# Joint Estimation of Location and Scatter in Complex Elliptically Symmetric Distributions

## A robust semiparametric and computationally efficient $R$-estimator of the shape matrix

**Stefano Fortunati** · **Alexandre Renaux** ·
**Frédéric Pascal**

**Abstract** The joint estimation of the location vector and the shape matrix of a set of independent and identically Complex Elliptically Symmetric (CES) distributed observations is investigated from both the theoretical and computational viewpoints. This joint 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. In the first part of the paper, a computationally efficient and memory saving implementation of the robust and semiparmaetric efficient $R$-estimator for shape matrices is derived. Building upon this result, in the second part, a joint estimator, relying on the Tyler's $M$-estimator of location and on the $R$-estimator of shape matrix, is proposed and its Mean Squared Error (MSE) performance compared with the Semiparametric Cramér-Rao Bound (SCRB).

**Keywords:** Semiparametric models, robust estimators, efficient estimators, covariance matrix, Complex Elliptically Symmetric (CES) distributions.

## 1 Introduction

Inferring the correlation structure of a set of centered data is a key step in many signal processing and machine learning procedures. Among others, radar/sonar

Stefano Fortunati
Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France & DR2I-IPSA, 94200, Ivry sur Seine, France
E-mail: stefano.fortunati@centralesupelec.fr

Alexandre Renaux and Frédéric Pascal
Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France
E-mail: alexandre.renaux@universite-paris-saclay.fr, frederic.pascal@centralesupelec.fr.

detection, image segmentation, dimension reduction, distance learning and clustering rely on the estimation of the covariance/correlation matrix of an acquired data set [2]. Along with the need of an estimation of the covariance matrix, there is another common aspect in all the above-mentioned applications: the non-Gaussian and heavy-tailed nature of the data. As a consequence, popular Gaussian and pseudo-Gaussian inference procedures may present a dramatic performance decay as extensively shown in statistics and signal processing literature (see e.g. [33] and the references therein).

Motivated by a wide range of experimental evidences and measurement campaigns, the (Real or Complex) Elliptically Symmetric (ES) model has been recently adopted to characterize the statistical data behavior. The RES and CES models in fact have been proved to be able to catch the data heavy-tailedness in a large variety of applications such us radar/sonar [23, 29], hyperspectral imaging [8, 17] and image/data classification [28, 30] just to cite a few. From now on, in this paper, we will focus our attention only on complex-valued, CES distributed, datasets. This choice allows us to work in the most general framework, since the obtained results can be easily "brought back" to the real-valued case.

Together with its generality, the second feature that has placed the elliptical model in the spotlight of signal processing and machine learning communities is its "parsimony" in terms of required parameters. In fact, CES model is fully specified by two *finite dimensional* parameters, i.e. a location vector and a covariance/scatter matrix (as the classical Gaussian model) and by an *infinite dimensional* functional parameter, usually called *density generator*, characterizing the data heavy-tailedness. To better understand the role of this infinite dimensional term, let us take a step back to introduce the notion of *semiparametric* model.

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 probability density function (pdf), i.e. $\mathbb{C}^N \ni \mathbf{z}_l \sim p_Z$, $\forall l$. A parametric model $\mathcal{P}_{\boldsymbol{\theta}} \triangleq \{p_Z \,|\, p_Z(\mathbf{z}_l; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ is then defined as a family of pdfs parameterized by a finite dimensional vector $\boldsymbol{\theta} \in \Theta$. As a classical example, in multivariate Gaussian-based inference, $\boldsymbol{\theta}$ is set up by the mean vector $\boldsymbol{\mu}$ and by the covariance/scatter matrix $\boldsymbol{\Sigma}$.

However, a parametric model is generally too "narrow" and it fails to take into account all the actual uncertainty about the data distribution that is generally present in practical scenarios. Semiparametric models have been then introduced to provide with an additional (functional) degree of freedom [1]. A semiparametric model $\mathcal{P}_{\boldsymbol{\theta},h} \triangleq \{p_Z \,|\, p_Z(\mathbf{z}_l; \boldsymbol{\theta}, h), \boldsymbol{\theta} \in \Theta, h \in \mathcal{G}\}$ is a family of pdf parameterized by a finite dimensional vector $\boldsymbol{\theta} \in \Theta$ (as in the classical parametric case) and by a function $h$ belonging to some suitable function space $\mathcal{G}$. Usually, in most applications, $\boldsymbol{\theta} \in \Theta$ is the parameter vector of interest, while $h \in \mathcal{G}$ can be considered as a *nuisance* function that "contains" the missing knowledge of the functional form of the data pdf $p_Z$. Consequently, inference procedures in semiparametric models aim at estimating/testing for $\boldsymbol{\theta} \in \Theta$ in the presence of an unknown function $h \in \mathcal{G}$ whose estimation is not strictly required.

It is now immediate to realize that the CES model can be framed as a semiparametric model [3, 10, 12]. Formally, the pdf $p_Z$ of a CES-distributed random vector $\mathbf{z}_l \in \mathbb{C}^N$ can be expressed as [22]: [1]

$$p_Z(\mathbf{z}_l|\boldsymbol{\mu}, \boldsymbol{\Sigma}, h) = |\boldsymbol{\Sigma}|^{-1} h\left((\mathbf{z}_l - \boldsymbol{\mu})^{\mathsf{H}} \boldsymbol{\Sigma}^{-1}(\mathbf{z}_l - \boldsymbol{\mu})\right), \tag{1}$$

where, as said before, the final dimensional parameter of interest $\boldsymbol{\theta} \in \Theta$ is composed of a location vector $\boldsymbol{\mu}$ and by the scatter matrix $\boldsymbol{\Sigma}$ that represents the correlation structure of the data, while the *nuisance* density generator $h$ belongs to the set

$$\mathcal{G} = \left\{ h : \mathbb{R}_0^+ \to \mathbb{R}^+ \,\Big|\, \int_0^\infty t^{N-1} h(t) dt < \infty, \int p_Z = 1 \right\}. \tag{2}$$

Two considerations are now in order:

- *The identifiability issue*: the scatter matrix $\boldsymbol{\Sigma}$ and the density generator $h$ are not jointly identifiable. Consequently, only scaled versions, usually called *shape matrix*, $\mathbf{V} \triangleq \boldsymbol{\Sigma}/s(\boldsymbol{\Sigma})$ can be estimated [22]. According to our recent work [6], from now on we consider the shape matrix

$$\mathbf{V}_1 \triangleq \boldsymbol{\Sigma}/[\boldsymbol{\Sigma}]_{1,1}, \tag{3}$$

  i.e. the one obtained form the scatter matrix by constraining its first top-left element to be equal to one.
- *Augmented complex representation of $\boldsymbol{\theta}$*: Following the rules of the Wirtinger calculus [13, 15, 27], in order to take into account the complex-value nature of the location vector $\boldsymbol{\mu}$ and of the shape matrix $\mathbf{V}_1$, the finite-dimensional parameter $\boldsymbol{\theta}$ has to be built up as [6]: [2]

$$\boldsymbol{\theta} \triangleq (\boldsymbol{\mu}^{\top}, \boldsymbol{\mu}^{\mathsf{H}}, \underline{\mathrm{vec}}(\mathbf{V}_1)^{\top})^{\top} \in \Theta \subseteq \mathcal{C}^q, \tag{4}$$

  where $q = N(N+2) - 1 \,(= 2N + N^2 - 1)$ and $\mathcal{C}^q$ is a complex-vector space on real field of dimension $q$ [13, 15]. Note that the "$-1$" is due to the fact that we constraint the first top-left element of $\mathbf{V}_1$ to be equal to 1, so it does not have to be estimated.

Building upon the previous considerations, the semiparamentric CES model [3] can be cast as:

$$\mathcal{P}_{\boldsymbol{\theta},h} = \left\{ p_Z | p_Z(\mathbf{z}_l|\boldsymbol{\theta}, h) = |\mathbf{V}_1|^{-1} h\left((\mathbf{z}_l - \boldsymbol{\mu})^{\mathsf{H}} \mathbf{V}_1^{-1}(\mathbf{z}_l - \boldsymbol{\mu})\right); \boldsymbol{\theta} \in \Theta, h \in \mathcal{G} \right\}, \tag{5}$$

---

[1] It is worth highlighting that the "definition" of density generator used here is a slightly different from the one used in [22]. Unlike [22], we include the normalization factor $C_{m,h}$ in $h$, i.e. our density generator is the product $C_{m,h}h(\cdot)$. We do this for a specific reason: the normalization term $C_{m,h}$ depends on $h$ itself. Then, in a semiparametric framework, we need to define the whole product $C_{m,h}h(\cdot)$ as nuisance function and not just $h(\cdot)$.

[2] The operator $\underline{\mathrm{vec}}(\mathbf{A})$ defines the $N^2 - 1$-dimensional vector obtained from $\mathrm{vec}\,(\mathbf{A})$ by deleting its first element, i.e. $\mathrm{vec}\,(\mathbf{A}) \triangleq [a_{11}, \underline{\mathrm{vec}}(\mathbf{A})^T]^T$. In general, in this paper, we always adopt the same notation used in our previous work [6].

Estimating $\boldsymbol{\theta} \in \Theta$ in the presence of different "degrees of uncertainty" on the density generator $h$ is a well-known problem in robust statistics and signal processing. The most popular class of robust estimators for $\boldsymbol{\theta} \in \Theta$ belongs to the family of $M$-estimators [14] and has been firstly proposed by Maronna in [18] and further developed and investigated by Tyler [31]. However, if on one hand the Maronna/Tyler $M$-estimators have the remarkable robustness property, they fail to be *semiparametrically efficient*, i.e their Mean Square Error (MSE) does not achieve the Semiparametric Cramér-Rao Bound [3, 4].

In order to fill this gap, in their seminal work Hallin, Oja and Paindaveine [12] proposed a new class of rank-based $R$-estimators of the shape matrix for a set of centered RES-distributed data able to be both *distributionally robust* and (almost) *semiparametric efficient*. [3] The real-valued $R$-estimator proposed in [12] has been expended to the case of CES-distributed data in our recent work [6] where a theoretical and simulative analysis of its "finite-sample" performance has been provided as well.

Following the trail of [6, 12], the present paper has two main goals:

1. Derive a "computationally efficient" version of the complex-valued shape matrix $R$-estimator proposed in [6],
2. Investigate 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 $h$. [4]

Having a "computationally efficient" implementation of an estimator is of fundamental importance in real-time applications or in high-dimensional data sets. The new version of the $R$-estimator of the shape matrix proposed in Section 2 of this paper is computationally faster and "memory saving" than the previous implementation in [6] making its exploitation possible in a wider range of applications. The paper continues with an exhaustive theoretical investigation of the statistical interrelation underlying the joint semiparametric estimation of the location vector $\boldsymbol{\mu}$ and of the shape matrix $\mathbf{V}_1$ provided in Section 3 while a robust semiparametric (and computationally) efficient joint estimator of location and shape is discussed in Section 4. The numerical analysis of the performance of the proposed joint estimator is presented in Section 5. Finally, some concluding remarks are collected in Section 6.

We conclude this nonproductive Section with three paragraphs summarizing some useful notations and definitions that serve as prerequisites to the comprehension of the material presented in the rest of the paper.

*Algebraic notation*: For the sake of consistency with our previous works, in the rest of this paper, we adopt the same notation already introduced in [6, 7]. In addition to the list of symbols detailed in [6], we will make extensive use of some specific matrices whose definitions are collected below. In particular:

$$\mathbf{P} = [\mathbf{e}_2|\mathbf{e}_3|\cdots|\mathbf{e}_{N^2}]^\top ,\qquad(6)$$

---

[3] The interested reader can find the Matlab and Python code related to our implementation of this $R$-estimator in both RES- and CES- distributed data at `https://github.com/StefanoFor`.

[4] This part has been partially addressed in our related conference paper [7].

where $\mathbf{e}_i$ is the $i$-th vector of the canonical basis of $\mathbb{R}^{N^2}$, the projection matrix

$$\Pi^{\perp}_{\mathrm{vec}(\mathbf{I}_N)} = \mathbf{I}_{N^2} - N^{-1}\mathrm{vec}(\mathbf{I}_N)\mathrm{vec}(\mathbf{I}_N)^{\top}, \tag{7}$$

and

$$\mathbf{L}_{\mathbf{V}_1} = \mathbf{P}\left(\mathbf{V}_1^{-\top/2} \otimes \mathbf{V}_1^{-1/2}\right)\Pi^{\perp}_{\mathrm{vec}(\mathbf{I}_N)}, \tag{8}$$

where $\mathbf{V}_1$ is the shape matrix previously introduced.

*CES-related notation*: Without any claim of completeness, we collect below the basic properties and notation on CES distributed random vectors. We refer the reader to the excellent review paper [22] for additional material. Let $\boldsymbol{\theta}_0 \triangleq (\boldsymbol{\mu}_0^{\top}, \boldsymbol{\mu}_0^{\mathsf{H}}, \mathrm{vec}(\mathbf{V}_{1,0})^{\top})^{\top}$ 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 parameterized 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:

$$\mathbf{z} =_d \boldsymbol{\mu}_0 + \sqrt{\mathcal{Q}}\boldsymbol{\Sigma}_0^{1/2}\mathbf{u}, \tag{9}$$

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 (s.t.):

$$\mathcal{Q} =_d (\mathbf{z} - \boldsymbol{\mu}_0)^H \boldsymbol{\Sigma}_0^{-1}(\mathbf{z} - \boldsymbol{\mu}_0) \triangleq Q, \tag{10}$$

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

$$p_{\mathcal{Q},0}(q) = \pi^N \Gamma(N)^{-1} q^{N-1} h_0(q), \tag{11}$$

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

$$\psi(t) \triangleq d\ln h(t)/dt. \tag{12}$$

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

*Ranks*: The concept of *ranks* of a set of relevant random variables are a useful tool in non-parametric statistics and numerous works can be found on this topic (see [9], [32, Ch. 13] and references therein). Far be it from us to propose a comprehensive overview of the use of ranks in robust statistics, in the following we limit ourselves to introduce their definition since they will play a crucial role in the definition of the $R$-estimator of the shape matrix. Let $\{a_1, a_2, \ldots, a_L\}$ be a set of $L$ continuous i.i.d. random variables with unspecified distribution. Let us rearrange the variables $a_l, l = 1, \ldots, L$ in an ascending order $a_{L(1)} < a_{L(2)} < \cdots < a_{L(L)}$ and, consequently build the vector of *order statistics* as $\mathbf{v}_A \triangleq [a_{L(1)}, a_{L(2)}, \ldots, a_{L(L)}]^{\top}$. Then, the rank $r_l \in \mathbb{N}/\{0\}$ of $a_l$ is the position index of $a_l$ in $\mathbf{v}_A$.

## 2 A computationally efficient implementation of the $R$-estimator for shape matrices

Building upon the the seminal work of Hallin, Oja and Paindaveine [12], in our recent papers [5, 6] a robust and semiparametric efficient $R$-estimator $\widehat{\mathbf{V}}_{1,R}$ for the shape matrix $\mathbf{V}_{1,0}$ of CES distributed data has been proposed and its properties investigated. This $R$-estimator has its roots in the Le Cam's theory of efficient "one-step" estimator [16] and consequently it can be expressed as a linear combination of two terms:

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^\star) + L^{-1/2}\widehat{\boldsymbol{\Upsilon}}_\mathbb{C}^{-1} \widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}^\mathbb{C}, \tag{13}$$

where $\widehat{\mathbf{V}}_1^\star$ is a $\sqrt{L}$-consistent preliminary estimator that provides $\widehat{\mathbf{V}}_{1,R}$ with the consistency property while the linear correction term $L^{-1/2}\widehat{\boldsymbol{\Upsilon}}_\mathbb{C}^{-1} \widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}^\mathbb{C}$ makes $\widehat{\mathbf{V}}_{1,R}$ semiparametric efficient. Let us have a closer look at the two quantities which constitute the linear correction term (all the details can be found in [6]).

The $(N^2 - 1)$-dimensional vector $\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}^\mathbb{C}$ is the "distribution-free" version of the *efficient central sequence* [1] and it can be explicitly expressed as:

$$\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}^\mathbb{C} \triangleq \frac{1}{\sqrt{L}}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \sum_{l=1}^L K_h\left(\frac{r_l^\star}{L+1}\right) \mathrm{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^\mathsf{H}), \tag{14}$$

where $\{r_l^\star\}_{l=1}^L$ are the ranks of the random variables $\{\hat{Q}_l^\star\}_{l=1}^L$ defined as:

$$\hat{Q}_l^\star \triangleq (\mathbf{z}_l - \widehat{\boldsymbol{\mu}}^\star)^\mathsf{H}[\widehat{\mathbf{V}}_1^\star]^{-1}(\mathbf{z}_l - \widehat{\boldsymbol{\mu}}^\star), \tag{15}$$

$\widehat{\boldsymbol{\mu}}^\star$ and $\widehat{\mathbf{V}}_1^\star$ are two $\sqrt{L}$-consistent preliminary estimators [5] of the location vector $\boldsymbol{\mu}_0$ and of the shape matrix $\mathbf{V}_{1,0}$. The random vectors $\{\hat{\mathbf{u}}_l^\star\}_{l=1}^L$ are given by:

$$\hat{\mathbf{u}}_l^\star \triangleq (\hat{Q}_l^\star)^{-1/2}[\widehat{\mathbf{V}}_1^\star]^{-1/2}(\mathbf{z}_l - \widehat{\boldsymbol{\mu}}^\star). \tag{16}$$

The function $K_h : (0,1) \to \mathbb{R}^+$ is the so-called *score function* and it is a key element to guarantee the robustness of $\widehat{\mathbf{V}}_{1,R}$. We refer the reader to [6] for further details on the assumptions that a score function has to satisfy and on how to built it starting from the set of density generators $\mathcal{G}$.

The $(N^2 - 1) \times (N^2 - 1)$ matrix $\widehat{\boldsymbol{\Upsilon}}_\mathbb{C}$ represents the "distribution-free" approximation of the *semiparametric Fisher Information Matrix* (SFIM) and it is given by [6, Eq. (52)]:

$$\widehat{\boldsymbol{\Upsilon}} \triangleq \hat{\alpha}_\mathbb{C}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^\mathsf{H}, \tag{17}$$

where $\hat{\alpha}_\mathbb{C}$ is a complex scalar that can be obtained as [6, Eq. (53)]:

$$\hat{\alpha}_\mathbb{C} = \frac{||\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star + L^{-1/2}\mathbf{H}_\mathbb{C}^0}^\mathbb{C} - \widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}^\mathbb{C}||}{||\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^\mathsf{H}\underline{\mathrm{vec}}(\mathbf{H}_\mathbb{C}^0)||}, \tag{18}$$

---

[5] The choice of these preliminary estimators and of their impact on the asymptotic performance of $\widehat{\mathbf{V}}_{1,R}$ will be extensively discussed in the next Sections.

and $\mathbf{H}_{\mathbb{C}}^0$ is a "small perturbation", Hermitian, matrix s. t. $[\mathbf{H}_{\mathbb{C}}^0]_{1,1} = 0$.

By substituting Eqs. (14) and (17) in Eq. (13), the $R$-estimator $\widehat{\mathbf{V}}_{1,R}$ can be explicitly re-written as (see [6, Eq. (54)]):

$$
\begin{aligned}
\underline{\text{vec}}(\widehat{\mathbf{V}}_{1,R}) = &\underline{\text{vec}}(\widehat{\mathbf{V}}_1^\star) + \\
&\frac{1}{L\hat{\alpha}_{\mathbb{C}}} \left[\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}}\right]^{-1} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right) \text{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}).
\end{aligned}
\tag{19}
$$

For an in-depth discussion about the semiparametric efficiency and the robustness property characterizing $\widehat{\mathbf{V}}_{1,R}$, we refer the readers to our previous works [5, 6] and to the related statistical literature [10, 11, 12]. Here we focus our attention on an important aspect that has not been fully addressed yet: the computational cost underlying the calculation of Eq. (19). As already noted in [6, Sec. V.C], there is a main (computational) drawback in Eqs. (19) and (18) that really stands out: to evaluate the $N \times N$ matrix $\widehat{\mathbf{V}}_{1,R}$ (or equivalently its $(N^2 - 1)$-dimensional vectorized counterpart $\underline{\text{vec}}(\widehat{\mathbf{V}}_{1,R})$), we have to calculate the $(N^2-1)\times(N^2-1)$ matrix $\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$. This may become a cumbersome bottleneck in many practical applications.

Fortunately, as proved in Appendix A of this paper, it is possible to recast Eqs. (19) and (18) in order to avoid the calculation of $\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$. In particular, a computationally efficient "matrix version" of the $R$-estimator $\widehat{\mathbf{V}}_{1,R}$ can be expressed as:

$$
\boxed{\widehat{\mathbf{V}}_{1,R} = \widehat{\mathbf{V}}_1^\star + \frac{1}{\hat{\alpha}_{\mathbb{C}}}\left(\mathbf{W} - [\mathbf{W}]_{1,1}\,\widehat{\mathbf{V}}_1^\star\right)},
\tag{20}
$$

$$
\boxed{\hat{\alpha}_{\mathbb{C}} = \frac{\|\mathbf{z}_{\widehat{\mathbf{V}}_1^\star + L^{-1/2}\mathbf{H}_{\mathbb{C}}^0} - \mathbf{z}_{\widehat{\mathbf{V}}_1^\star}\|}{\left\|\underline{\text{vec}}\left((\widehat{\mathbf{V}}_1^\star)^{-1}\mathbf{H}_{\mathbb{C}}^0(\widehat{\mathbf{V}}_1^\star)^{-1} - N^{-1}\text{tr}\left((\widehat{\mathbf{V}}_1^\star)^{-1}\mathbf{H}_{\mathbb{C}}^0\right)(\widehat{\mathbf{V}}_1^\star)^{-1}\right)\right\|}},
\tag{21}
$$

where:

$$
\mathbf{W} \triangleq L^{-1/2}(\widehat{\mathbf{V}}_1^\star)^{1/2}\mathbf{R}(\widehat{\mathbf{V}}_1^\star)^{1/2}.
\tag{22}
$$

$$
\mathbf{z}_{\widehat{\mathbf{V}}_1^\star} \triangleq \underline{\text{vec}}\left((\widehat{\mathbf{V}}_1^\star)^{-1/2}\mathbf{R}(\widehat{\mathbf{V}}_1^\star)^{-1/2} - \zeta(\widehat{\mathbf{V}}_1^\star)^{-1}\right),
\tag{23}
$$

$$
\mathbf{R} \triangleq \frac{1}{\sqrt{L}}\sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right)\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^{\mathsf{H}},
\tag{24}
$$

$$
\zeta \triangleq \frac{1}{N\sqrt{L}}\sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right).
\tag{25}
$$

It is worth stressing that Eqs. (20) and (21) involve matrix and vector quantities of (linear) dimension equal at most to $N^2$ and this fact leads to a great reduction in terms of computational load with respect to Eqs. (19) and (18) that, on the contrary, rely on the calculation of matrices whose linear dimension is $N^4$. A quantitative analysis of the reduction of the computational load will be provided in Sec. 5. We conclude this Section by noticing that an

expression similar to Eq. (20) characterizing the $R$-estimator of the shape matrix of a set of Real Elliptically Symmetric (RES) distributed data has been firstly provided in [12, Eq. (3.9)].

*Remark*: The interested reader can find our Matlab implementation of the computationally efficient $R$-estimator $\widehat{\mathbf{V}}_{1,R}$ for the shape matrix of both Real and Complex Elliptically Symmetric distributed data at `https://github.com/StefanoFor`.

## 3 Semiparametric joint estimation of location and shape: the role of the nuisance density generator

After having introduced a computationally efficient version of the $R$-estimator of the shape matrix $\mathbf{V}_{1,0}$, in this Section we will focus on a different still interrelated topic: which is the impact of not knowing the location vector $\boldsymbol{\mu}_0$ when estimating $\mathbf{V}_{1,0}$? Along with its theoretical implication, the answer to this question has a practical importance as well. As shown in the previous Section in fact (see Eqs. (15) and (16)), the $R$-estimator $\widehat{\mathbf{V}}_{1,R}$ relies on $\widehat{\boldsymbol{\mu}}^\star$, i.e. a $\sqrt{L}$-consistent preliminary estimator of $\boldsymbol{\mu}_0$. However, Section 2 does not provide any suggestion on which specific estimator $\widehat{\boldsymbol{\mu}}^\star$ should we choose among all the possible $\sqrt{L}$-consistent ones. In this Section then we are going to provide with the necessary theoretical framework that will allow us to make the good choice for $\widehat{\boldsymbol{\mu}}^\star$.

Let us start by formalizing the problem. Let $\{\mathbf{z}_l\}_{l=1}^L$ be a set of CES distributed vectors such that $\mathbb{C}^N \ni \mathbf{z}_l \sim p_0 \equiv CES_N(\boldsymbol{\mu}_0, \mathbf{V}_{1,0}, h_0)$, $\forall l$ where $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$ have to be considered as two finite-dimensional parameters of interest while $h_0$ is a functional nuisance term.

The two fundamental questions underlying the above mentioned joint estimation problem are:

1. What is the impact of not knowing $h_0$ on the joint estimation of $(\boldsymbol{\mu}_0, \mathbf{V}_{1,0})$?
2. What 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 *semiparamatric Fisher Information Matrix* (SFIM) $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)$. As discussed in the relevant statistical literature for a generic semiparametric model [1] and recently investigated for the specific CES model [3, 4], the semiparametric efficient score vector for the joint estimation of location and shape matrix of a set of CES distributed data is given by:

$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_0} = [\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}^\top, \bar{\mathbf{s}}_{\boldsymbol{\mu}_0^*}^\top, \bar{\mathbf{s}}_{\underline{\text{vec}}(\mathbf{V}_{1,0})}^\top]^\top = \mathbf{s}_{\boldsymbol{\theta}_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0}|\mathcal{T}_{h_0}), \tag{26}$$

where $\mathbf{s}_{\boldsymbol{\theta}_0}$ is the "classical" score vector defined, by means of the Wirtinger derivatives, as [13]:

$$[\mathbf{s}_{\boldsymbol{\theta}_0}]_i \triangleq \partial \ln p_Z(\mathbf{z}; \boldsymbol{\theta}, h_0)/\partial \theta_i^* |_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \; i = 1, \ldots, q \tag{27}$$

and $\boldsymbol{\theta}$ is given in Eq. (4) 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 Eq. (5) 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 $h_0$. In our previous work [3, Sec. III.A], we proved the following facts:

1. The projection of $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}$ and $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0^*}$ onto $\mathcal{T}_{h_0}$ is equal to zero:

$$\Pi(\mathbf{s}_{\boldsymbol{\mu}_0}|\mathcal{T}_{h_0}) = \Pi(\mathbf{s}_{\boldsymbol{\mu}_0^*}|\mathcal{T}_{h_0}) = \mathbf{0}_N, \tag{28}$$

   This implies that the lack of knowledge of $h_0$ does not have any impact on the (asymptotic) estimation of the location parameter $\boldsymbol{\mu}_0$.
2. The projection of $\bar{\mathbf{s}}_{\underline{\text{vec}}(\mathbf{V}_{1,0})}$ onto $\mathcal{T}_{h_0}$ is generally different from zero and it is given by:

$$\Pi(\mathbf{s}_{\underline{\text{vec}}(\mathbf{V}_{1,0})}|\mathcal{T}_{h_0}) = -(1 + N^{-1}\mathcal{Q}\psi_0(\mathcal{Q}))\underline{\text{vec}}(\mathbf{V}_{1,0}^{-1}), \tag{29}$$

   where $\psi_0$ is defined in Eq. (12) while $\mathcal{Q}$ is given in Eq. (10). Consequently, not knowing $h_0$ does have an impact on the estimation of the shape matrix $\mathbf{V}_{1,0}$.

Points 1) and 2) answer the first question.

To address the second question 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 Eq. (5) has been evaluated in [3, Sec. III.C] as:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0) \triangleq E_0\{\bar{\mathbf{s}}_{\boldsymbol{\theta}_0}\bar{\mathbf{s}}_{\boldsymbol{\theta}_0}^{\mathsf{H}}\} = \begin{pmatrix} \bar{\mathbf{I}}(\boldsymbol{\mu}_0|h_0) & \mathbf{0}_{2N\times(N^2-1)} \\ \mathbf{0}_{(N^2-1)\times 2N} & \bar{\mathbf{I}}(\mathbf{V}_{1,0}|h_0) \end{pmatrix}, \tag{30}$$

$$\bar{\mathbf{I}}(\boldsymbol{\mu}_0|h_0) = \frac{E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\}}{N} \begin{pmatrix} \mathbf{V}_{1,0}^{-1} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{V}_{1,0}^{-*} \end{pmatrix}, \tag{31}$$

$$\begin{aligned} \bar{\mathbf{I}}(\underline{\text{vec}}(\mathbf{V}_{1,0})|h_0) &= \frac{E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}}{N(N+1)}\mathbf{L}_{\mathbf{V}_{1,0}}\mathbf{L}_{\mathbf{V}_{1,0}}^{H} = \\ &= \frac{E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}}{N(N+1)}\mathbf{P}\left[\mathbf{V}_{1,0}^{-\top}\otimes\mathbf{V}_{1,0}^{-1} - N^{-1}\text{vec}(\mathbf{V}_{1,0}^{-1})\text{vec}(\mathbf{V}_{1,0}^{-1})^{\mathsf{H}}\right]\mathbf{P}^{\top}, \end{aligned} \tag{32}$$

where, as before, the function $\psi_0$ is given in Eq. (12) while $\mathcal{Q}$ is given in Eq. (10). Eq. (30) clearly shows that the efficient SFIM $\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)$ is a block-diagonal matrix, i.e. 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. This fact greatly simplify the implementation of a practical joint estimation algorithm. In fact, 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 $\boldsymbol{\mu}_0$. This important theoretical result will be exploited in the next Section, to implement a *robust, semiparametric efficient* joint estimator for the location and shape matrix in CES distributed data.

## 4 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 [18]. In particular, in [18], 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:

$$\sum_{l=1}^{L} u_1(\hat{Q}_l^{1/2})(\mathbf{z}_l - \hat{\boldsymbol{\mu}}) = \mathbf{0}, \tag{33}$$

$$L^{-1} \sum_{l=1}^{L} u_2(\hat{Q}_l)(\mathbf{z}_l - \hat{\boldsymbol{\mu}})(\mathbf{z}_l - \hat{\boldsymbol{\mu}})^{\mathsf{H}} = \widehat{\mathbf{V}}_1, \tag{34}$$

where, according to Eq. (10), $\hat{Q}_l = (\mathbf{z}_l - \hat{\boldsymbol{\mu}})^{\mathsf{H}}\widehat{\mathbf{V}}_1^{-1}(\mathbf{z}_l - \hat{\boldsymbol{\mu}})$, and $\{\mathbf{z}_l\}_{l=1}^{L}$ is the set of available CES distributed observations such that $\mathbb{C}^N \ni \mathbf{z}_l \sim p_0 \equiv CES_N(\boldsymbol{\mu}_0, \mathbf{V}_{1,0}, h_0)$, $\forall l$. 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 Eqs. (33) and (34) (see [18] for the real case and [22] for the extension to the complex one).

4.1 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 [31] (see also [8], [20] and [28]) 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

$$\hat{Q}_l^{(k)} = (\mathbf{z}_l - \hat{\boldsymbol{\mu}}^{(k)})^{\mathsf{H}}[\widehat{\mathbf{V}}_1^{(k)}]^{-1}(\mathbf{z}_l - \hat{\boldsymbol{\mu}}^{(k)}), \tag{35}$$

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:

$$\hat{\boldsymbol{\mu}}_{Ty}^{(k+1)} = \left[\sum_{l=1}^{L}[\hat{Q}_l^{(k)}]^{-1/2}\right]^{-1} \sum_{l=1}^{L}\left(\hat{Q}_l^{(k)}\right)^{-1/2}\mathbf{z}_l, \tag{36}$$

$$\begin{cases} \widehat{\mathbf{V}}_{Ty}^{(k+1)} = \dfrac{N}{L}\sum_{l=1}^{L}\dfrac{(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}^{(k)})(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}^{(k)})^{\mathsf{H}}}{\hat{Q}_l^{(k)}}. \\ \widehat{\mathbf{V}}_{1,Ty}^{(k+1)} \triangleq \widehat{\mathbf{V}}_{Ty}^{(k+1)}/[\widehat{\mathbf{V}}_{Ty}^{(k+1)}]_{1,1}. \end{cases} \tag{37}$$

Note that, even if a formal proof of the joint convergence of Eqs. (36) and (37) is still an open problem, this iterative algorithm has been shown to provide reliable estimates in most of the scenarios of possible interest in practical applications. We refer to [8] where joint $M$-estimators of the form (33)-(34) have been exploited in hyperspectral anomaly detection problems and to [28] 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 $h \in \mathcal{G}$ (see [31] for the real-valued case and [20] for the complex-valued case). Consistency, however, is only one of the properties that good robust estimators should have. Another important property is the (semiparametric) efficiency.

4.2 The Semiparametric Cramér-Rao Bound (SCRB)

A robust estimator is said to be *semiparametric efficient* if its Mean Square Error (MSE) achieves the Semiparametric Cramér-Rao Bound (SCRB) [1] as the number of available observations $L$ goes to infinity. The SCRB for the joint estimation of location and shape in CES distributed data has been derived in [3] as the inverse of the SFIM in Eq. (30). We refer the reader to [3] for all the details about its calculation. Here, for the sake of conciseness, we report only the final expression. As discussed before, since the efficient score vectors for the location, i.e. $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}$ and $\bar{\mathbf{s}}_{\boldsymbol{\mu}_0^*}$, 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

$$\mathrm{SCRB}(\boldsymbol{\mu}_0|h_0) \triangleq \bar{\mathbf{I}}(\boldsymbol{\mu}_0|h_0)^{-1} = \frac{N}{E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\}} \begin{pmatrix} \mathbf{V}_{1,0} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{V}_{1,0}^* \end{pmatrix}. \qquad (38)$$

On the other hand, since as previously shown in Eq. (29), $\Pi(\mathbf{s}_{\underline{\mathrm{vec}}(\mathbf{V}_{1,0})}|\mathcal{T}_{h_0}) \neq \mathbf{0}$, 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:

$$\mathrm{SCRB}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0) \triangleq \bar{\mathbf{I}}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0)^{-1} = \frac{N(N+1)}{E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}} \left[ \mathbf{L}_{\mathbf{V}_{1,0}} \mathbf{L}_{\mathbf{V}_{1,0}}^H \right]^{-1},$$
$$(39)$$

where the matrix $\mathbf{L}_{\mathbf{V}_{1,0}}$ is defined in Eq. (8). It is worth mentioning that the expression of the SCRB given in Eq. (39) is valid only if the shape matrix is defined through the constraint in Eq. (3), i.e. when the first-top left element of $\mathbf{V}_{1,0}$ is forced to be equal to 1. The interested reader may find the general form of the SCRB for the shape matrix estimation under any constraints (e.g. constraints on its trace or determinant) in [3, 4].

In [3, 4], 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 $R$-estimator of the shape given in Eqs. (20) in Sec. 2.

4.3 An $R$-estimator of $\mathbf{V}_{1,0}$ in non-centered CES data

In this subsection, we finally put all our previous results together to provide a robust and semiparametric efficient joint estimation of the location vector $\boldsymbol{\mu}_0$ and of the shape matrix $\mathbf{V}_{1,0}$ of a set $\{\mathbf{z}_l\}_{l=1}^{L}$ of CES-distributed observations. As previously discussed, in order to gain the semiparametric efficiency, we will exploit the $R$-estimator in Eq. (20). As amply discussed in Section 2, to implement this estimator we need:

1. A preliminary $\sqrt{L}$-consistent estimator $\widehat{\mathbf{V}}_1^{\star}$ for the shape matrix and another $\sqrt{L}$-consistent estimator of the location vector $\hat{\boldsymbol{\mu}}^{\star}$,
2. A score function $K_h : (0,1) \to \mathbb{R}^{+}$.

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 in Eqs. (36) and (37) are perfect candidates for this role, i.e. $\hat{\boldsymbol{\mu}}^{\star} \equiv \hat{\boldsymbol{\mu}}_{Ty}$ and $\widehat{\mathbf{V}}_1^{\star} \equiv \widehat{\mathbf{V}}_{1,Ty}$. Specifically, the general expression of the (computationally efficient) $R$-estimator given in Section 2 can be recast as:

$$\widehat{\mathbf{V}}_{1,R} = \widehat{\mathbf{V}}_{1,Ty} + \frac{1}{\hat{\alpha}_{\mathbb{C}}} \left( \mathbf{W} - [\mathbf{W}]_{1,1} \widehat{\mathbf{V}}_{1,Ty} \right). \tag{40}$$

Note that, all the other related quantities reported in Eqs. (22) - (25) have to be evaluated by substituting to the generic preliminary estimators $\hat{\boldsymbol{\mu}}^{\star}$ and $\widehat{\mathbf{V}}_1^{\star}$ with the Tyler's estimators for location and scale, $\hat{\boldsymbol{\mu}}_{Ty}$ and $\widehat{\mathbf{V}}_{1,Ty}$ respectively.

Regarding the second point, i.e. the choice of a score function $K_h$, we will exploit two different options [6]:

– The complex *van der Waerden* score function:

$$K_{vdW}(u) \triangleq \Phi_G^{-1}(u), \quad u \in (0,1), \tag{41}$$

where $\Phi_G^{-1}$ indicates the inverse function of the cdf of a Gamma-distributed random variable with parameters $(N, 1)$.

– The complex $t_\nu$-score given by:

$$K_{t_\nu}(u) = \frac{N(2N + \nu)F_{2N,\nu}^{-1}(u)}{\nu + 2NF_{2N,\nu}^{-1}(u)}, \quad u \in (0,1), \tag{42}$$

where $F_{2N,\nu}(u)$ stands for the Fisher cdf with $2N$ and $\nu \in (0, \infty)$ degrees of freedom.

The *van der Waerden* score $K_{vdW}$ has been proved to have excellent performance in terms of efficiency in the estimation of the shape matrix in centered CES data [6, 24] while the $t_\nu$-score $K_{t_\nu}$ is able to provide with a better robustness to the presence of possible outliers thanks to the presence of the additional "tuning" parameter $\nu$.

To conclude this sub section, the pseudocode for the implementation of the $R$-estimator in Eq. (40) is provided in the following while the related Matlab code can be downloaded at `https://github.com/StefanoFor`.

---

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

---

**Input:** $\mathbf{z}_1, \ldots, \mathbf{z}_L$.
**Output:** $\widehat{\mathbf{V}}_{1,R}$.
1: Evaluate the preliminary Tyler's joint estimators:
$\quad \hat{\boldsymbol{\mu}}_{Ty} \leftarrow \lim_{k \to \infty} \hat{\boldsymbol{\mu}}_{Ty}^{(k)}$ in (36),
$\quad \widehat{\mathbf{V}}_{1,Ty} \leftarrow \lim_{k \to \infty} \widehat{\mathbf{V}}_{1,Ty}^{(k+1)}$ in (37),
2: Data centring: $\{\mathbf{z}_l\}_{l=1}^{L} \leftarrow \{\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}\}_{l=1}^{L}$,
3: **for** $l = 1$ to $L$ **do**
4: $\quad \hat{Q}_l^\star \leftarrow \mathbf{z}_l^H \widehat{\mathbf{V}}_{1,Ty}^{-1} \mathbf{z}_l$,
5: $\quad \hat{\mathbf{u}}_l^\star \leftarrow (\hat{Q}_l^\star)^{-1/2} \widehat{\mathbf{V}}_{1,Ty}^{-1/2} \mathbf{z}_l$,
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: Select a score function $K_h$ ($K_{vdW}$ in (41) and $K_{t_\nu}$ in (42) are two options),
9: $\mathbf{R} \leftarrow \frac{1}{\sqrt{L}} \sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right) \hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}$,
10: $\zeta \leftarrow \frac{1}{N\sqrt{L}} \sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right)$,
11: $\mathbf{W} \leftarrow L^{-1/2} (\widehat{\mathbf{V}}_{1,Ty})^{1/2} \mathbf{R} (\widehat{\mathbf{V}}_{1,Ty})^{1/2}$,
12: $\underline{\mathbf{z}}_{\widehat{\mathbf{V}}_{1,Ty}} \leftarrow \underline{\mathrm{vec}}\left((\widehat{\mathbf{V}}_{1,Ty})^{-1/2} \mathbf{R} (\widehat{\mathbf{V}}_{1,Ty})^{-1/2} - \zeta (\widehat{\mathbf{V}}_{1,Ty})^{-1}\right)$,
13: Obtain $\hat{\alpha}_{\mathbb{C}}$ through the following two steps:
$\quad$ 1) Generate a random Hermitian matrix $\mathbf{H}_{\mathbb{C}}^0$ s.t. $[\mathbf{H}_{\mathbb{C}}^0]_{1,1} = 0$,
$\quad$ 2) $\hat{\alpha}_{\mathbb{C}} \leftarrow \dfrac{||\underline{\mathbf{z}}_{\widehat{\mathbf{V}}_{1,Ty} + L^{-1/2} \mathbf{H}_{\mathbb{C}}^0} - \underline{\mathbf{z}}_{\widehat{\mathbf{V}}_{1,Ty}}||}{||\underline{\mathrm{vec}}((\widehat{\mathbf{V}}_{1,Ty})^{-1} \mathbf{H}_{\mathbb{C}}^0 (\widehat{\mathbf{V}}_{1,Ty})^{-1} - N^{-1}\mathrm{tr}((\widehat{\mathbf{V}}_{1,Ty})^{-1} \mathbf{H}_{\mathbb{C}}^0)(\widehat{\mathbf{V}}_{1,Ty})^{-1})||}$,
$\quad$ *Remark:* The entries of $\mathbf{H}_{\mathbb{C}}^0$ should be "small enough" to guarantee that $\widehat{\mathbf{V}}_{1,Ty} + L^{-1/2}\mathbf{H}_{\mathbb{C}}^0$ is a positive definite matrix.
14: The last step: $\widehat{\mathbf{V}}_{1,R} \leftarrow \widehat{\mathbf{V}}_{1,Ty} + \frac{1}{\hat{\alpha}_{\mathbb{C}}}\left(\mathbf{W} - [\mathbf{W}]_{1,1} \widehat{\mathbf{V}}_{1,Ty}\right)$,
15: **return** $\widehat{\mathbf{V}}_{1,R}$

---

## 5 Numerical results

This Section will be basically divided in two parts. In the first one, we discuss the computational advantages that the "matrix version" of the $R$-estimator provided in Eq. (20) has with respect to the "vectorized version" derived in [6] and recalled in Eq. (19). In the second part we finally assess, through numerical simulations, the semiparametric efficiency of the joint estimator $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,R})$, given in Eqs. (36) and (40), respectively.

*Data generation*: In both the two parts, we generate the set of non-centered CES distributed data $\{\mathbf{z}_l\}_{l=1}^{L}$ according to a Generalized Gaussian (GG) distribution [25], such that $\mathbb{C}^N \ni \mathbf{z}_l \sim p_0, \forall l$ where:

$$p_0(\mathbf{z}_l) = |\boldsymbol{\Sigma}_0|^{-1} h_0\left((\mathbf{z}_l - \boldsymbol{\mu}_0)^{\mathsf{H}} \boldsymbol{\Sigma}_0^{-1}(\mathbf{z}_l - \boldsymbol{\mu}_0)\right), \tag{43}$$

while the relevant density generator is given by:

$$h_0(t) \triangleq \frac{s\Gamma(N)b^{-N/s}}{\pi^N \Gamma(N/s)} \exp\left(-\frac{t^s}{b}\right), \ t \in \mathbb{R}^+. \tag{44}$$

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 (44) 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 $[1, \rho, \ldots, \rho^{N-1}]^T$; $\rho = 0.8e^{j2\pi/5}$ 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: $b = [\sigma_Z^2 N \Gamma(N/s) / \Gamma((N+1)/s)]^s$ in Eq. (44) and $\sigma_Z^2 = E\{\mathcal{Q}\}/N = 4$.
– Numbers of observations: $L = 5N$. This clearly defines a "finite-sample" regime.

## 5.1 Computational efficiency of the proposed "matrix version" of the $R$-estimator for shape

As amply discussed in Section 2, the crucial difference between the "vectorized" implementation of the $R$-estimator given in Eq. (19) and its "matrix version" provided in Eq. (20) is in the fact that, while the first one relies on the calculation of $N^2 \times N^2$ matrix quantities, the latter only involves $N \times N$ matrices. Clearly, this will lead to a huge gain in term of computational efficiency, in particular when the data dimension $N$ increases. In order to highlight this fact, in Fig. 1 we report the time (in seconds) required for the calculation of three shape matrix estimators as function of the data dimension $N$:

– The Tyler's estimator $\widehat{\mathbf{V}}_{1,Ty}$ in Eq. (37),
– The "vectorized version" of the $R$-estimator in Eq. (19) exploiting $\widehat{\mathbf{V}}_{1,Ty}$ as preliminary estimator for the shape matrix,
– The "matrix version" of the $R$-estimator in Eq. (20) exploiting $\widehat{\mathbf{V}}_{1,Ty}$ as preliminary estimator for the shape matrix.

The curves in Fig. 1 are clear: the proposed "matrix version" of the $R$-estimator is more than two orders of magnitude faster that the "vectorized" one derived in [6]. The gap between the two clearly increases as the data dimension $N$ increases. Moreover, it can be noted that the computational time of the "matrix version" of the $R$-estimator in Eq. (20) is similar to the one of the Tyler's estimator. These considerations provide us with hard evidence in favor of the computational effectiveness of the "matrix version" of the $R$-estimator derived in Section 2 and suggest us to adopt it as standard form of the $R$-estimator.

## 5.2 Statistical efficiency of the joint estimator for location and shape

As previously anticipated, this last sub session is devoted to the assessment of the semiparametric efficiency of the joint estimator $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,R})$, where $\hat{\boldsymbol{\mu}}_{Ty}$ is
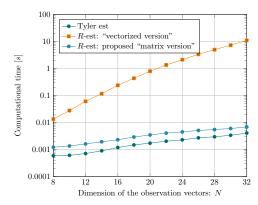
**Fig. 1** Comparison among the time riquired to calculate the Tyler's estimator, the "vectorized" and the proposed "matrix" versions of the $R$-estimator.

the Tyler's estimator in Eq. (36) of the location vector $\boldsymbol{\mu}_0$, while $\widehat{\mathbf{V}}_{1,R}$ is the $R$-estimator in Eq. (40) 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 estimators.

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:

$$\hat{\boldsymbol{\mu}}_{SM} \triangleq L^{-1} \sum_{l=1}^{L} \mathbf{z}_l, \tag{45}$$

$$\begin{cases} \widehat{\boldsymbol{\Sigma}}_{SCM} \triangleq L^{-1} \sum_{l=1}^{L} (\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{SM})(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{SM})^H \\ \widehat{\mathbf{V}}_{1,SCM} \triangleq \widehat{\boldsymbol{\Sigma}}_{SCM}/[\widehat{\boldsymbol{\Sigma}}_{SCM}]_{1,1}. \end{cases} \tag{46}$$

The performance assessment will be performed in terms of the following indices:

*Bias indices*

– Bias index for the estimation of the location vector $\boldsymbol{\mu}_0$:

$$\beta_\gamma \triangleq ||E\{\hat{\boldsymbol{\mu}}_\gamma - \boldsymbol{\mu}_0\}||_2, \tag{47}$$

where $\gamma \in \{SM, Ty\}$ indicates the sample mean in Eq. (45) or the Tyler's estimator in Eq. (36).

– Bias index for the estimation of the shape matrix $\mathbf{V}_{1,0}$:

$$\varphi_\gamma \triangleq ||E\{\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,\gamma} - \mathbf{V}_{1,0})\}||_2, \tag{48}$$

where $\gamma \in \{SCM, Ty, R-vdW, R-t_5\}$ indicates a specif estimator among the SCM in Eq. (46), the Tyler's shape estimator in Eq. (37) and the $R$-estimator in Eq. (40) that relies on the Tyler's one as preliminary estimator. Note that for the $R$-estimator we have two options: $\widehat{\mathbf{V}}_{1,R-vdW}$ indicates the $R$-estimator in Eq. (40) exploiting the *van der Waerden* score in Eq. (41) while $\widehat{\mathbf{V}}_{1,R-t_5}$ indicates again $R$-estimator in Eq. (40) but exploiting the $t$-score in Eq. (42) with $\nu = 5$.

*Mean Squared Error (MSE) indices*

– MSE index for the estimation of the location vector $\boldsymbol{\mu}_0$:

$$\varrho_\gamma \triangleq ||E\{(\hat{\boldsymbol{\mu}}_\gamma^a - \boldsymbol{\mu}_0^a)(\hat{\boldsymbol{\mu}}_\gamma^a - \boldsymbol{\mu}_0^a)^H\}||_F, \tag{49}$$

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

– MSE index for the estimation of the shape matrix $\mathbf{V}_{1,0}$:

$$\varsigma_\gamma \triangleq ||E\{\underline{\text{vec}}(\widehat{\mathbf{V}}_{1,\gamma} - \mathbf{V}_{1,0})\underline{\text{vec}}(\widehat{\mathbf{V}}_{1,\gamma} - \mathbf{V}_{1,0})^H\}||_F, \tag{50}$$

and $\gamma \in \{SCM,\ Ty, R - vdW, R - t_5\}$, as before, indicates the relevant estimator at hand.

As lower bound indices, we use

$$\varepsilon_{SCRB,\boldsymbol{\mu}_0} \triangleq ||\text{SCRB}(\boldsymbol{\mu}_0|h_0)||_F, \tag{51}$$

$$\varepsilon_{SCRB,\mathbf{V}_{1,0}} \triangleq ||\text{SCRB}(\underline{\text{vec}}(\mathbf{V}_{1,0})|h_0)||_F, \tag{52}$$

where $\text{SCRB}(\boldsymbol{\mu}_0|h_0)$ is given in (38) and $\text{SCRB}(\underline{\text{vec}}(\mathbf{V}_{1,0})|h_0)$ in (39).

The bias indices of the sample mean in Eq. (45) and of the Tyler's estimator in Eq. (37) are reported in Fig. 1. As we can note, the bias is on the order of $10^{-3}$, so it can be considered negligible and the two estimators unbiased. Fig. 3 shows the MSE performance of the sample mean $\hat{\boldsymbol{\mu}}_{SM}$ estimator in (45) and of the Tyler's estimator $\hat{\boldsymbol{\mu}}_{Ty}$ in (36) compared to the lover bound in (38). As wee can see, $\hat{\boldsymbol{\mu}}_{Ty}$ is almost efficient with respect to $\text{SCRB}(\boldsymbol{\mu}_0|h_0)$ 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.
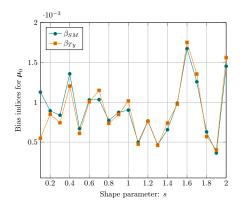


**Fig. 2** Bias in the estimation of $\boldsymbol{\mu}_0$ for the sample mean and for the Tyler's estimator.

**Fig. 3** MSE in the estimation of $\boldsymbol{\mu}_0$ for the sample mean and for the Tyler's estimator.
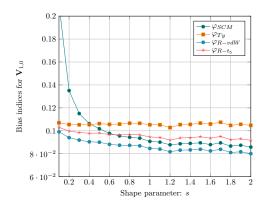


**Fig. 4** Bias in the estimation of $\mathbf{V}_{1,0}$ for the SCM, the Tyler's estimator and the $R$-estimator exploiting both the van der Waerden and the $t_5$-score functions.

As far it concerns the shape matrix estimation, the simulation results are shown in Fig. 4 for the bias and Fig. 5 for the MSE. The main fact here is that the two $R$-estimators $\widehat{\mathbf{V}}_{1,R-vdW}$ and $\widehat{\mathbf{V}}_{1,R-t_5}$ in Eq. (40) outperforms the Tyler's estimator $\widehat{\mathbf{V}}_{1,Ty}$ in (37) for every values of $s$, i.e. for both heavy-tailed and light-tailed data. Moreover, as expected, $\widehat{\mathbf{V}}_{1,R-vdW}$ and $\widehat{\mathbf{V}}_{1,R-t_5}$ greatly outperform the sample covariance matrix $\widehat{\mathbf{V}}_{1,SCM}$ in Eq. (46) in the presence of heavy-tailed data $(0 < s < 1)$, while their MSE is of the same order for $s > 1$. Between the two $R$-estimators, we can notice that $\widehat{\mathbf{V}}_{1,R-vdW}$ has better performance than $\widehat{\mathbf{V}}_{1,R-t_5}$ in terms of both bias and MSE. 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 $\widehat{\mathbf{V}}_{1,R-vdW}$ and $\widehat{\mathbf{V}}_{1,R-t_5}$ and the SCRB. 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,
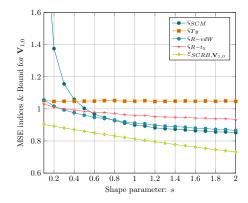
**Fig. 5** MSE in the estimation of $\mathbf{V}_{1,0}$ for the SCM, the Tyler's estimator and the $R$-estimator exploiting both the van der Waerden and the $t_5$-score functions..

by letting $L \rightarrow \infty$, it can be shown that both the two $R$-estimators $\widehat{\mathbf{V}}_{1,R-vdW}$ and $\widehat{\mathbf{V}}_{1,R-t_5}$ achieve the bound $\mathrm{SCRB}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0)$ in Eq. (39) as predicted by theoretical considerations [6, 12].

In summary, previous simulations highlight 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 [19, 21].

## 6 Conclusions

This paper dealt with the fundamental problem of estimating the location vector and the shape matrix of a set of CES distributed data. In the first part of this work, we derived a computationally efficient version of the robust and semiparametric efficient $R$-estimator already proposed in [6]. Remarkably, the new "matrix version" of the $R$-estimator can provide the same estimate of the shape matrix but at a computational time that is more than two order of magnitude smaller with respect to the "vectorized version" previously derived in [6]. This fundamental property suggests us to use the new version given in Eq. (20) as the default version of the $R$-estimator. In the second part of this paper, the joint estimation of the location vector $\boldsymbol{\mu}_0$ and the shape matrix $\mathbf{V}_{1,0}$ 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 $\mathbf{V}_{1,0}$ has been discussed and its performance, in terms of both bias and MSE, 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.

## A Appendix: Proof of the Eqs. (20) and (21)

The main aim of this Appendix is to show how to obtain a closed from expression of the (complex-valued) $R$-estimator for the shape matrix $\mathbf{V}_1$, provided in (19) (see also [6, Eq. (54)]) by avoiding the use of the matrix $\mathbf{L}_{\mathbf{V}_1}$. The matrix $\mathbf{L}_{\mathbf{V}_1}$ is in fact a structured $(N^2 - 1) \times (N^2 - 1)$ that is built upon the $N \times N$ matrix $\mathbf{V}_1$. Consequently, if we are able to obtain a expression of the $R$-estimator that relies only on $\mathbf{V}_1$ and not on $\mathbf{L}_{\mathbf{V}_1}$, we would gain a lot in terms of computational efficiency.

At first, let us recall here the expression of the $R$-estimator introduced in Eqs. (19) and (18):

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^\star) +$$
$$\frac{1}{L\hat{\alpha}_{\mathbb{C}}} \left[ \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}} \right]^{-1} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \sum_{l=1}^{L} K_h \left( \frac{r_l^\star}{L+1} \right) \mathrm{vec}(\hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}).$$

where

$$\hat{\alpha}_{\mathbb{C}} = ||\widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star + L^{-1/2} \mathbf{H}^0_{\mathbb{C}}} - \widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star} || / ||\mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}} \underline{\mathrm{vec}}(\mathbf{H}^0_{\mathbb{C}})||,$$

and $\mathbf{H}^0_{\mathbb{C}}$ is a "small perturbation", Hermitian, matrix s. t. $[\mathbf{H}^0_{\mathbb{C}}]_{1,1} = 0$.

In the calculation proposed below, we make extensive use of the following properties holding for conforming matrices (see e.g. [26]):

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{vec}(\mathbf{A}^\top)^\top \mathrm{vec}(\mathbf{B}) \tag{A.1}$$

$$\mathrm{vec}(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^\top \otimes \mathbf{A})\mathrm{vec}(\mathbf{X}) \tag{A.2}$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}) \tag{A.3}$$

$$(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top, \quad (\mathbf{A} \otimes \mathbf{B})^{\mathsf{H}} = \mathbf{A}^{\mathsf{H}} \otimes \mathbf{B}^{\mathsf{H}} \tag{A.4}$$

A.1 Matrix version of the central sequence

The "distribution-free" version of the (complex-valued) efficient central sequence $\widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star}$ is defined in Eq. (14) as:

$$\widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star} \triangleq \frac{1}{\sqrt{L}} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \sum_{l=1}^{L} K_h \left( \frac{r_l^\star}{L+1} \right) \mathrm{vec}(\hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}).$$

Let us start by showing how to re-write $\widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star}$ without using $\mathbf{L}_{\mathbf{V}_1}$.

For notation simplicity, we introduce the scalar $\kappa_l$ as:

$$\kappa_l \triangleq \frac{1}{\sqrt{L}} K_h \left( \frac{r_l^\star}{L+1} \right) \tag{A.5}$$

Let us now define the $N^2$-dimensianl vector $\mathbf{z}_{\widehat{\mathbf{V}}_1^\star}$ as:

$$\mathbf{z}_{\widehat{\mathbf{V}}_1^\star} = \left[ z_1, \quad (\underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star})^\top \right]^\top = \left[ z_1, \quad (\widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star})^\top \right]^\top. \tag{A.6}$$

where $z_1$ is an unspecified complex number. Then, from Eq. (14) and from the definition of the matrices $\mathbf{P}$ and $\mathbf{L}_{\mathbf{V}_1}$ in Eqs. (6) and (8) respectively, the vector $\mathbf{z}_{\widehat{\mathbf{V}}_1^\star}$ can be cast as:

$$
\begin{aligned}
\mathbf{z}_{\widehat{\mathbf{V}}_1^\star} &= \left[ (\widehat{\mathbf{V}}_1^\star)^{-\top/2} \otimes (\widehat{\mathbf{V}}_1^\star)^{-1/2} \right] \Pi^\perp_{\text{vec}(\mathbf{I}_N)} \sum_{l=1}^{L} \kappa_l \text{vec}(\hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}) \\
&= \left[ (\widehat{\mathbf{V}}_1^\star)^{-\top/2} \otimes (\widehat{\mathbf{V}}_1^\star)^{-1/2} \right] \left[ \sum_{l=1}^{L} \kappa_l \text{vec}(\hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}) - \frac{1}{N} \text{vec}(\mathbf{I}_N) \sum_{l=1}^{L} \kappa_l \right] \\
&= \text{vec} \left( (\widehat{\mathbf{V}}_1^\star)^{-1/2} \mathbf{R} (\widehat{\mathbf{V}}_1^\star)^{-1/2} - \zeta (\widehat{\mathbf{V}}_1^\star)^{-1} \right)
\end{aligned}
$$

$$\tag{A.7}$$

where:

$$\mathbf{R} \triangleq \sum_{l=1}^{L} \kappa_l \hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}, \tag{A.8}$$

and

$$\zeta \triangleq \frac{1}{N} \sum_{l=1}^{L} \kappa_l. \tag{A.9}$$

Consequently, we have that:

$$\widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star} = \underline{\text{vec}} \left( (\widehat{\mathbf{V}}_1^\star)^{-1/2} \mathbf{R} (\widehat{\mathbf{V}}_1^\star)^{-1/2} - \zeta (\widehat{\mathbf{V}}_1^\star)^{-1} \right) \triangleq \underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star}. \tag{A.10}$$

Note that in (A.7), we have used the fact that:

$$\text{tr}((\hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}) = (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}} \hat{\mathbf{u}}_l^\star = 1, \quad \forall l. \tag{A.11}$$

To avoid confusion, we indicate as $\underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star}$ the "$\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$-free" version of $\widetilde{\boldsymbol{\Delta}}^{\mathbb{C}}_{\widehat{\mathbf{V}}_1^\star}$ obtained in Eq. (A.10). It is important to underline in fact that $\underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star}$ does not make use of the "unnecessary large" $(N^2 - 1) \times (N^2 - 1)$ matrix $\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$, while only $N \times N$ matrices are involved.

A.2 An "$\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$-free" version of the scalar $\hat{\alpha}_\mathbb{C}$

Let us define the $N^2$-dimensional vector as:

$$\mathbf{r}_{\widehat{\mathbf{V}}_1^\star} = \begin{bmatrix} r_1, & (\underline{\mathbf{r}}_{\widehat{\mathbf{V}}_1^\star})^\top \end{bmatrix}^\top = \begin{bmatrix} r_1, & (\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}}\underline{\mathrm{vec}}(\mathbf{H}_\mathbb{C}^0))^\top \end{bmatrix}^\top, \qquad (A.12)$$

where $r_1$ is an unspecified complex scalar.

By using the fact that, by definition, $[\mathbf{H}_\mathbb{C}^0]_{1,1} = 0$, we have that:

$$\begin{aligned}
\mathbf{r}_{\widehat{\mathbf{V}}_1^\star}^0 &= \begin{bmatrix} r_1^0, & (\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}}\underline{\mathrm{vec}}(\mathbf{H}_\mathbb{C}^0))^\top \end{bmatrix}^\top \\
&= \left[ \left( (\widehat{\mathbf{V}}_1^\star)^{-\top/2} \otimes (\widehat{\mathbf{V}}_1^\star)^{-1/2} \right) \Pi^\perp_{\mathrm{vec}(\mathbf{I}_N)} \left( (\widehat{\mathbf{V}}_1^\star)^{-\top/2} \otimes (\widehat{\mathbf{V}}_1^\star)^{-1/2} \right)^{\mathsf{H}} \right] \mathrm{vec}(\mathbf{H}_\mathbb{C}^0) \\
&= \left[ (\widehat{\mathbf{V}}_1^\star)^{-\top} \otimes (\widehat{\mathbf{V}}_1^\star)^{-1} - N^{-1}\mathrm{vec}\left( (\widehat{\mathbf{V}}_1^\star)^{-1} \right) \mathrm{vec}\left( (\widehat{\mathbf{V}}_1^\star)^{-1} \right)^{\mathsf{H}} \right] \mathrm{vec}(\mathbf{H}_\mathbb{C}^0) \\
&= \mathrm{vec}\left( (\widehat{\mathbf{V}}_1^\star)^{-1}\mathbf{H}_\mathbb{C}^0(\widehat{\mathbf{V}}_1^\star)^{-1} - N^{-1}\mathrm{tr}\left( (\widehat{\mathbf{V}}_1^\star)^{-1}\mathbf{H}_\mathbb{C}^0 \right) (\widehat{\mathbf{V}}_1^\star)^{-1} \right)
\end{aligned}$$
$$(A.13)$$

Then, from Eq. (A.12), we get:

$$\underline{\mathbf{r}}_{\widehat{\mathbf{V}}_1^\star}^0 = \underline{\mathrm{vec}}\left( (\widehat{\mathbf{V}}_1^\star)^{-1}\mathbf{H}_\mathbb{C}^0(\widehat{\mathbf{V}}_1^\star)^{-1} - N^{-1}\mathrm{tr}\left( (\widehat{\mathbf{V}}_1^\star)^{-1}\mathbf{H}_\mathbb{C}^0 \right) (\widehat{\mathbf{V}}_1^\star)^{-1} \right), \qquad (A.14)$$

that depends only on $N \times N$ matrix quantities.

Finally, by using the previous results, an "$\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$-free" version of $\hat{\alpha}_\mathbb{C}$ can be expressed as:

$$\hat{\alpha}_\mathbb{C} = \frac{||\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star + L^{-1/2}\mathbf{H}_\mathbb{C}^0}^\mathbb{C} - \widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}^\mathbb{C}||}{||\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}}\underline{\mathrm{vec}}(\mathbf{H}_\mathbb{C}^0)||} = \frac{||\underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star + L^{-1/2}\mathbf{H}_\mathbb{C}^0} - \underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star}||}{||\underline{\mathbf{r}}_{\widehat{\mathbf{V}}_1^\star}^0||}. \qquad (A.15)$$

A.3 An "$\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$-free" matrix version of the $R$-estimator

By using the scalar $\kappa_l$, previously defined in Eq. (A.5), we can re-write the expression of the $R$-estimator in Eq. (19) as:

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^\star) + \frac{1}{\sqrt{L}\hat{\alpha}_\mathbb{C}} \left[ \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}} \right]^{-1} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \sum_{l=1}^{L} \kappa_l \mathrm{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^{\mathsf{H}}). \quad (A.16)$$

By using the "$\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}$-free" version of $\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}^\mathbb{C}$ obtained in Eq. (A.10), the expression in Eq. (A.16) can be rewritten as:

$$\begin{aligned}
\mathrm{vec}(\widehat{\mathbf{V}}_{1,R}) = \mathrm{vec}(\widehat{\mathbf{V}}_1^\star) + \frac{1}{\sqrt{L}\hat{\alpha}_\mathbb{C}}\mathbf{P}^\top \left[ \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}} \right]^{-1} \mathbf{P} \times \\
\mathrm{vec}\left( (\widehat{\mathbf{V}}_1^\star)^{-1/2}\mathbf{R}(\widehat{\mathbf{V}}_1^\star)^{-1/2} - \zeta(\widehat{\mathbf{V}}_1^\star)^{-1} \right).
\end{aligned}$$
$$(A.17)$$

To get rid of the "unnecessary large" $N^2 \times N^2$ matrix $\mathbf{P}^\top \left[ \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}} \right]^{-1} \mathbf{P}$ we make use of the extension to the complex field of the result obtained in Lemma 3.1 of [10]. Specifically, it can be shown that (see [12, Appendix A.2]):

$$
\begin{aligned}
& \mathbf{P}^\top \left[ \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}} \right]^{-1} \mathbf{P} \\
& = \left[ \mathbf{I}_{N^2} - \mathrm{vec}(\widehat{\mathbf{V}}_1^\star) \mathbf{e}_1^\top \right] \left[ (\widehat{\mathbf{V}}_1^\star)^\top \otimes \widehat{\mathbf{V}}_1^\star \right] \left[ \mathbf{I}_{N^2} - \mathrm{vec}(\widehat{\mathbf{V}}_1^\star) \mathbf{e}_1^\top \right]^{\mathsf{H}},
\end{aligned}
\tag{A.18}
$$

where $\mathbf{e}_1$ is the first vector of the canonical basis of $\mathbb{R}^{N^2}$.

Through direct calculation, we can easily show that:

$$
\begin{aligned}
& \mathbf{P}^\top \left[ \mathbf{L}_{\widehat{\mathbf{V}}_1^\star} \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^{\mathsf{H}} \right]^{-1} \mathbf{P} \mathrm{vec} \left( (\widehat{\mathbf{V}}_1^\star)^{-1/2} \mathbf{R} (\widehat{\mathbf{V}}_1^\star)^{-1/2} - \zeta (\widehat{\mathbf{V}}_1^\star)^{-1} \right) \\
& = \left[ \mathbf{I}_{N^2} - \mathrm{vec}(\widehat{\mathbf{V}}_1^\star) \mathbf{e}_1^\top \right] \left[ (\widehat{\mathbf{V}}_1^\star)^\top \otimes \widehat{\mathbf{V}}_1^\star \right] \times \\
& \qquad\qquad \mathrm{vec} \left( (\widehat{\mathbf{V}}_1^\star)^{-1/2} \mathbf{R} (\widehat{\mathbf{V}}_1^\star)^{-1/2} - \zeta (\widehat{\mathbf{V}}_1^\star)^{-1} \right) \\
& = \left[ \mathbf{I}_{N^2} - \mathrm{vec}(\widehat{\mathbf{V}}_1^\star) \mathbf{e}_1^\top \right] \mathrm{vec} \left( (\widehat{\mathbf{V}}_1^\star)^{1/2} \mathbf{R} (\widehat{\mathbf{V}}_1^\star)^{1/2} - \zeta \widehat{\mathbf{V}}_1^\star \right) \\
& = \mathrm{vec} \left( \mathbf{W} - [\mathbf{W}]_{1,1} \widehat{\mathbf{V}}_1^\star \right),
\end{aligned}
\tag{A.19}
$$

where

$$
\begin{aligned}
\mathbf{W} & \triangleq L^{1/2} (\widehat{\mathbf{V}}_1^\star)^{1/2} \mathbf{R} (\widehat{\mathbf{V}}_1^\star)^{1/2} \\
& = (\widehat{\mathbf{V}}_1^\star)^{1/2} \left[ \frac{1}{L} \sum_{l=1}^{L} K_h \left( \frac{r_l^\star}{L+1} \right) \hat{\mathbf{u}}_l^\star (\hat{\mathbf{u}}_l^\star)^{\mathsf{H}} \right] (\widehat{\mathbf{V}}_1^\star)^{1/2}.
\end{aligned}
\tag{A.20}
$$

By collecting the previous results, the $R$-estimator in Eq. (19) can be expressed as:

$$
\mathrm{vec}(\widehat{\mathbf{V}}_{1,R}) = \mathrm{vec}(\widehat{\mathbf{V}}_1^\star) + \frac{1}{\hat{\alpha}_{\mathbb{C}}} \mathrm{vec}(\mathbf{W} - [\mathbf{W}]_{1,1} \widehat{\mathbf{V}}_1^\star).
\tag{A.21}
$$

Finally, in matrix form, we have:

$$
\boxed{\widehat{\mathbf{V}}_{1,R} = \widehat{\mathbf{V}}_1^\star + \frac{1}{\hat{\alpha}_{\mathbb{C}}} \left( \mathbf{W} - [\mathbf{W}]_{1,1} \widehat{\mathbf{V}}_1^\star \right),}
\tag{A.22}
$$

where

$$
\boxed{\hat{\alpha}_{\mathbb{C}} = \frac{\| \underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star + L^{-1/2} \mathbf{H}_{\mathbb{C}}^0} - \underline{\mathbf{z}}_{\widehat{\mathbf{V}}_1^\star} \|}{\| \underline{\mathbf{r}}_{\widehat{\mathbf{V}}_1^\star}^0 \|}.}
\tag{A.23}
$$

This concludes the proof of Eqs. (20) and (21).

## References

1. Bickel P, Klaassen C, Ritov Y, Wellner J (1993) Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press
2. Bishop CM (2007) Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. Springer
3. Fortunati S, Gini F, Greco MS, Zoubir AM, Rangaswamy M (2019) Semiparametric CRB and Slepian-Bangs formulas for complex elliptically symmetric distributions. IEEE Transactions on Signal Processing 67(20):5352–5364
4. Fortunati S, Gini F, Greco MS, Zoubir AM, Rangaswamy M (2019) Semiparametric inference and lower bounds for real elliptically symmetric distributions. IEEE Transactions on Signal Processing 67(1):164–177
5. Fortunati S, Renaux A, Pascal F (2020) Properties of a new $R$-estimator of shape matrices. EUSIPCO 2020
6. Fortunati S, Renaux A, Pascal F (2020) Robust semiparametric efficient estimators in complex elliptically symmetric distributions. IEEE Transactions on Signal Processing 68:5003–5015
7. Fortunati S, Renaux A, Pascal F (2020) Robust semiparametric joint estimators of location and scatter in elliptical distributions. In: 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), pp 1–6, DOI 10.1109/MLSP49062.2020.9231865
8. Frontera-Pons J, Veganzones MA, Pascal F, Ovarlez J (2016) Hyperspectral anomaly detectors using robust estimators. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9(2):720–731
9. Hájek J (1968) Asymptotic normality of simple linear rank statistics under alternatives. Ann Math Statist 39(2):325–346
10. Hallin M, Paindaveine D (2006) Semiparametrically efficient rank-based inference for shape I. Optimal rank-based tests for sphericity. The Annals of Statistics 34(6):2707–2756
11. Hallin M, Paindaveine D (2009) Parametric and semiparametric inference for shape: the role of the scale functional. Statistics & Decisions 24(3):327–350
12. Hallin M, Oja H, Paindaveine D (2006) Semiparametrically efficient rank-based inference for shape II. Optimal $R$-estimation of shape. The Annals of Statistics 34(6):2757–2789
13. Hjørungnes A (2011) Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications. Cambridge University Press
14. Huber PJ, Ronchetti EM (2011) Robust Statistics (Second Edition). John Wiley & Sons
15. Kreutz-Delgado K (2009) The complex gradient operator and the CR-calculus. URL https://arxiv.org/abs/0906.4835
16. Le Cam L, Yang GL (2000) Asymptotics in Statistics: Some Basic Concepts (second edition). Springer series in statistics

17. Manolakis DG, Marden D, Kerekes JP, Shaw GA (2001) Statistics of hyperspectral imaging data. In: Shen SS, Descour MR (eds) Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII, International Society for Optics and Photonics, SPIE, vol 4381, pp 308 – 316

18. Maronna RA (1976) Robust $M$-estimators of multivariate location and scatter. Ann Statist 4(1):51–67

19. Mériaux B, Ren C, El Korso MN, Breloy A, Forster P (2019) Robust estimation of structured scatter matrices in (mis)matched models. Signal Processing 165:163 – 174

20. Mériaux B, Ren C, Korso MNE, Breloy A, Forster P (2019) Asymptotic performance of complex $m$-estimators for multivariate location and scatter estimation. IEEE Signal Processing Letters 26(2):367–371

21. Mériaux B, Ren C, Breloy A, El Korso MN, Forster P (2020) Matched and mismatched estimation of kronecker product of linearly structured scatter matrices under elliptical distributions. IEEE Transactions on Signal Processing DOI 10.1109/TSP.2020.3042946

22. Ollila E, Tyler DE, Koivunen V, Poor HV (2012) Complex elliptically symmetric distributions: Survey, new results and applications. IEEE Transactions on Signal Processing 60(11):5597–5625

23. Ollila E, Tyler DE, Koivunen V, Poor HV (2012) Compound-gaussian clutter modeling with an inverse gaussian texture distribution. IEEE Signal Processing Letters 19(12):876–879

24. Paindaveine D (2006) A Chernoff-Savage result for shape:on the nonadmissibility of pseudo-Gaussian methods. Journal of Multivariate Analysis 97(10):2206 – 2220

25. Pascal F, Bombrun L, Tourneret J, Berthoumieu Y (2013) Parameter estimation for multivariate generalized gaussian distributions. IEEE Transactions on Signal Processing 61(23):5960–5971

26. Petersen KB, Pedersen MS (2012) The matrix cookbook. URL `http://www2.imm.dtu.dk/pubdb/p.php?3274`, version 20121115

27. Remmert R (1991) Theory of Complex Functions. New York: Springer

28. Roizman V, Jonckheere M, Pascal F (2020) A flexible EM-like clustering algorithm for noisy data. In: Arxiv, URL `https://arxiv.org/abs/1907.01660`

29. Sangston KJ, Gini F, Greco MS (2012) Coherent radar target detection in heavy-tailed compound-gaussian clutter. IEEE Transactions on Aerospace and Electronic Systems 48(1):64–77

30. Schroth C, Muma M (2021) Robust $M$-estimation based bayesian cluster enumeration for real elliptically symmetric distributions. IEEE Transactions on Signal Processing DOI 10.1109/TSP.2021.3072482

31. Tyler DE (1987) A distribution-free $M$-estimator of multivariate scatter. The Annals of Statistics 15(1):234–251

32. van der Vaart AW (1998) Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press

33. Zoubir AM, Koivunen V, Ollila E, Muma M (2018) Robust Statistics for Signal Processing. Cambridge University Press