Supporting material for the paper:

Robust Semiparametric Efficient Estimators in Complex Elliptically Symmetric Distributions

Stefano Fortunati, Alexandre Renaux, Frédéric Pascal

I. Le Cam's one-step estimators in a nutshell

The aim of this first section is to provide the reader of our paper with some additional discussion about the general theory of *efficient one-step estimators*. This class of estimators has its root in the concept of Local Asymptotic Normality (LAN) of a statistical model. The LAN property has been introduced for the first time by Le Cam in his fundamental work [1] (see also [2, Ch. 6]) and it has since established itself as a milestone in modern statistics. Leaving aside the deep theoretical implications that the LAN property has for a given family of distributions, there is at least one outcome of great interest for any practitioner working in signal processing (SP) and related fields. As Le Cam showed, if a statistical model is Locally Asymptotic Normal, then it is possible to derive asymptotically efficient estimators that, unlike the Maximum Likelihood (ML) one, do not search for the maxima of the log-likelihood function. This fact is of great importance in practical applications, where the ML estimator can present computational difficulties in the resulting optimization problem or even existence/uniqueness issues [3, Ch. 6].

We start by introducing the concept of *Hellinger differentiability*, or *differentiability in quadratic mean*. Then, the definition of the LAN property for *parametric models* will be given and its exploitation, in deriving efficient one-step estimators, discussed. Finally, the generalization of the previously developed theory to *semiparametric models* will be provided.

Algebraic notation: Throughout this document, italics indicates scalar quantities (a), lower case and upper case boldface indicate column vectors ($\mathbf{a}$) and matrices ($\mathbf{A}$), respectively. Each entry of a matrix $\mathbf{A}$ is indicated as $a_{ij} \triangleq [\mathbf{A}]_{i,j}$. $\mathbf{I}_N$ defines the $N \times N$ identity matrix. The superscript $\top$ indicates the transpose operator, then $\mathbf{A}^{-\top} \triangleq (\mathbf{A}^{-1})^{\top} = (\mathbf{A}^{\top})^{-1}$. The Euclidean norm of a vector $\mathbf{a}$ is indicated as $||\mathbf{a}||$. The determinant and the Frobenius norm of a matrix $\mathbf{A}$ are indicated as $|\mathbf{A}|$ and $||\mathbf{A}||_F$, respectively.

Small o notation: Given a real-valued function $f(x)$ and a strictly positive real-valued function $g(x)$, $f(x) = o(g(x))$ if for every positive real number a, there exists a real number x_0 such that $|f(x)| \leq a g(x),\ \forall x \geq x_0$.

Statistical notation: Let x_l be a sequence of random variables in the same probability space. We write:

- $x_l = o_P(1)$ if $\lim_{l\to\infty} \Pr\{|x_l| \geq \epsilon\} = 0, \forall \epsilon > 0$ (*convergence in probability to 0*),
- $x_l = O_P(1)$ if for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $L > 0$, s.t. $\Pr\{|x_l| > M\} < \epsilon, \forall l > L$ (*stochastic boundedness*).

The cumulative distribution function (cdf) and the related probability density function (pdf) of a random variable x or a random vector $\mathbf{x}$ are indicated as P_X and p_X, respectively. For random variables and vectors, $\overset{d}{=}$ stands for "has the same distribution as". The symbol $\underset{L\to\infty}{\sim}$ indicates the convergence in distribution. We indicate the *true* pdf as $p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\boldsymbol{\phi}_0, g_0)$, where $\boldsymbol{\phi}_0$ and g_0 indicate the true parameter vector to be estimated and the true nuisance function, respectively. We define as $E_{\boldsymbol{\phi},g}\{f(\mathbf{x})\} = \int f(\mathbf{x}) p_X(\mathbf{x}|\boldsymbol{\phi}, g) d\mathbf{x}$ the expectation operator of a *measurable* function f of a random vector $\mathbf{x}$. Moreover, we simply indicate as $E_0\{\cdot\}$ the expectation with

respect to (w.r.t.) the true pdf $p_0(\mathbf{x})$. The superscript $\star$ indicates a $\sqrt{L}$-consistent, *preliminary*, estimator $\hat{\boldsymbol{\phi}}^{\star}$ of $\boldsymbol{\phi}_0$, s.t. $\sqrt{L}\,(\boldsymbol{\phi}^{\star} - \boldsymbol{\phi}_0) = O_P(1)$.

Let $\mathbf{x} \in \mathbb{R}^N$ be a real-valued random vector and let p_X be its probability density function (pdf). A parametric model, characterizing the statistical behavior of $\mathbf{x}$, will be indicated as:

$$\mathcal{P}_{\boldsymbol{\phi}} = \{p_X | p_X(\mathbf{x}|\boldsymbol{\phi}); \boldsymbol{\phi} \in \Omega \subseteq \mathbb{R}^q\}, \tag{62}$$

while a semiparametric model will be described as:

$$\mathcal{P}_{\boldsymbol{\phi},g} = \{p_X | p_X(\mathbf{x}|\boldsymbol{\phi}, g); \boldsymbol{\phi} \in \Omega \subseteq \mathbb{R}^q, g \in \mathcal{G}\}, \tag{63}$$

where $\mathcal{G}$ is a suitable set of functions.

A. Hellinger differentiability

Let $\boldsymbol{\phi} \in \Omega \subseteq \mathbb{R}^q$ be the parameter vector and let $p_X(\mathbf{x}|\boldsymbol{\phi}) \in \mathcal{P}_{\boldsymbol{\phi}}$ be a pdf belonging to the *parametric model* $\mathcal{P}_{\boldsymbol{\phi}}$ in Eq. (62). We define $u_{\boldsymbol{\phi}}(\mathbf{x})$ as the following parametric map:

$$\begin{aligned} u_{\boldsymbol{\phi}} &: \Omega \to \mathcal{L}_2 \\ &\boldsymbol{\phi} \mapsto u_{\boldsymbol{\phi}}(\mathbf{x}) \triangleq \sqrt{p_X(\mathbf{x}|\boldsymbol{\phi})}, \end{aligned} \tag{64}$$

where $\mathcal{L}_2$ indicates the set of all the square integrable functions. We say that $u_{\boldsymbol{\phi}}$ is Hellinger differentiable in $\boldsymbol{\phi} \in \Omega$ if there exists a vector $\dot{\mathbf{u}}_{\boldsymbol{\phi}} \equiv \dot{\mathbf{u}}_{\boldsymbol{\phi}}(\mathbf{x})$ such that [4, Ch. 2, Def. 1], [5, Ch. 5.5]:

$$\int \left[u_{\boldsymbol{\phi}+\mathbf{h}}(\mathbf{x}) - u_{\boldsymbol{\phi}}(\mathbf{x}) - \mathbf{h}^{\top}\dot{\mathbf{u}}_{\boldsymbol{\phi}}(\mathbf{x})\right]^2 d\mathbf{x} = o(||\mathbf{h}||), \quad \mathbf{h} \in \Omega, \quad ||\mathbf{h}|| \to 0. \tag{65}$$

Then $\dot{\mathbf{u}}_{\boldsymbol{\phi}} \equiv \dot{\mathbf{u}}_{\boldsymbol{\phi}}(\mathbf{x})$ is the Hellinger derivative of $u_{\boldsymbol{\phi}}$ in $\boldsymbol{\phi} \in \Omega$. According to [4, Ch. 2, Def. 2], a parametric model $\mathcal{P}_{\boldsymbol{\phi}}$ is said to be *regular* if each $p_X(\mathbf{x}|\boldsymbol{\phi}) \in \mathcal{P}_{\boldsymbol{\phi}}$ is Hellinger differentiable at every $\boldsymbol{\phi} \in \Omega$.

The Hellinger differentiability was introduced by Le Cam as the weakest regularity condition required to develop the LAN theory. However, even if extremely useful for theoretical purposes, the Hellinger differentiability is not really suitable to derive practical inference algorithms. Fortunately, statistical models involved in practical signal processing (SP) applications can generally satisfy more stringent assumptions than the one in Eq. (65). This allows us to link the regularity "*á la Le Cam*" of a parametric model to more familiar quantities, e.g. the *score vector* and the *Fisher Information Matrix* (FIM), as detailed in the following Proposition (see [4, Ch. 2, Prop. 1] for the proof).

Proposition 1. *Let $\mathbf{x}$ be a set of N-dimensional, real-valued, random vector sampled from a pdf $p_X \in \mathcal{P}_{\boldsymbol{\phi}}$ in Eq.* (62)*. Let $\mathbf{s}_{\boldsymbol{\phi}} \equiv \mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x})$ be the score vector defined as:*

$$\mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x}) \triangleq \nabla_{\boldsymbol{\phi}} \ln p_X(\mathbf{x}|\boldsymbol{\phi}) \tag{66}$$

and let $\mathbf{I}(\boldsymbol{\phi})$ be the Fisher Information Matrix (FIM):

$$\mathbf{I}(\boldsymbol{\phi}) \triangleq E_{\boldsymbol{\phi}}\left\{\mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x})\mathbf{s}_{\boldsymbol{\phi}}^{\top}(\mathbf{x})\right\}. \tag{67}$$

Then, the parametric model $\mathcal{P}_{\boldsymbol{\phi}}$ is regular "á la Le Cam" if the following three sufficient (but not necessary) conditions are satisfied:

i) $p_X(\mathbf{x}|\boldsymbol{\phi})$ is continuously differentiable in $\boldsymbol{\phi} \in \Omega$ for almost all $\mathbf{x}$ with gradient $\nabla_{\boldsymbol{\phi}} p_X(\mathbf{x}|\boldsymbol{\phi})$,

ii) $E_{\boldsymbol{\phi}}\{\mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x})^{\top}\mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x})\} < \infty$,

iii) *The FIM in Eq.* (67) *is non-singular and continuous in* $\boldsymbol{\phi} \in \Omega$.

If i), ii) and iii) hold true, the Hellinger derivative $\dot{\mathbf{u}}_{\boldsymbol{\phi}}$ *defined in Eq.* (65) *can be explicitly expressed as function of the score vector* $\mathbf{s}_{\boldsymbol{\phi}}$ *in Eq.* (66) *as:*

$$\dot{\mathbf{u}}_{\boldsymbol{\phi}}(\mathbf{x}) = \frac{1}{2}\sqrt{p_X(\mathbf{x}|\boldsymbol{\phi})}\mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x}). \tag{68}$$

The regularity conditions *i*), *ii*) and *iii*) in Prop. 1 requires, among others, the pointwise differentiability of the pdf and consequently they are more stringent than the integral condition in Eq. (65). However, they are generally satisfied by the vast majority of the statistical models exploited in practical inference problems. For this reason, in the following discussion, we will assume them for granted but we will always indicate when the obtained results can be derived starting form the weaker regularity condition in Eq. (65).

B. LAN property and ES distributions

The following Proposition introduces the fundamental LAN property ([1], [2, Ch. 6], [5, Ch. 7.6]) of a parametric model satisfying the regularity conditions stated in Prop. 1.

Proposition 2. *Let* $\{\mathbf{x}_l\}_{l=1}^{L}$ *be a set of real-valued, i.i.d. observations sampled from a pdf* p_X *belonging to a regular parametric model* $\mathcal{P}_{\boldsymbol{\phi}}$ *in Eq.* (62). *Let* $\boldsymbol{\Delta}_{\boldsymbol{\phi}}(\mathbf{x}_1, \ldots, \mathbf{x}_L)$ *be a random vector, usually referred to as central sequence, defined as:*

$$\boldsymbol{\Delta}_{\boldsymbol{\phi}}(\mathbf{x}_1, \ldots, \mathbf{x}_L) \equiv \boldsymbol{\Delta}_{\boldsymbol{\phi}} \triangleq L^{-1/2}\sum\nolimits_{l=1}^{L} \mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x}_l), \tag{69}$$

where $\mathbf{s}_{\boldsymbol{\phi}}(\mathbf{x}_l)$ *is the score vector given in Eq.* (66).

Then, any $p_X(\mathbf{x}|\boldsymbol{\phi}) \in \mathcal{P}_{\boldsymbol{\phi}}$ *satisfies the following LAN property:*

$$\ln \frac{\prod_{l=1}^{L} p_X(\mathbf{x}_l|\boldsymbol{\phi} + L^{-1/2}\mathbf{h})}{\prod_{l=1}^{L} p_X(\mathbf{x}_l|\boldsymbol{\phi})} = \mathbf{h}^{\top}\boldsymbol{\Delta}_{\boldsymbol{\phi}} - \frac{1}{2}\mathbf{h}^{\top}\mathbf{I}(\boldsymbol{\phi})\mathbf{h} + o_P(1), \quad \forall \boldsymbol{\phi}, \mathbf{h} \in \Omega, \tag{70}$$

where $\mathbf{I}(\boldsymbol{\phi})$ *is the FIM given in Eq.* (67).

Moreover $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$ *satisfies the following two properties:*

C1 *Asymptotic differentiability (or asymptotic linearity): for all* $\boldsymbol{\phi}, \mathbf{h} \in \Omega$

$$\boldsymbol{\Delta}_{\boldsymbol{\phi}+L^{-1/2}\mathbf{h}} - \boldsymbol{\Delta}_{\boldsymbol{\phi}} = -\mathbf{I}(\boldsymbol{\phi})\mathbf{h} + o_P(1), \tag{71}$$

C2 *Asymptotic normality:*

$$\boldsymbol{\Delta}_{\boldsymbol{\phi}} \underset{L\to\infty}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}(\boldsymbol{\phi})), \quad \forall \boldsymbol{\phi} \in \Omega. \tag{72}$$

Remark: The proof of Prop. 2 and extensive in-depth discussion about the LAN property can be found in [1], [2, Ch. 6], and [5, Ch. 7.6].

Before moving on, it is important to stress that the LAN property can be defined in much more general settings, e.g. for non-i.i.d. observations and for statistical models that do not admit a FIM or even a score vector. Actually, under the regularity conditions in Prop. 1, the expansion in Eq. (70) can be thought as the second-order Taylor approximation of the log-likelihood function [5, Ch. 7.2]. Anyway, as said before, even if they are not the weakest ones, the assumptions made in Prop. 2 are satisfied by many data generating processes in SP applications. In

particular, they are met by the Elliptical Symmetric (ES) distributions. Specifically, let us define the parametric model of the Real ES (RES) distributions as:

$$\mathcal{P}_{\boldsymbol{\phi}} = \left\{ p_X | p_X(\mathbf{x}|\boldsymbol{\phi}) = 2^{-N/2} |\mathbf{V}_1|^{-1/2} g_0 \left((\mathbf{x}_l - \boldsymbol{\mu})^\top \mathbf{V}_1^{-1} (\mathbf{x}_l - \boldsymbol{\mu}) \right) ; \boldsymbol{\phi} \in \Omega \right\}, \tag{73}$$

and the parameter vector $\boldsymbol{\phi}$ is defined in Eq. (6) of our paper as $\boldsymbol{\phi} \triangleq \left(\boldsymbol{\mu}^\top, \underline{\text{vecs}}(\mathbf{V}_1)^\top \right)^\top$, where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the location vector and $\mathbf{V}_1 \in \mathcal{M}_N^{\mathbb{R}}$ is the shape matrix s.t. $[\mathbf{V}_1]_{1,1} = 1$. The general proof of the fact that the RES model in Eq. (73) is regular and satisfies the LAN property in Prop. 2 has been provided by Hallin and Paindaveine in [6, Prop. 2.1] (see also [6, Appendix 1]). As mentioned above, this is of great practical importance because, as proved by Le Cam in [1], [2, Ch. 6], if a parametric model is Local Asymptotic Normal, then asymptotically efficient estimators of the parameter of interest $\boldsymbol{\phi}$ can be built using a "one-step linear correction" to any preliminary $\sqrt{L}$-consistent estimator $\hat{\boldsymbol{\phi}}^\star$ of the true parameter vector $\boldsymbol{\phi}_0$.

C. Efficient one-step parametric estimators

In parametric setting, the standard procedure to derive efficient estimators is given by the Maximum Likelihood theory. Specifically, given a set of i.i.d. data $\{\mathbf{x}_l\}_{l=1}^L$, an asymptotically efficient estimate of the true parameter vector $\boldsymbol{\phi}_0 \in \Omega \subseteq \mathbb{R}^q$, if it exists, can be obtained as:

$$\hat{\boldsymbol{\phi}}_{ML} \triangleq \underset{\boldsymbol{\phi} \in \Omega}{\text{argmax}} \sum\nolimits_{l=1}^{L} \ln p_X(\mathbf{x}_l|\boldsymbol{\phi}). \tag{74}$$

As every practitioner knows, solving the optimization problem in Eq. (74) may result to be a prohibitive task and, in some cases, $\hat{\boldsymbol{\phi}}_{ML}$ may not even exist or may not be unique [3, Ch. 6]. So, it would be useful to figure out a different methodology to derive efficient estimates.

Under the regularity conditions stated in Prop. 1, if $\hat{\boldsymbol{\phi}}_{ML}$ exists, then it satisfies:

$$\boldsymbol{\Delta}_{\boldsymbol{\phi}}(\mathbf{x}_1, \ldots, \mathbf{x}_M)|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}_{ML}} \equiv \boldsymbol{\Delta}_{\hat{\boldsymbol{\phi}}_{ML}} = \mathbf{0}, \tag{75}$$

where $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$ is the central sequence defined in Eq. (69). Eq. (75) can be thought as a set of q nonlinear equations, then we can define a new estimator $\hat{\boldsymbol{\phi}}$ given by the one-step Newton-Raphson approximate solution of Eq. (75) as:

$$\hat{\boldsymbol{\phi}} = \tilde{\boldsymbol{\phi}} - [\mathbf{J}_{\boldsymbol{\Delta}}(\tilde{\boldsymbol{\phi}})]^{-1} \boldsymbol{\Delta}_{\tilde{\boldsymbol{\phi}}}, \tag{76}$$

where $\tilde{\boldsymbol{\phi}}$ is a "good" starting point and $\mathbf{J}_{\boldsymbol{\Delta}}(\tilde{\boldsymbol{\phi}})$ indicates the Jacobian matrix of $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$ evaluated at $\tilde{\boldsymbol{\phi}}$. Note that the approximation in Eq. (76) is valid even if $\hat{\boldsymbol{\phi}}_{ML}$ does not exists. In [1] and [2, Ch. 6], Le Cam formalized and generalized this intuitive procedure by providing an asymptotic characterization of the class of efficient one-step estimators. This fundamental result is summarized in the following theorem (see also [5, Ch. 5.7]).

Theorem 1. *Let $\{\mathbf{x}_l\}_{l=1}^L$ be a set of i.i.d. observations sampled from the "true" pdf $p_0 \in \mathcal{P}_{\boldsymbol{\phi}}$ satisfying the LAN property as in Prop. 2. Let $\hat{\boldsymbol{\phi}}^\star$ any preliminary $\sqrt{L}$-consistent estimator of the true parameter vector $\boldsymbol{\phi}_0 \in \Omega$. Then, the one-step estimator*

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}^\star + L^{-1/2} \mathbf{I}(\hat{\boldsymbol{\phi}}^\star)^{-1} \boldsymbol{\Delta}_{\hat{\boldsymbol{\phi}}^\star}, \tag{77}$$

has the following properties:

P1 *$\sqrt{L}$-consistency*

$$\sqrt{L} \left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 \right) = O_P(1), \tag{78}$$

P2 *Asymptotic normality and efficiency*

$$\sqrt{L}\left(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}_0\right) \underset{L\to\infty}{\sim} \mathcal{N}(\mathbf{0},\mathbf{I}(\boldsymbol{\phi}_0)^{-1}), \tag{79}$$

where $\mathbf{I}(\boldsymbol{\phi}_0)^{-1}=\mathrm{CRB}(\boldsymbol{\phi}_0)$ *is the Cramér-Rao Bound.*

Proof: Let us start by showing that the expression defining the one-step estimator in Eq. (77) can be derived directly from the Newton-Raphson approximation in Eq. (76), using the asymptotic differentiability property C1, given in Eq. (71), of the central sequence. Specifically, in analogy with the definition of Jacobian matrix, we have that:

$$\mathbf{J}_{\boldsymbol{\Delta}}(\boldsymbol{\phi}) \equiv -L^{1/2}\mathbf{I}(\boldsymbol{\phi})+o_P(1), \quad \forall \boldsymbol{\phi}\in\Omega. \tag{80}$$

Finally, substituting Eq. (80) in Eq. (76), and noticing that $\hat{\boldsymbol{\phi}}^{\star}$ is a good starting point since it is, by definition, in the $\sqrt{L}$-neighborhood of $\boldsymbol{\phi}_0$, yields the expression Eq. (77).

The proof of the $\sqrt{L}$-consistency property P1 of $\hat{\boldsymbol{\phi}}$ can be found in [4, Sec. 2.5, Th. 2]. To prove the property P2, we start from the intermediate result provided in [4, Sec. 2.3, Th. 1], that is $\mathbf{I}(\boldsymbol{\phi})^{-1}\boldsymbol{\Delta}_{\boldsymbol{\phi}} \underset{L\to\infty}{\sim} \mathcal{N}(\mathbf{0},\mathbf{I}(\boldsymbol{\phi})^{-1})$. Consequently, using the fact that $\hat{\boldsymbol{\phi}}^{\star}$ is $\sqrt{L}$-consistent, the asymptotic normality and efficiency of $\hat{\boldsymbol{\phi}}$ in Eq. (77) follows form a direct application of the Slutsky's theorem [5, Lemma 2.8]. Note that the same warning raised up for Prop. 2 holds here for Theorem 1. In fact, in [4, Sec. 2.3, Th. 1 and Sec. 2.5, Th. 2] only the Hellinger differentiability is required, while here we need to assume the existence of the gradient (w.r.t. $\boldsymbol{\phi}\in\Omega$) of the log-likelihood function. ■

Since, as shown in [6, Prop. 2.1], the RES model in Eq. (73) satisfies the LAN property, Theorem 1 can be readily applied to derive a one-step efficient estimator of the true parameter vector $\boldsymbol{\phi}_0 \triangleq \left(\boldsymbol{\mu}_0^{\top}, \underline{\mathrm{vecs}}(\mathbf{V}_{1,0})^{\top}\right)^{\top}$. The closed form expressions of the score vector $\mathbf{s}_{\boldsymbol{\phi}}$ (and consequently the one of the central sequence $\boldsymbol{\Delta}_{\boldsymbol{\phi}}$) and of the FIM $\mathbf{I}(\boldsymbol{\phi})$, needed to implement the estimator in Eq. (77), can be directly obtained by the ones already derived in our previous work [7]. Moreover, as preliminary $\sqrt{L}$-consistent estimator we may use:

$$\hat{\boldsymbol{\phi}}^{\star}=\left(\hat{\boldsymbol{\mu}}_{Ty}^{\top}, \underline{\mathrm{vecs}}(\widehat{\mathbf{V}}_{1,Ty})^{\top}\right)^{\top}, \tag{81}$$

where $\hat{\boldsymbol{\mu}}_{Ty}^{\top}$ and $\widehat{\mathbf{V}}_{1,Ty}$ are the joint Tyler's estimates of the location vector and of the shape matrix constrained to have $[\widehat{\mathbf{V}}_{1,Ty}]_{1,1}=1$ [8], [9].

The result in Theorem 1 would be enough to derive original, asymptotically efficient, estimators of the location vector $\boldsymbol{\mu}_0$ and of the shape matrix $\mathbf{V}_{1,0}$ in the classical parametric context. Here however, we want to go one step further towards the semiparametric framework.

D. One-step, semiparametric estimators

A semiparametric model $\mathcal{P}_{\boldsymbol{\phi},g}$ is a set of pdfs parameterized by a finite-dimensional parameter vector $\boldsymbol{\phi}\in\Omega\subseteq\mathbb{R}^q$ and by a function $g\in\mathcal{G}$ that usually plays the role of an infinite-dimensional nuisance parameter [4], [10]. As amply discussed in [7] and [11] the ES distributions are a perfect candidate to be modeled as a semiparametric model, since we generally do not have any *a priori* information on the actual density generator g_0 characterizing the specific distribution of the observations. Specifically, the RES semiparametric model can be expressed as:

$$\mathcal{P}_{\boldsymbol{\phi},g}=\left\{p_X | p_X(\mathbf{x}|\boldsymbol{\phi},g)=2^{-N/2}|\mathbf{V}_1|^{-1/2}g\left((\mathbf{x}_l-\boldsymbol{\mu})^{\top}\mathbf{V}_1^{-1}(\mathbf{x}_l-\boldsymbol{\mu})\right); \boldsymbol{\phi}\in\Omega, g\in\mathcal{G}\right\}, \tag{82}$$

where, as for the parametric case, $\boldsymbol{\phi} = \left(\boldsymbol{\mu}^\top, \underline{\text{vecs}}(\mathbf{V}_1)^\top\right)^\top$ while $\mathcal{G}$ is the set of all the "proper" density generators, i.e. $\mathcal{G} = \left\{ g : \mathbb{R}^+ \to \mathbb{R}^+ \,\middle|\, \int_0^\infty t^{N/2-1} g(t) dt < \infty, \int p_X d\mathbf{x} = 1 \right\}$ [12].

The question that we are going to address here is the following: *is it possible to generalize the concept of one-step estimators, as formalized in Theorem 1, to semiparametric inference problems?* To answer to this important point, let us start by focusing on the main building blocks needed to derive the one-step estimator $\hat{\boldsymbol{\phi}}$ given, for the parametric case, in Eq. (77). As already discussed in the dedicated statistical literature (see e.g. [4], [10], [13]) and in our recent works [7], [11], [14], the semiparametric counterpart of the score vector $\mathbf{s}_{\boldsymbol{\phi}}$ is the *efficient* score vector $\bar{\mathbf{s}}_{\boldsymbol{\phi},g_0}$ defined as (see [14] and [7, Th. IV.1]):

$$\bar{\mathbf{s}}_{\boldsymbol{\phi},g_0}(\mathbf{x}) \equiv \bar{\mathbf{s}}_{\boldsymbol{\phi},g_0} \triangleq \mathbf{s}_{\boldsymbol{\phi}} - \Pi(\mathbf{s}_{\boldsymbol{\phi}}|\mathcal{T}_{g_0}), \tag{83}$$

where $\Pi(\mathbf{s}_{\boldsymbol{\phi}}|\mathcal{T}_{g_0})$ is the orthogonal projection of the score vector $\mathbf{s}_{\boldsymbol{\phi}}$ in Eq. (66) on the semiparametric nuisance tangent space $\mathcal{T}_{g_0}$ [15], [5, Ch. 25.4]. The semiparametric counterpart of the FIM $\mathbf{I}(\boldsymbol{\phi})$ is the *efficient* semiparametric FIM (SFIM) [14], [7, Th. IV.1]:

$$\bar{\mathbf{I}}(\boldsymbol{\phi}|g_0) \triangleq E_{\boldsymbol{\phi},g_0}\{\bar{\mathbf{s}}_{\boldsymbol{\phi},g_0}(\mathbf{x})\bar{\mathbf{s}}_{\boldsymbol{\phi},g_0}(\mathbf{x})^\top\}. \tag{84}$$

On the same line of Eq. (69), we introduce the *efficient* central sequence $\overline{\boldsymbol{\Delta}}_{\boldsymbol{\phi},g}$ simply as:

$$\overline{\boldsymbol{\Delta}}_{\boldsymbol{\phi},g}(\mathbf{x}_1, \ldots, \mathbf{x}_L) \equiv \overline{\boldsymbol{\Delta}}_{\boldsymbol{\phi},g} \triangleq L^{-1/2} \sum\nolimits_{l=1}^{L} \bar{\mathbf{s}}_{\boldsymbol{\phi},g}(\mathbf{x}_l), \quad \forall \boldsymbol{\phi} \in \Omega,\ g \in \mathcal{G}. \tag{85}$$

The natural "semiparametric" generalization of the ML estimating equations in Eq. (75) would be [5, Ch. 25.8]

$$\boldsymbol{\Delta}_{\boldsymbol{\phi},g}(\mathbf{x}_1, \ldots, \mathbf{x}_M)|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}_{ML}, g=\hat{g}^\star} \equiv \boldsymbol{\Delta}_{\hat{\boldsymbol{\phi}}_{ML},\hat{g}^\star} = \mathbf{0}. \tag{86}$$

It must be readily noted that the critical difference between the ML estimating equation in Eq. (75) and their semiparametric generalization in Eq. (86) is that the latter involve a preliminary $\sqrt{L}$-consistent, *non-parametric*, estimator $\hat{g}^\star$ of the nuisance function g. Unfortunately, as discussed in [5, Ch. 25.8] and in [4, Ch. 7], it is generally impossible to find an estimator of the infinite-dimensional nuisance g that converge to the true function g_0 at the $O_P(L^{-1/2})$ rate characterizing most of the parametric estimators. Roughly speaking, the non-parametric estimation of a function requires much more data then the ones needed to estimate a finite-dimensional parameter.

For the specific problem of the semiparametric shape matrix estimation in RES distributions, in their seminal work [16], Hallin, Oja and Paindaveine proposed a different approach that does not involve the non-parametric estimation of g_0, still providing *nearly* semiparametric efficient estimator of $\boldsymbol{\phi} = \left(\boldsymbol{\mu}^\top, \underline{\text{vecs}}(\mathbf{V}_1)^\top\right)^\top$. The basic idea developed in [16] is to split the semiparametric estimation problem at hand in two parts:

1) Assume that the true density generator g_0 is known and solve Eq. (86) to derive a "clairvoyant" semiparametric estimatior $\hat{\boldsymbol{\phi}}_s$.
2) Robustify $\hat{\boldsymbol{\phi}}_s$ by using a distribution-free, rank based, procedure.

To better understand this approach, let us start by analyzing the properties of the clairvoyant efficient central sequence $\boldsymbol{\Delta}_{\boldsymbol{\phi},g_0}$ of a set of RES distributed data.

Proposition 3. *Let $\{\mathbf{x}_l\}_{l=1}^L$ be a set of i.i.d. observations sampled from a RES pdf $p_0 \in \mathcal{P}_{\boldsymbol{\phi},g}$ in Eq.* (82)*. Then, the clairvoyant efficient central sequence $\boldsymbol{\Delta}_{\boldsymbol{\phi},g_0}$ satisfies the following two properties:*

CS1 *Asymptotic differentiability (or asymptotic linearity): for all $\boldsymbol{\phi}, \mathbf{h} \in \Omega$*

$$\overline{\boldsymbol{\Delta}}_{\boldsymbol{\phi}+L^{-1/2}\mathbf{h},g_0} - \overline{\boldsymbol{\Delta}}_{\boldsymbol{\phi},g_0} = -\bar{\mathbf{I}}(\boldsymbol{\phi}|g_0)\mathbf{h} + o_P(1), \tag{87}$$

CS2 *Asymptotic normality*

$$\overline{\boldsymbol{\Delta}}_{\boldsymbol{\phi},g_0} \underset{L\to\infty}{\sim} \mathcal{N}(\mathbf{0}, \bar{\mathbf{I}}(\boldsymbol{\phi}|g_0)), \quad \forall \boldsymbol{\phi} \in \Omega. \tag{88}$$

Remark: The proof can be found in [6, Sec. 3].

The result in Prop. 3 suggests us that, for the semiparametric RES estimation problem at hand, it may be possible to derive semiparametric and asymptotically efficient estimators using a procedure similar to the one provided in Theorem 1, simply by substituting the parametric score vector and FIM with their semiparametric counterparts. This intuition is formalized by the next theorem that is also given in our main paper as Theorem 1.

Theorem 2. *Let $\{\mathbf{x}_l\}_{l=1}^{L}$ be a set of i.i.d. observations sampled from a RES distribution with pdf $p_0 \in \mathcal{P}_{\boldsymbol{\phi},g}$ in Eq. (82). Let $\hat{\boldsymbol{\phi}}^\star$ be any preliminary $\sqrt{L}$-consistent estimator of the true parameter vector $\boldsymbol{\phi}_0 = \left(\boldsymbol{\mu}_0^\top, \underline{\mathrm{vecs}}(\mathbf{V}_{1,0})^\top\right)^\top$. Then, the clairvoyant semiparametric one-step estimator*

$$\hat{\boldsymbol{\phi}}_s = \hat{\boldsymbol{\phi}}^\star + L^{-1/2}\bar{\mathbf{I}}(\hat{\boldsymbol{\phi}}^\star|g_0)^{-1}\overline{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\phi}}^\star,g_0}, \tag{89}$$

has the following properties:

PS1 $\sqrt{L}$*-consistency*

$$\sqrt{L}\left(\hat{\boldsymbol{\phi}}_s - \boldsymbol{\phi}_0\right) = O_P(1), \tag{90}$$

PS2 *Asymptotic normality and efficiency*

$$\sqrt{L}\left(\hat{\boldsymbol{\phi}}_s - \boldsymbol{\phi}_0\right) \underset{L\to\infty}{\sim} \mathcal{N}(\mathbf{0}, \bar{\mathbf{I}}(\boldsymbol{\phi}_0|g_0)^{-1}), \tag{91}$$

where $\bar{\mathbf{I}}(\boldsymbol{\phi}_0|g_0)^{-1} = \mathrm{CSCRB}(\boldsymbol{\phi}_0|g_0) = \mathrm{CSCRB}(\boldsymbol{\mu}_0, \mathbf{V}_{1,0}|g_0)$ and the constrained semiparametric CRB (CSCRB) [7] is evaluated for the constraint $[\mathbf{V}_{1,0}]_{1,1} = 1$.

Proof: The expression of the semiparametric one-step estimator in Eq. (89) can be obtained using the same arguments discussed in Theorem 1. The proof of the $\sqrt{L}$-consistency property PS1 of $\hat{\boldsymbol{\phi}}_s$ can be found in [4, Sec. 7.8, Th. 1]. To prove the asymptotic normality, we start from the intermediate result, given in [4, Sec. 3.3, Th. 2], that $\bar{\mathbf{I}}(\boldsymbol{\phi}|g_0)^{-1}\overline{\boldsymbol{\Delta}}_{\boldsymbol{\phi},g_0} \underset{L\to\infty}{\sim} \mathcal{N}(\mathbf{0}, \bar{\mathbf{I}}(\boldsymbol{\phi}|g_0)^{-1})$. Then, from the expression Eq. (89) and from the fact that $\hat{\boldsymbol{\phi}}^\star$ is $\sqrt{L}$-consistent, the asymptotic normality and efficiency property PS2 of $\hat{\boldsymbol{\phi}}_s$ follows from a direct application of the Slutsky's theorem (see also [4, Sec. 7.8, Cor. 1]). Again, here we need to assume the existence of the gradient (w.r.t. $\boldsymbol{\phi} \in \Omega$) of the log-likelihood function, while in the proof [4, Sec. 7.8, Th. 1] only the Hellinger differentiability is required. ■

As previously underlined and as we can see from its closed form expression in Eq. (89), the clairvoyant estimator $\hat{\boldsymbol{\phi}}_s$ relies on the true density generator g_0, so it is not useful for inference problems in the semiparametric model (82) where the density generator is an unknown nuisance function. However, it has the fundamental role to link the parametric one-step Le Cam's estimator in Eq. (77) with a distributionally robust estimator of the shape matrix, as shown in [16] and recalled in Section III of our paper.

II. Numerical analysis for real t-distributed data

This Section mimics Sec. V of the main paper and provides a numerical investigation about the statistical performance of the *real* R-estimator in Eq. (38) in *real* t-distributed data.

As in the main paper, in order to distinguish different estimators, each of them will be indicated as $\widehat{\mathbf{V}}_{1,\gamma}^{\varphi}$ where γ and φ specify the estimator at hand. Moreover, we re-normalized $\widehat{\mathbf{V}}_{1,\gamma}^{\varphi}$ in order to have $\mathrm{tr}(\widehat{\mathbf{V}}_{1,\gamma}^{\varphi}) = N$, i.e. $\widehat{\mathbf{V}}_{\gamma}^{\varphi} = N\widehat{\mathbf{V}}_{1,\gamma}^{\varphi}/\mathrm{tr}(\widehat{\mathbf{V}}_{1,\gamma}^{\varphi})$.

As a reference, in the figures we also report the Constrained Semiparametric CRB (CSCRB) derived, in closed form, in [7]. As performance index for the shape matrix estimators, we use

$$\varsigma_{\gamma}^{\varphi} = ||E\{\text{vecs}(\widehat{\mathbf{V}}_{\gamma}^{\varphi} - \mathbf{V}_0)\text{vecs}(\widehat{\mathbf{V}}_{\gamma}^{\varphi} - \mathbf{V}_0)^{\top}\}||_F, \tag{92}$$

Similarly, as performance bound, we adopt the index:

$$\varepsilon_{CSCRB} = |||[\text{CSCRB}(\mathbf{\Sigma}_0, g_0)]||_F. \tag{93}$$

Unlike the main paper, where a set of complex GG-distributed data are considered, here we generate the dataset according to a real t-distribution. The density generator for the t-distribution is [12]: [1]

$$g_0(t) = \frac{2^{N/2}\Gamma(\frac{\lambda+N}{2})}{(\lambda\pi)^{N/2}\Gamma(\lambda/2)}\left(1+\frac{t}{\lambda}\right)^{-\frac{\lambda+N}{2}},\ t \in \mathbb{R}^+ \tag{94}$$

and the degrees of freedom $\lambda \in (0, \infty)$ controls the non-Gaussianity of the data. In particular, for small values of λ the data are highly non-Gaussian while, as $\lambda \to \infty$, the distribution collapses into the Gaussian one. The simulation parameters for this study case are:

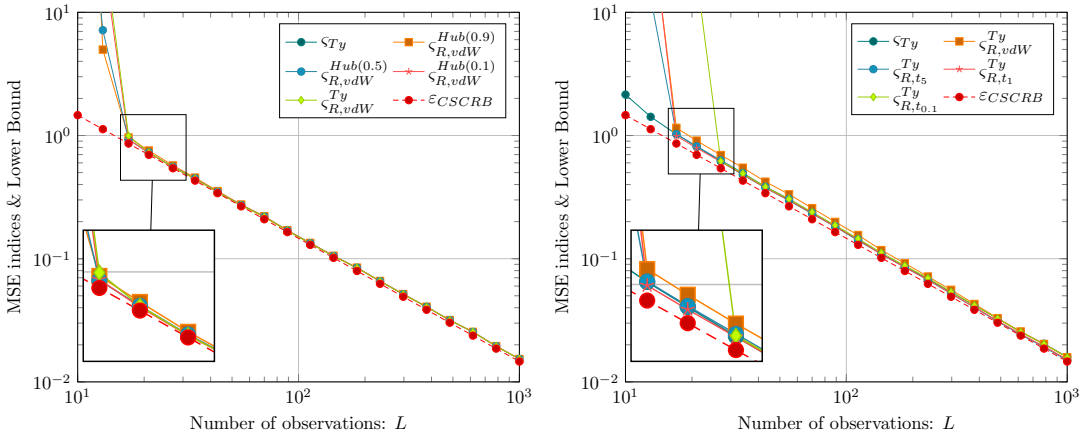
- $[\mathbf{\Sigma}_0]_{i,j} = \rho^{|i-j|},\ i,j = 1,\ldots,N$; $\rho = 0.8$ and $N = 8$.
- The "small perturbation" matrix $\mathbf{H}^0$ is chosen to be a symmetric random matrix s.t. $\mathbf{H}^0 = (\mathbf{G} + \mathbf{G}^T)/2$ where $[\mathbf{G}]_{i,j} \sim \mathcal{N}(0, \upsilon^2)$, $[\mathbf{G}]_{1,1} = 0$ and $\upsilon = 0.01$. Note that υ should be small enough to guarantee that $\widehat{\mathbf{V}}_1^{\star} + L^{-1/2}\mathbf{H}^0 \in \mathcal{M}_N^{\mathbb{R}}$.

As discussed in the main paper, the R-estimator in Eq. (38) depends on two "user-defined" quantities: 1) the preliminary estimator $\widehat{\mathbf{V}}_1^{\star}$ and 2) the score function K_g. In order to assess the impact of their choice on the performance of the R-estimator, we perform our simulations by using the Tyler's and the Huber's estimators as preliminary estimators. Moreover, for the Huber's estimator, three different values of the tuning parameter q (i.e. $q = 0.9, 0.5, 0.1$) have been adopted [17, Sec. V.C]. Moreover, as score functions, we exploit the *van der Waerden* one and the t_ν-score for $\nu = 0.1, 1, 5$, given in Eqs. (34) and (35) of the main paper. As we will see in the following, the simulation results obtained for the real case are perfectly in line with the one reported in the main paper for the complex case.

A. Semiparametric efficiency

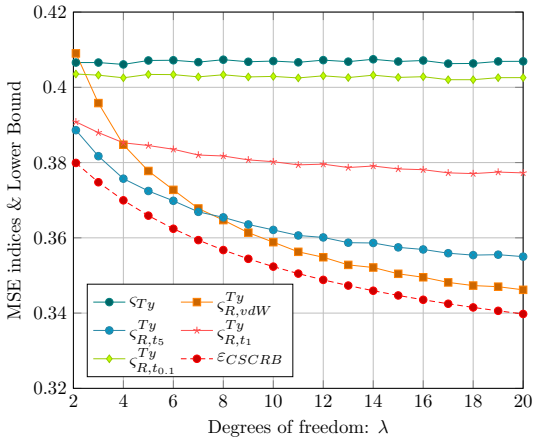
In Figs. 1(a) and 1(b), MSE indices of the real R-estimator in Eq. (38) are plotted as function of the number L of t-distributed observations with $\lambda = 5$ and then compared with the CSCRB. Specifically, in Fig. 1(a) the asymptotic efficiency of the R-estimator, exploiting a *van der Waerden* score, is investigated for the two considered preliminary estimators, i.e. Tyler's and Huber's one. As for the complex case, the impact of the choice of the preliminary estimator on the efficiency of the R-estimator is negligible. Similarly, the asymptotic impact of the choice of the score functions is also negligible, as shown in Fig. 1(b). However, as for the complex case, the score function plays a role in the "finite-sample" performance of the estimator. To see this, in Fig. 1(c), we report the MSE indices obtained for the *van der Waerden* and t_ν- scores as function of the degrees of freedom λ for $L = 5N$. Note that, for $\lambda = 5$, the t_5-score is perfectly specified and then it provides the lowest MSE value at $\lambda = 5$. However, as for the complex case, the *van der Waerden* score confirms its surprisingly good performance (see the discussion on the "Chernoff-Savage result" provided in the main paper).

[1]Note that the expression of the density generator in Eq. (94) can be obtained from the one given in [7, Eq. (75)] by putting $\eta = 1$.



(a) MSE indices vs preliminary Tyler's and Huber's estimators as function of L ($\lambda = 5$).

(b) MSE indices vs different score functions K_g as function of L ($\lambda = 5$).



(c) MSE indices vs different score functions K_g as function of λ ($L = 5N$).

Fig. 1: MSE performance of the real R-estimator.

The t_ν-scores are more flexible since the additional parameter ν can be used to tune the desired trade-off between semiparametric efficiency and robustness to outliers, as we will see ahead. In particular, t_ν-scores characterized by a small value of ν increases the robustness of the resulting R-estimator at the price of a loss of efficiency. On the other hand, larger values of ν will provide a better efficiency, sacrificing the robustness as addressed in the next section.

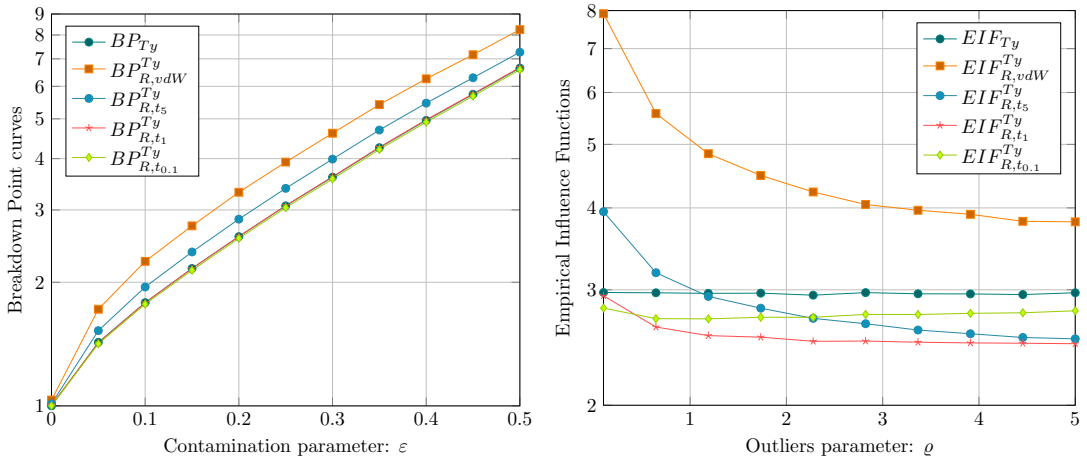
B. Robustness to outliers

Following Sec. V.B of the main paper, in this subsection we evaluate the "finite-sample" Breakdown Point (BP) [18] and the Empirical Influence Function (EIF) [19] for the real R-estimator in Eq. (38).

We indicate with $X = \{\mathbf{x}_l\}_{l=1}^{L} \sim RES(\mathbf{0}, \mathbf{V}_1, g_0)$ the "pure" t-distributed data set whose g_0 is given in Eq. (94) and with $X_\varepsilon = \{\mathbf{x}_l\}_{l=1}^{L} \sim f_{X_\varepsilon}$ the *ε-contaminated* data set s.t.:

$$f_{X_\varepsilon}(\mathbf{x}|\mathbf{V}_1, g_0, \varrho) = (1-\varepsilon)RES(\mathbf{0}, \mathbf{V}_1, h_0) + \varepsilon q_X(\varrho), \tag{95}$$

where $\varepsilon \in [0, 1/2]$ is a contamination parameter. The function $q_X(\varrho)$ represents the pdf of an outlier $\tilde{\mathbf{x}}$ that we arbitrary choose to be as $\tilde{\mathbf{x}} = \tau^{-1}\mathbf{u}$ where $\mathbf{u} \sim \mathcal{U}(S_{\mathbb{R}^N})$ while $\tau \sim \mathrm{Gam}(\varrho, 1/\varrho)$ and Gam indicates the Gamma



(a) BP vs different score functions K_g as function of ε ($L = 5N$,$\varrho = 0.1$,$\lambda = 5$).

(b) EIF vs different score functions K_g as function of ϱ ($L = 1000$,$\lambda = 5$).

Fig. 2: BP and EIF of the real R-estimator in t-distributed data.

distribution. The reader can find additional discussion about this model in Sec. V.B of the main paper.

Let $\widehat{\mathbf{V}}^{\varphi}_{\gamma}(X)$ and $\widehat{\mathbf{V}}^{\varphi}_{\gamma}(X_\varepsilon)$ be two shape matrix estimators evaluated from the pure and the ε-contaminated data sets, respectively. As for the complex case, the finite-sample BP curves can be evaluated as [18]:

$$BP^{\varphi}_{\gamma}(\varepsilon) \triangleq \max\left\{\lambda^{\varphi}_{\gamma,1}(\varepsilon), 1/\lambda^{\varphi}_{\gamma,N}(\varepsilon)\right\}, \tag{96}$$

where $\lambda^{\varphi}_{\gamma,i}(\varepsilon)$ is the i-th ordered eigenvalue of the matrix $[\widehat{\mathbf{V}}^{\varphi}_{\gamma}(X)]^{-1}\widehat{\mathbf{V}}^{\varphi}_{\gamma}(Z_\varepsilon)$, s.t. $\lambda^{\varphi}_{\gamma,1}(\varepsilon) \geq \cdots \geq \lambda^{\varphi}_{\gamma,N}(\varepsilon)$. Note that $BP^{\varphi}_{\gamma}(0) = 1$.

Fig. 2(a) reports the BP curves of the real R-estimator in Eq. (38) built upon the *van der Waerden* and three t_ν- scores ($\nu = 0.1, 1, 5$). Since $BP^{\varphi}_{\gamma}(\varepsilon)$ depends on X and X_ε, we plot its averaged value over 10^4 realizations of these data sets. For the sake of comparison, we report also the BP value of Tyler's estimator. The BP of the non-robust Sample Covariance Matrix (SCM) estimator explodes to 10^{17} as soon as $\varepsilon \neq 0$, so we do not include it in the plot. As for the complex case, all the BP curves, related to the R-estimator in Eq. (38) are bounded (w.r.t. the one of the non robust SCM) and close to the Tyler's one for every value of ε.

Let us now focus on the EIF. Similarly to the complex case discussed in our paper, the EIF can be defined as:

$$EIF^{\varphi}_{\gamma} \triangleq (L+1)||\widehat{\mathbf{V}}^{\varphi}_{\gamma}(X) - \widehat{\mathbf{V}}^{\varphi}_{\gamma}(X, \tilde{\mathbf{x}})||_F, \tag{97}$$

where $\tilde{\mathbf{x}}$ is an outlier distributed according to the pdf $q_X(\varrho)$ defined in Eq. (95). We refer the reader to the main paper for additional discussion on the definition of the EIF in Eq. (97). In Fig. 2(b), we report the EIF of the real R-estimator in Eq. (38) built upon the *van der Waerden* and three t_ν- scores ($\nu = 0.1, 1, 5$). As benchmark, the EIF of the Tyler's estimator is adopted since it is known that the relevant IF is continuous and bounded [17]. On the other hand, the EIF of the non-robust SCM grows rapidly to 10^4 as the norm of the outlier $\tilde{\mathbf{x}}$ increases (i.e. when $\varrho \to 0$), so we do not include it in the plot. As for the complex case, Fig. 2(b) shows that the EIFs of the R-estimator Eq. (38) remain bounded and close to the one of the Tyler's estimator for arbitrary large vale of $||\tilde{\mathbf{x}}||$ ($\varrho \to 0$).

REFERENCES

[1] L. Le Cam, "Locally asymptotically normal families of distributions," in *Univ. California Publ. Statist.*, vol. 3, 1960, pp. 37–98.

[2] L. Le Cam and G. L. Yang, *Asymptotics in Statistics: Some Basic Concepts (second edition)*. Springer series in statistics, 2000.

[3] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998.

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

[5] A. W. van der Vaart, *Asymptotic Statistics*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[6] M. Hallin and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape I. optimal rank-based tests for sphericity," *The Annals of Statistics*, vol. 34, no. 6, pp. 2707–2756, 2006.

[7] 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.

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

[9] J. Frontera-Pons, J. Ovarlez, and F. Pascal, "Robust ANMF detection in noncentered impulsive background," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1891–1895, Dec 2017.

[10] A. Tsiatis, *Semiparametric Theory and Missing Data*. Springer series in statistics, 2006.

[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] S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions," *Journal of Multivariate Analysis*, vol. 11, no. 3, pp. 368 – 385, 1981.

[13] M. Hallin and B. J. M. Werker, "Semi-parametric efficiency, distribution-freeness and invariance," *Bernoulli*, vol. 9, no. 1, pp. 137–165, 2003.

[14] S. Fortunati, F. Gini, M. Greco, A. M. Zoubir, and M. Rangaswamy, "A fresh look at the semiparametric Cramér-Rao bound," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 261–265.

[15] S. Fortunati and F. Gini, "Misspecified and semiparametric lower bounds and their application to inference problems with complex elliptically symmetric distributed data (part II)," Tutorial at EUSIPCO 2019. [Online]. Available: https://github.com/StefanoFor

[16] 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.

[17] 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.

[18] L. Dümberg and D. E. Tyler, "On the breakdown properties of some multivariate m-functionals," *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 247–264, 2005.

[19] C. Croux, "Limit behavior of the empirical influence function of the median," *Statistics & Probability Letters*, vol. 37, no. 4, pp. 331 – 340, 1998.