigma}_0^{-1}(\mathbf{z}_l - \boldsymbol{\mu}_0)\right), \quad (25)
$$

---

[3]Matlab code at https://github.com/StefanoFor
[4]Let $\Omega \triangleq \{\omega_l\}_{l=1}^L$ be a set of continuous random variables and let $\Omega_o \triangleq \{\omega_{L(1)} < \omega_{L(2)} < \ldots < \omega_{L(L)}\}$ be a set containing the same elements of Ω ordered in an ascending way. Then, the rank r_l of $\omega_l \in \Omega$ is its position index in Ω_o.

Algorithm 1 R-estimator for $\mathbf{V}_{1,0}$

Input: $\mathbf{z}_1, \ldots, \mathbf{z}_L$.
Output: $\widehat{\mathbf{V}}_{1,R}$.
1: Obtain the preliminary Tyler's joint estimators:
   $\hat{\boldsymbol{\mu}}_{Ty} \leftarrow \lim_{k\to\infty} \hat{\boldsymbol{\mu}}_{Ty}^{(k)}$ in (16),
   $\widehat{\mathbf{V}}_{1,Ty} \leftarrow \lim_{k\to\infty} \widehat{\mathbf{V}}_{1,Ty}^{(k+1)}$ in (17),
2: Data centring: $\{\mathbf{z}_l\}_{l=1}^L \leftarrow \{\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}\}_{l=1}^L$,
3: **for** $l = l$ to L **do**
4:    $\hat{Q}_l^\star \leftarrow \mathbf{z}_l^H \widehat{\mathbf{V}}_{1,Ty}^{-1} \mathbf{z}_l$,
5:    $\hat{\mathbf{u}}_l^\star \leftarrow (\hat{Q}_l^\star)^{-1/2} \widehat{\mathbf{V}}_{1,Ty}^{-1/2} \mathbf{z}_l$,
6: **end for**
7: Evaluate the ranks $\{r_1^\star, \ldots, r_L^\star\}$ of $\{\hat{Q}_1^\star, \ldots, \hat{Q}_L^\star\}$,
8: $\mathbf{L}_{\widehat{\mathbf{V}}_{1,Ty}} \leftarrow \mathbf{P}(\widehat{\mathbf{V}}_{1,Ty}^{-T/2} \otimes \widehat{\mathbf{V}}_{1,Ty}^{-1/2})\Pi_{\mathrm{vec}(\mathbf{I}_N)}^\perp$,
9: Evaluate $\hat{\alpha}$ as in [8, Sec. IV.B].
10: $\widehat{\boldsymbol{\Upsilon}} \leftarrow \hat{\alpha}\mathbf{L}_{\widehat{\mathbf{V}}_{1,Ty}} \mathbf{L}_{\widehat{\mathbf{V}}_{1,Ty}}^H$,
11: $\boldsymbol{\Delta}_{\widehat{\mathbf{V}}_{1,Ty}} \leftarrow \mathbf{L}_{\widehat{\mathbf{V}}_{1,Ty}} \sum_{l=1}^L K_{\mathrm{vdW}}\left(\frac{r_l^\star}{L+1}\right) \mathrm{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H)$,
12: $\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) \leftarrow \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,Ty}) + L^{-1/2}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{\Delta}_{\widehat{\mathbf{V}}_{1,Ty}}$,
13: Reshape $\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R})$ as an $N \times N$ matrix,
14: **return** $\widehat{\mathbf{V}}_{1,R}$

while the relevant density generator is given by:

$$h_0(t) \triangleq \frac{s\Gamma(N)b^{-N/s}}{\pi^N\Gamma(N/s)} \exp\left(-\frac{t^s}{b}\right), \, t \in \mathbb{R}^+. \quad (26)$$

We chose the GG distribution to assess the performance of the proposed joint estimator because of its flexibility in characterizing the data "heavy-tailness" with respect to the Gaussian one. In fact, according to the value of the shape parameter $s > 0$, the GG density generator in (26) is able to define a distribution with both heavier tails ($0 < s < 1$) and lighter tails ($s > 1$) compared to the Gaussian one ($s = 1$).

The parameters adopted in our simulations are:

- $\boldsymbol{\Sigma}_0$ is a Toeplitz Hermitian matrix whose first column is given by $[1, \rho, \ldots, \rho^{N-1}]^T$; $\rho = 0.8e^{j2\pi/5}$ and $N = 8$.

- Shape matrix: $\mathbf{V}_{1,0} \triangleq \boldsymbol{\Sigma}_0/[\boldsymbol{\Sigma}_0]_{1,1}$.

- Location vector: $[\boldsymbol{\mu}_0]_n \triangleq 0.5e^{j1\pi/7(n-1)}, n = 1, \ldots, N$.

- Scale parameter: $b = [\sigma_X^2 N\Gamma(N/s)/\Gamma((N+1)/s)]^s$ in (26) and $\sigma_X^2 = E\{\mathcal{Q}\}/N = 4$.

- Numbers of observations: $L = 5N$. This clearly defines a "finite-sample" regime.

The performance assessment will be performed in terms of the following Mean Squared Error (MSE) indices:

$$\varrho_\gamma \triangleq ||E\{(\hat{\boldsymbol{\mu}}_\gamma^a - \boldsymbol{\mu}_0^a)(\hat{\boldsymbol{\mu}}_\gamma^a - \boldsymbol{\mu}_0^a)^H\}||_F, \quad (27)$$

where, for a given $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{x}^a \triangleq (\mathbf{x}^T, \mathbf{x}^H)^T \in \mathbb{C}^{2N}$,

$$\varsigma_\gamma \triangleq ||E\{\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,\gamma} - \mathbf{V}_{1,0})\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,\gamma} - \mathbf{V}_{1,0})^H\}||_F, \quad (28)$$

and γ indicates the relevant estimator at hand. As lower bound indices, we use $\varepsilon_{CRB,\boldsymbol{\mu}_0} \triangleq ||\mathrm{CRB}(\boldsymbol{\mu}_0|h_0)||_F$ and $\varepsilon_{SCRB,\mathbf{V}_{1,0}} \triangleq ||\mathrm{SCRB}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0)||_F$, where $\mathrm{CRB}(\boldsymbol{\mu}_0|h_0)$ is given in (18) and $\mathrm{SCRB}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0)$ in (19).
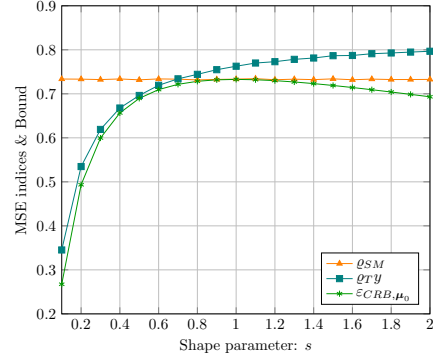

Fig. 1: Estimation of the location vector: MSE indices and CRB vs s ($L = 5N$ and 10^6 Monte Carlo runs).
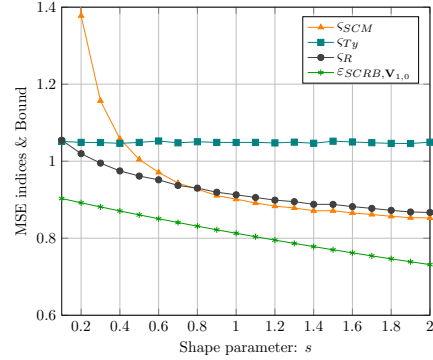

Fig. 2: Estimation of the shape matrix: MSE indices and SCRB vs s ($L = 5N$ and 10^6 Monte Carlo runs).

Fig. 1 shows the performance of the sample mean $\hat{\boldsymbol{\mu}}_{SM}$ estimator in (23) and of the Tyler's estimator $\hat{\boldsymbol{\mu}}_{Ty}$ in (16) compared to the lover bound in (18). As wee can see, $\hat{\boldsymbol{\mu}}_{Ty}$ is almost efficient with respect to $\mathrm{CRB}(\boldsymbol{\mu}_0|h_0)$ in heavy-tailed data ($0 < s < 1$) and outperforms $\hat{\boldsymbol{\mu}}_{SM}$ that it is known to be non robust. On the other hand, $\hat{\boldsymbol{\mu}}_{SM}$ is efficient in the Gaussian case ($s = 1$), and tends to have better performance than $\hat{\boldsymbol{\mu}}_{Ty}$ for $s > 1$. However, in this light-tails scenario, the MSE of $\hat{\boldsymbol{\mu}}_{Ty}$ does not explode and remains close to the $\hat{\boldsymbol{\mu}}_{SM}$'s one.

As far it concern the shape matrix estimation, the simulation results are shown in Fig. 2. The main fact here is that the R-estimator $\widehat{\mathbf{V}}_{1,R}$ in (20) outperforms the Tyler's estimator $\widehat{\mathbf{V}}_{1,Ty}$ in (17) for every values of s, i.e. for both heavy-tailed and light-tailed data. Moreover, as expected, $\widehat{\mathbf{V}}_{1,R}$ greatly outperforms the sample covariance matrix $\widehat{\mathbf{V}}_{1,SCM}$ in (24) in the presence of heavy-tailed data ($0 < s < 1$), while their MSE is of the same order for $s > 1$. Finally, a comment

on the efficiency of the above-mentioned estimator is in order. As we can see, there is a gap between the MSE indices of $\widehat{\mathbf{V}}_{1,SCM}$, $\widehat{\mathbf{V}}_{1,Ty}$ and $\widehat{\mathbf{V}}_{1,R}$. However, it is worth to underline that our aim here is to compare the performance of shape matrix estimators in a "finite-sample" regime, i.e. with a number of observations equal to $L = 5N$ that represents a reasonable value in many practical applications. Of course, by letting $L \to \infty$, it can be shown that $\widehat{\mathbf{V}}_{1,R}$ achieves the bound $\text{SCRB}(\underline{\text{vec}}(\mathbf{V}_{1,0})|h_0)$ in (19) as predicted by theoretical considerations [7, 8].

In summary, previous simulations highlights the benefits that the proposed robust R-estimator can bring. Specifically, it always outperforms the Tyler's estimator in both heavy- and light-tails scenarios. Moreover its estimation performance is way better that the SCM one in heavy-tailed data while it is almost similar in light-tailed scenarios: high gain, very small loss. These very promising results promote the use of the R-estimator to other problems, as the structured shape estimation discussed in [15].

5. CONCLUSIONS

The joint estimation of the location vector $\boldsymbol{\mu}_0$ and the shape matrix $\mathbf{V}_1$ of a set of i.i.d. CES-distributed, multivariate observations has been addressed. Building upon the asymptotic decorrelation of the location and shape estimation problems, a joint estimator that relies on the Tyler's M-estimator $\hat{\boldsymbol{\mu}}_{Ty}$ for $\boldsymbol{\mu}_0$ and on a recently proposed R-estimator $\widehat{\mathbf{V}}_{1,R}$ for $\mathbf{V}_1$ has been discussed and its MSE performance assessed and compared with the relevant Semiparametric Cramér-Rao Bound. Our simulation results, obtained for GG-distributed data, have shown that joint estimator $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,R})$ of location and shape represents a good alternative to the classical Maronna's joint M-estimators. In particular, in terms of shape matrix estimation, the proposed joint estimator outperforms the joint Tyler's estimator in both heavy-tailed and light-tailed data. Future works will investigate the application of the proposed estimator in robust clustering and distance learning problems.

6. REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. Springer, 2007.

[2] R. Remmert, *Theory of Complex Functions*. New York: Springer, 1991.

[3] J. Frontera-Pons, M. A. Veganzones, F. Pascal, and J. Ovarlez, "Hyperspectral anomaly detectors using robust estimators," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 2, pp. 720–731, 2016.

[4] V. Roizman, M. Jonckheere, and F. Pascal, "A flexible EM-like clustering algorithm for noisy data." [Online]. Available: https://arxiv.org/abs/1907.01660

[5] 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. 11, pp. 5597–5625, 2012.

[6] M. Hallin and D. Paindaveine, "Parametric and semi-parametric inference for shape: the role of the scale functional," *Statistics & Decisions*, vol. 24, no. 3, pp. 327–350, 2009.

[7] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. Optimal R-estimation of shape," *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.

[8] S. Fortunati, A. Renaux, and F. Pascal, "Robust semiparametric efficient estimators in elliptical distributions," *Submitted to IEEE Transactions on Signal Processing*. [Online]. Available: https://arxiv.org/abs/2002.02239

[9] P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, 1993.

[10] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "Semiparametric inference and lower bounds for real elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 164–177, Jan 2019.

[11] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "Semiparametric CRB and Slepian-Bangs formulas for complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5352–5364, Oct 2019.

[12] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, 01 1976.

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

[14] F. Pascal, L. Bombrun, J. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized gaussian distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5960–5971, 2013.

[15] B. Mériaux, C. Ren, M. N. E. Korso, A. Breloy, and P. Forster, "Robust estimation of structured scatter matrices in (mis)matched models," *Signal Processing*, vol. 165, pp. 163 – 174, 2019.