

# Semiparametric Inference and Lower Bounds for Real Elliptically Symmetric Distributions

Stefano Fortunati, *Member, IEEE*, Fulvio Gini, *Fellow, IEEE*, Maria S. Greco, *Fellow, IEEE*,  
Abdelhak M. Zoubir, *Fellow, IEEE* and Muralidhar Rangaswamy, *Fellow, IEEE*

**Abstract**—This paper has a twofold goal. The first aim is to provide a deeper understanding of the family of the Real Elliptically Symmetric (RES) distributions by investigating their intrinsic semiparametric nature. The second aim is to derive a semiparametric lower bound for the estimation of the parametric component of the model. The RES distributions represent a semiparametric model where the parametric part is given by the mean vector and by the scatter matrix while the non-parametric, infinite-dimensional, part is represented by the density generator. Since, in practical applications, we are often interested only in the estimation of the parametric component, the density generator can be considered as nuisance. The first part of the paper is dedicated to conveniently place the RES distributions in the framework of the semiparametric group models. The second part of the paper, building on the mathematical tools previously introduced, the Constrained Semiparametric Cramér-Rao Bound (CSCRb) for the estimation of the mean vector and of the constrained scatter matrix of a RES distributed random vector is introduced. The CSCRb provides a lower bound on the Mean Squared Error (MSE) of any robust  $M$ -estimator of mean vector and scatter matrix when no a-priori information on the density generator is available. A closed form expression for the CSCRb is derived. Finally, in simulations, we assess the statistical efficiency of the Tyler’s and Huber’s scatter matrix  $M$ -estimators with respect to the CSCRb.

**Index Terms**—Parametric model, semiparametric model, Cramér-Rao Bound, Semiparametric Cramér-Rao Bound, Elliptically Symmetric distributions, scatter matrix estimation, robust estimation.

## I. INTRODUCTION

A prerequisite for any statistical inference method is the notion of a *statistical model*, say  $\mathcal{P}$ , i.e. a collection, or a family, of probability density functions (pdfs) that is able to characterize random phenomena based on their observations. The most widely used models are *parametric models*. A parametric model is a family of pdfs parametrized by the elements of a subset  $\Gamma$  of a finite-dimensional Euclidean space  $\mathbb{R}^q$ . The popularity of parametric models is due to the ease of derivation of inference algorithms. On the other hand, a major drawback is their “narrowness” that can lead to misspecification problems [1], [2]. Their counterpart are *nonparametric models*,

a wide family of pdfs that can be required to satisfy some functional constraints, e.g. symmetry, smoothness or moment constraints. While the use of a nonparametric model minimizes the risk of model misspecification, the amount of data needed for nonparametric inference may represent an insurmountable obstacle in practical applications. *Semiparametric models* have been introduced as a compromise between the “narrowness” of parametric models and the cost of using nonparametric ones [3]. More formally, let  $\mathbf{x}$  be a random vector taking values in the sample space  $\mathcal{X} \subseteq \mathbb{R}^N$ . Then, a semiparametric model  $\mathcal{P}_{\gamma,l}$  is a family of pdfs parametrized by a finite-dimensional parameter vector of interest  $\gamma \in \Gamma \subseteq \mathbb{R}^q$ , along with an infinite-dimensional nuisance parameter  $l \in \mathcal{L}$ , where  $\mathcal{L}$  is a set of functions:

$$\mathcal{P}_{\gamma,l} \triangleq \{p_X | p_X(\mathbf{x} | \gamma, l), \gamma \in \Gamma, l \in \mathcal{L}\}. \quad (1)$$

There is a rich statistical literature on semiparametric models and their applications. For a comprehensive and detailed list of the main contributions in this field, we refer the reader to [3] and [4] and to the seminal book [5]. However, this profound theoretical understanding of semiparametric models has not been fully exploited in Signal Processing (SP) problems as yet. Two, among the very few, examples of SP applications of the semiparametric inference are the references [6] and [7], where the semiparametric theory has been applied to blind source separation and nonlinear regression, respectively.

This paper aims at improving the understanding of potential applications of semiparametric models. Specifically, we focus our attention on the joint estimation of the mean vector  $\boldsymbol{\mu}$  and of the (constrained) scatter matrix  $\boldsymbol{\Sigma}$  in the family of Real Elliptically Symmetric (RES) distributions by providing a closed form expression, up to a (numerically performed) singular value decomposition, for the Constrained Semiparametric Cramér-Rao Bound (CSCRb) on the MSE of any robust estimator of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ . As we will discuss below, a constraint on  $\boldsymbol{\Sigma}$  is required to avoid the scale ambiguity that characterizes the definition of scatter matrix in RES distributions. The RES class represents a wide family of distributions that includes the Gaussian, the  $t$ , the Generalized Gaussian and all the real Compound-Gaussian distributions as special cases ([8]–[14], and [15, Ch. 4]). The elliptical distributions are of fundamental importance in many practical applications since they can be successfully exploited to statistically characterize the non-Gaussian behavior of noisy data. Moreover, as we will discuss below, RES distributions represent an example of a semiparametric model where the parametric part is represented by the mean vector and by the scatter matrix. They should be

S. Fortunati is with Università di Pisa, Dipartimento di Ingegneria dell’Informazione, Pisa, Italy and with Technische Universität Darmstadt, Signal Processing Group, Darmstadt, Germany (e-mail: stefano.fortunati@iet.unipi.it).

F. Gini and M. S. Greco are with Università di Pisa, Dipartimento di Ingegneria dell’Informazione, Pisa, Italy (e-mail: f.gini,m.greco@iet.unipi.it).

A. M. Zoubir is with Technische Universität Darmstadt, Signal Processing Group, Darmstadt, Germany (e-mail: zoubir@spg.tu-darmstadt.de).

M. Rangaswamy is with U.S. AFRL, Sensors Directorate, Wright-Patterson AFB, OH, USA (e-mail: muralidhar.rangaswamy@us.af.mil).

estimated in the presence of an infinite-dimensional nuisance parameter, i.e. the density generator, which is generally unknown.

This paper is the natural follow on of our previous work [16]. In [16], we provided, in a tutorial and accessible manner, a general introduction to semiparametric inference framework and to the underlying mathematical tools needed for its development. Moreover, the extension of the classical CRB in the presence of a finite-dimensional nuisance deterministic vector to the SCRb where the nuisance parameter belongs to a certain, infinite-dimensional, function space has been reported in Theorem 1 in [16]. As discussed in the statistical literature and summarized in [16], this generalization can be carried out by means of three key elements:

- the Hilbert space  $\mathcal{H}^q$  of all the  $q$ -dimensional, zero-mean, vector-valued function of the data vector,
- a notion of *tangent space*  $\mathcal{T}$  for a statistical model,
- an orthogonal projection operator on  $\mathcal{T}$ , i.e.  $\Pi(\cdot|\mathcal{T})$ .

A formal definition of the Hilbert space  $\mathcal{H}^q$  and of the projection operator  $\Pi(\cdot|\mathcal{T})$  can be found in Appendix A, while the tangent space for both parametric and semiparametric models has been defined, in a tutorial manner, in [16].

This paper aims at investigating the applications of the general concepts introduced in [16] to the semiparametric model generated by RES distributions. We start by introducing the semiparametric group model and then we continue by showing that the RES class actually possesses this structure. Building on the mathematical framework that characterizes semiparametric group models and, in particular, their tangent space and projection operator, we then show how to derive the CSCRb for the joint estimation of the mean vector  $\boldsymbol{\mu}$  and the scatter matrix  $\boldsymbol{\Sigma}$  of a RES distributed random vector.

The problem of establishing a semiparametric lower bound for the joint estimation of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  in the RES class has been investigated firstly by Bickel in [17], where a bound on the estimation error of the inverse of the scatter matrix has been derived. More discussions and analyses have also been presented in [5] (Sec. 4.2 and Sec. 6.3). More recently, in a series of papers ([18], [19], [20] and [21]), Hallin and Paindaveine rediscovered the RES class as a semiparametric model and, by using two approaches based on Le Cam's theory [22] and on rank-based invariance [23], they presented the SCRb for the joint estimation of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  in its most general form. However, even if valuable and profound, Hallin and Paindaveine's work requires a deep understanding of Le Cam's theory on Local Asymptotic Normality [22]. For this reason, starting from the results in [5] (Sec. 4.2 and Sec. 6.3), we propose here an alternative derivation of the SCRb by using a simpler, even if less general, approach.

Along with the derivation of a lower bound on the estimation performance, we always have to specify the class of estimators to which such bound applies. It can be shown that the SCRb is a lower bound to the MSE of any *regular* and *asymptotic linear* (RAL) estimator (see [5, Sec. 2.2], [17], [24], [25], [26], [27, Ch. 3] and [28, Ch. 4] for additional details). Even if we do not address this issue here, it must be highlighted that the class of RAL estimators is a very wide

family that encompasses the Maximum Likelihood estimator and all the  $R$ -,  $S$ -, and in particular,  $M$ -robust estimators.

The rest of the paper is organized as follows. In Sec. II the semiparametric group model is presented with a particular focus on the calculation of the tangent space and projection operator. Sec. III collects the basic notions on RES distributions and their intrinsic semiparametric-group structure is investigated. The step-by-step derivation of the CSCRb for the estimation of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  is provided in Sec. IV. The efficiency of the Sample Covariance Matrix and of two robust scatter matrix  $M$ -estimators, Tyler's and the Huber's estimators, is assessed in Sec. V using the previously derived CSCRb. Finally, some concluding remarks are collected in Sec. VI.

*Notation:* Throughout this paper, italics indicates scalars or scalar-valued functions ( $a, A$ ), lower case and upper case boldface indicate column vectors ( $\mathbf{a}$ ) and matrices ( $\mathbf{A}$ ) respectively. Note that, since we deal with Hilbert spaces, the word "vector" indicates both Euclidean vectors and vector-valued functions. For clarity, we indicate sometimes a vector-valued function as  $\mathbf{a} \equiv \mathbf{a}(\mathbf{x})$ . Each entry of a matrix  $\mathbf{A}$  is indicated as  $a_{i,j} \triangleq [\mathbf{A}]_{i,j}$ . The superscript  $T$  indicates the transpose operator. Finally, for random variables or vectors, the notation  $=_d$  stands for "has the same distribution as".

## II. THE SEMIPARAMETRIC GROUP MODELS

This section introduces a particular semiparametric model, i.e. the *semiparametric group model*. As the name suggests, this class of semiparametric models is generated by the action of a group of invertible transformations on a random vector whose pdf is allowed to vary in a given set. As we will show in the sequel, this group-based data generating process allows for an easy calculation of the nuisance tangent space and of the orthogonal projection operator. Before introducing the definition of this class of semiparametric models, let us first introduce some related notation.

Let  $\mathcal{A}$  be a group of invertible transformations from  $\mathbb{R}^N$  into itself. Suppose that each transformation  $\alpha \in \mathcal{A}$  can be parametrized by means of a real vector  $\boldsymbol{\gamma} \in \Gamma \subseteq \mathbb{R}^q$ , i.e.

$$\mathcal{A} = \{\alpha|\alpha(\cdot; \boldsymbol{\gamma}) \triangleq \alpha_{\boldsymbol{\gamma}}(\cdot); \boldsymbol{\gamma} \in \Gamma\}. \quad (2)$$

We will indicate with  $\alpha_{\boldsymbol{\gamma}}^{-1}$  the inverse of  $\alpha_{\boldsymbol{\gamma}}$ . The operation  $\alpha_{\boldsymbol{\gamma}_2} \circ \alpha_{\boldsymbol{\gamma}_1}$  denotes the composition of  $\alpha_{\boldsymbol{\gamma}_1}$  and  $\alpha_{\boldsymbol{\gamma}_2}$  that can be explicitly expressed as  $(\alpha_{\boldsymbol{\gamma}_2} \circ \alpha_{\boldsymbol{\gamma}_1})(\cdot) \triangleq \alpha_{\boldsymbol{\gamma}_2}(\alpha_{\boldsymbol{\gamma}_1}(\cdot))$ . Finally,  $\boldsymbol{\gamma}_e$  indicates the parameter vector that characterizes the identity transformation  $\alpha_{\boldsymbol{\gamma}_e}$ , such that  $\alpha_{\boldsymbol{\gamma}_e}(\cdot) = \cdot$ .

**Definition II.1.** (see [5, Sec. 4.2]) Let  $\mathbf{z} \in \mathbb{R}^N$  be a real-valued random vector with pdf  $p_Z$ , i.e.  $\mathbf{z} \sim p_Z(\mathbf{z})$ . The parametric group model  $\mathcal{P}_{\boldsymbol{\gamma}}$ , generated by the action of the group of invertible parametric transformations  $\mathcal{A}$ , given in (2), on the random vector  $\mathbf{z} \sim p_Z(\mathbf{z})$ , is the set of parametric pdfs of the transformed random vector  $\alpha_{\boldsymbol{\gamma}}(\mathbf{z}) = \mathbf{x} \sim p_X(\mathbf{x}|\boldsymbol{\gamma})$ . Specifically,  $\mathcal{P}_{\boldsymbol{\gamma}}$  can be explicitly expressed as:

$$\mathcal{P}_{\boldsymbol{\gamma}} = \{p_X|p_X(\mathbf{x}|\boldsymbol{\gamma}) = |\mathbf{J}(\alpha_{\boldsymbol{\gamma}}^{-1})(\mathbf{x})|p_Z(\alpha_{\boldsymbol{\gamma}}^{-1}(\mathbf{x})); \boldsymbol{\gamma} \in \Gamma\}, \quad (3)$$

where  $[\mathbf{J}(\alpha_{\boldsymbol{\gamma}}^{-1})(\mathbf{x})]_{i,j} \triangleq \partial[\alpha^{-1}(\mathbf{x}; \boldsymbol{\gamma})]_i / \partial \gamma_j$  is the Jacobian matrix of the inverse transformation  $\alpha_{\boldsymbol{\gamma}}^{-1}$  and  $|\cdot|$  defines the

absolute value of the determinant of a matrix. The generalization to semiparametric models can be obtained by allowing the pdf  $p_Z$  to vary within a large set of density functions  $\mathcal{L}$ . Consequently, a semiparametric group model generated by the parametric group  $\mathcal{A}$  in (2) can be expressed as:

$$\mathcal{P}_{\gamma,p_Z} = \{p_X|p_X(\mathbf{x}|\gamma,p_Z) = |\mathbf{J}(\alpha_\gamma^{-1})(\mathbf{x})|p_Z(\alpha_\gamma^{-1}(\mathbf{x})); \gamma \in \Gamma, p_Z \in \mathcal{L}\}. \quad (4)$$

Using the notation introduced in [16], the *actual* “semiparametric vector” is indicated as  $(\gamma_0^T, p_{Z,0})^T$  and, consequently, the true pdf is given by:

$$p_0(\mathbf{x}) \triangleq p_X(\mathbf{x}|\gamma_0, p_{Z,0}) = |\mathbf{J}(\alpha_{\gamma_0}^{-1})(\mathbf{x})|p_{Z,0}(\alpha_{\gamma_0}^{-1}(\mathbf{x})). \quad (5)$$

Moreover, from now on, we denote by  $E_0\{\cdot\}$  the expectation operator with respect to the *true* pdf  $p_0(\mathbf{x})$ . Note that the word “true” here indicates the actual pdf, and consequently the actual semiparametric vector, that characterizes the data.

As mentioned before, the most useful feature of a semiparametric group model is the fact that its underlying group structure allows for a convenient derivation of the nuisance tangent space and of the relevant projection operator. The following proposition formalizes this concept.

Let  $\mathcal{P}_{\gamma,p_Z}$  be a semiparametric group model defined in (4). Let  $\mathcal{T}_{p_Z,0}(\gamma)$  be the semiparametric nuisance tangent space of  $\mathcal{P}_{\gamma,p_Z}$  evaluated at  $(\gamma^T, p_{Z,0})^T$ , where  $\gamma$  is a generic element of the finite-dimensional parameter space  $\Gamma$ . Let us indicate by  $\mathcal{T}_{p_Z,0}(\gamma_e)$  the semiparametric nuisance tangent space of  $\mathcal{P}_{\gamma,p_Z}$  evaluated at  $(\gamma_e^T, p_{Z,0})^T$ , where, as defined before,  $\gamma_e$  is the parameter vector that characterizes the identity transformation.

**Proposition II.1.** *Let  $\mathbf{t}$  be a generic  $q$ -dimensional vector-valued function belonging to  $\mathcal{T}_{p_Z,0}(\gamma_e)$ , then the semiparametric nuisance tangent space  $\mathcal{T}_{p_Z,0}(\gamma)$  can be obtained from  $\mathcal{T}_{p_Z,0}(\gamma_e)$  as follows:*

$$\mathcal{T}_{p_Z,0}(\gamma) = \{\mathbf{t} \circ \alpha_\gamma^{-1} | \mathbf{t} \in \mathcal{T}_{p_Z,0}(\gamma_e)\}, \forall \gamma \in \Gamma. \quad (6)$$

Moreover, let  $\mathbf{1}$  be a generic  $q$ -dimensional vector-valued function in  $\mathcal{H}^q$ , then the projection operator  $\Pi(\cdot|\mathcal{T}_{p_Z,0}(\gamma))$  on  $\mathcal{T}_{p_Z,0}(\gamma)$  (see Appendix A) can be obtained from the projection operator  $\Pi(\cdot|\mathcal{T}_{p_Z,0}(\gamma_e))$  on  $\mathcal{T}_{p_Z,0}(\gamma_e)$  as follows:

$$\Pi(\mathbf{1}|\mathcal{T}_{p_Z,0}(\gamma)) = \Pi(\mathbf{1} \circ \alpha_\gamma | \mathcal{T}_{p_Z,0}(\gamma_e)) \circ \alpha_\gamma^{-1}, \forall \gamma \in \Gamma. \quad (7)$$

The proof can be found in [5, Sec. 4.2, Lemma 3].

It is worth noticing that Proposition II.1 can be directly used to derive the nuisance tangent space at the true semiparametric vector  $(\gamma_0^T, p_{Z,0})^T$ , i.e.  $\mathcal{T}_{p_Z,0} \equiv \mathcal{T}_{p_Z,0}(\gamma_0)$  and the relevant projection operator  $\Pi(\cdot|\mathcal{T}_{p_Z,0})$ . This can be done by evaluating the relations (6) and (7) at the true parameter vector of interest  $\gamma_0$ . As discussed below, Proposition II.1 is of fundamental importance for the derivation of the Semiparametric Cramér-Rao Bound (SCR) for  $\gamma_0$  in the class of RES distributions.

### III. THE FAMILY OF RES DISTRIBUTIONS AS A SEMIPARAMETRIC MODEL

In this section, the semiparametric nature of the family of RES distributions is investigated. In particular, we show that the RES class can be conveniently interpreted as a

semiparametric group model. Here we restrict the discussion to the *absolutely continuous case* [14, Sec. III.D], i.e. we always assume that a RES distributed random vector admits a pdf. In what follows, we exploit the definition of the RES class provided in [29] and [9, Ch. 3], since it is particularly useful for our aims.

#### A. Spherically Symmetric (SS) distributions

As a prerequisite for the definition of the RES class, Spherically Symmetric (SS) distributions need to be introduced first.

**Definition III.1.** *Let  $\mathbf{z} \in \mathbb{R}^N$  be a real-valued random vector and let  $\mathcal{O}$  be the set of all the orthogonal transformations such that:*

$$\mathcal{O} \ni O : \mathbb{R}^N \rightarrow \mathbb{R}^N \quad (8)$$

$$\mathbf{z} \mapsto O(\mathbf{z}) = \mathbf{O}\mathbf{z},$$

for any orthogonal matrix  $\mathbf{O}$ , i.e. for any  $\mathbf{O} \in \mathbb{R}^{N \times N}$  such that  $\mathbf{O}^T \mathbf{O} = \mathbf{O} \mathbf{O}^T = \mathbf{I}$ . Then,  $\mathbf{z}$  is said to be SS-distributed if its distribution is invariant to any orthogonal transformations  $\mathbf{O} \in \mathcal{O}$ , i.e.

$$\mathbf{z} =_d \mathbf{O}\mathbf{z}. \quad (9)$$

We indicate with  $\mathcal{S}$  the class of all SS-distributions.

As a consequence, the following properties hold true (see, e.g. [29] or [9, Ch. 3] for the proof):

P1) The SS-distributed random vector  $\mathbf{z} \sim SS(g)$  has a pdf given by:

$$p_Z(\mathbf{z}) = 2^{-N/2} g(\|\mathbf{z}\|^2), \quad (10)$$

where  $\mathcal{G} \ni g$ , is a function, called *density generator*, that depends on  $\mathbf{z}$  only through  $\|\mathbf{z}\|$  and

$$\mathcal{G} = \left\{ g : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \mid \int_0^\infty t^{N/2-1} g(t) dt < \infty \right\} \quad (11)$$

where the integrability condition in (11) is required to guarantee the integrability of  $p_Z(\mathbf{z})$  (see [9, eq. 3.25]). Consequently, the set of all SS pdfs can be described as:

$$\mathcal{S} = \left\{ p_Z | p_Z(\mathbf{z}) = 2^{-N/2} g(\|\mathbf{z}\|^2), \forall g \in \mathcal{G} \right\}. \quad (12)$$

P2) Let  $s_N \triangleq 2\pi^{N/2}/\Gamma(N/2)$  be the surface area of the unit sphere in  $\mathbb{R}^N$ , then the pdf of the random variables  $\mathcal{Q} =_d \|\mathbf{z}\|^2$  and  $\mathcal{R} \triangleq \sqrt{\mathcal{Q}}$ , called *2nd-order modular variate* and *modular variate* respectively [14], are given by:

$$p_{\mathcal{Q}}(q) = s_N 2^{-N/2-1} q^{N/2-1} g(q), \quad (13)$$

$$p_{\mathcal{R}}(r) = s_N 2^{-N/2} r^{N-1} g(r^2). \quad (14)$$

P3) *The Stochastic Representation Theorem.* Let  $\mathbf{u} \sim \mathcal{U}(\mathbb{R}S^N)$  be a random vector uniformly distributed on the real unit sphere of dimension  $N$ , indicated as  $\mathbb{R}S^N$ . If  $\mathbf{z} \in \mathbb{R}^N$  is SS-distributed with pdf given by (10), then:

$$\mathbf{z} =_d \sqrt{\mathcal{Q}} \mathbf{u} =_d \mathcal{R} \mathbf{u}, \quad (15)$$

where  $\mathcal{Q} \sim p_{\mathcal{Q}}(q)$  in (13),  $\mathcal{R} \sim p_{\mathcal{R}}(r)$  in (14). Moreover,  $\mathcal{Q}$  and  $\mathbf{u}$  (or  $\mathcal{R}$  and  $\mathbf{u}$ ) are independent. Note that  $\mathbf{u}$  in

(15) satisfies the following three properties:  $\|\mathbf{u}\| = 1$ ,  $E\{\mathbf{u}\} = \mathbf{0}$  and  $E\{\mathbf{u}\mathbf{u}^T\} = N^{-1}\mathbf{I}$ .

P4) *Maximal Invariant Statistic.* The Stochastic Representation Theorem shows that there exists a one-to-one relationship between every  $\mathbf{z} \sim SS(g)$  and every couple  $(\mathcal{R}, \mathbf{u})$ . Moreover, it is easy to verify that the modular variate  $\mathcal{R}$  is a *maximal invariant statistic* for the set of the SS-distributed random vectors.<sup>1</sup>

We can now introduce the class of RES distributions as a *semiparametric group model*.

### B. The RES class as a semiparametric group model

At first, let us define the parameter space  $\Omega \subseteq \mathbb{R}^q$  of dimension  $q = N(N+3)/2$  as:

$$\Omega = \{\phi \in \mathbb{R}^q | \phi = (\boldsymbol{\mu}^T, \text{vecs}(\boldsymbol{\Sigma})^T)^T; \boldsymbol{\mu} \in \mathbb{R}^N, \boldsymbol{\Sigma} \in \mathcal{M}_N\}, \quad (16)$$

where  $\boldsymbol{\mu} \in \mathbb{R}^N$  is a real-valued  $N$ -dimensional vector while  $\boldsymbol{\Sigma}$  is an  $N \times N$  matrix belonging to the set  $\mathcal{M}_N$  of all the symmetric, positive-definite matrices of dimension  $N \times N$ . Note that the *vecs* operator maps the  $N \times N$  symmetric matrix  $\boldsymbol{\Sigma}$  in an  $N(N+1)/2$ -dimensional vector of the entries of the lower triangular sub-matrix of  $\boldsymbol{\Sigma}$  [31], [32]. We can now introduce the group  $\mathcal{A}$  of all affine transformations, parameterized by the parameter space  $\Omega$  in (16), such that

$$\begin{aligned} \mathcal{A} \ni \alpha_\phi : \mathbb{R}^N &\rightarrow \mathbb{R}^N, \quad \forall \phi \in \Omega \\ \mathbf{z} &\mapsto \alpha_\phi(\mathbf{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}. \end{aligned} \quad (17)$$

The identity element  $\alpha_{\phi_e}(\cdot)$  of the group  $\mathcal{A}$  is parameterized by the vector  $\phi_e = (\mathbf{0}^T, \text{vecs}(\mathbf{I})^T)^T$ , while the inverse transformation is given by:

$$\alpha_\phi^{-1}(\cdot) = \boldsymbol{\Sigma}^{-1/2}(\cdot - \boldsymbol{\mu}). \quad (18)$$

The class of RES distributions is defined as the class of distributions that is closed under the action of the group  $\mathcal{A}$  in (17) on any SS-distributed random vector. The next definition formalizes this statement.

**Definition III.2.** A real-valued random vector  $\mathbf{x} \in \mathbb{R}^N$  is said to be RES-distributed with mean value  $\boldsymbol{\mu}$  and scatter matrix  $\boldsymbol{\Sigma}$ , if it can be expressed as:

$$\mathbf{x} = \alpha_\phi(\mathbf{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z} =_d \boldsymbol{\mu} + \sqrt{Q}\boldsymbol{\Sigma}^{1/2}\mathbf{u}, \quad (19)$$

where  $\phi \in \Omega$  defined in (16),  $\mathbf{z} \sim SS(g)$  is an SS-distributed random vector, while  $\mathbf{u} \sim \mathcal{U}(\mathbb{R}S^N)$  and the 2nd-order modular variate  $Q$  have been already defined in (15). In particular:

$$Q =_d \|\mathbf{z}\|^2 = \|\alpha_\phi^{-1}(\mathbf{x})\|^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \triangleq Q. \quad (20)$$

<sup>1</sup>For the sake of clarity, let us recall the definition of maximal invariant statistic [30, Ch. 6]. Let  $\mathcal{D} = \{d\}$  be a group of one-to-one transformations on a sample space  $\mathcal{X}$  and let  $T$  be an invariant statistic such that  $T(\mathbf{x}) =_d T(d(\mathbf{x}))$ ,  $\forall \mathbf{x} \in \mathcal{X}$  and  $\forall d \in \mathcal{D}$ . Then,  $T$  is a maximal invariant on  $\mathcal{X}$  w.r.t.  $\mathcal{D}$  if  $T(\mathbf{x}_1) =_d T(\mathbf{x}_2)$  implies that  $\mathbf{x}_1 =_d d(\mathbf{x}_2)$ ,  $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$  and  $\forall d \in \mathcal{D}$ . Clearly, for any couple of SS-distributed random vectors  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , we have  $\|\mathbf{z}_1\| =_d \|\mathbf{z}_2\| \Rightarrow \mathbf{z}_1 =_d \mathbf{O}\mathbf{z}_2$ ,  $\forall \mathbf{O} \in \mathcal{O}$ , where  $\mathcal{O}$  is the group of orthogonal transformations defined in (8). Consequently,  $\mathcal{R} = \|\mathbf{z}\|$  is a maximal invariant statistic for the set of the SS-distributed random vectors.

We refer the reader to [29], [9, Ch. 3] and [14] for the proof.

Definition III.2 provides the link between the RES family and the semiparametric group model defined in Section II. As a consequence, the explicit expression of the pdf of an RES distributed random vector can be obtained as shown in the Definition II.1, Equation (4). In particular, the determinant of the Jacobian matrix of the inverse transformation in (18) is  $|\mathbf{J}(\alpha_\phi^{-1})| = |\boldsymbol{\Sigma}^{-1/2}| = |\boldsymbol{\Sigma}|^{-1/2}$ . Then, the pdf of any RES-distributed random vector  $\mathbf{x}$  can be obtained from the relevant SS-distributed random vector  $\mathbf{z}$ , i.e.  $p_Z(\mathbf{z})$  in (10), as <sup>2</sup>:

$$\begin{aligned} RES_N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, g) &\triangleq |\boldsymbol{\Sigma}|^{-1/2} p_Z(\alpha_\phi^{-1}(\mathbf{x})) \\ &= 2^{-N/2} |\boldsymbol{\Sigma}|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})), \quad \forall g \in \mathcal{G}. \end{aligned} \quad (21)$$

Moreover, the general description of a semiparametric group model given in (4) can be specialized for the RES case as:

$$\begin{aligned} \mathcal{P}_{\phi, g} &= \left\{ p_X | p_X(\mathbf{x} | \phi, g) = 2^{-N/2} |\boldsymbol{\Sigma}|^{-1/2} g(\|\alpha_\phi^{-1}(\mathbf{x})\|^2); \right. \\ &\quad \left. \phi \in \Omega, g \in \mathcal{G} \right\}, \end{aligned} \quad (22)$$

where  $\mathcal{G}$  is the set of density generators given in (11). Clearly, the mean vector of  $\mathbf{x} \sim RES_N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$  is given by  $E_0\{\mathbf{x}\} = \boldsymbol{\mu}$  while, if  $E\{Q\} < \infty$ , its covariance matrix  $\mathbf{M}$  is  $\mathbf{M} \triangleq E_0\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\} = N^{-1}E\{Q\}\boldsymbol{\Sigma}$

As extensively discussed in the literature on elliptically symmetric distributions, the representation of an RES distributed vector  $\mathbf{x}$  is not uniquely determined by (19). In fact,  $\mathbf{x} =_d \boldsymbol{\mu} + \sqrt{Q}\boldsymbol{\Sigma}^{-1/2}\mathbf{u} =_d \boldsymbol{\mu} + \sqrt{c^{-2}Q}(c\boldsymbol{\Sigma}^{-1/2})\mathbf{u}$ ,  $\forall c > 0$ . This scale ambiguity can also be seen as a consequence of the functional form of an RES pdf given in (21) since  $RES_N(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, g(t)) \equiv RES_N(\mathbf{x}; \boldsymbol{\mu}, c^2\boldsymbol{\Sigma}, g(t/c^2))$ ,  $\forall c > 0$ . To avoid this well-known identifiability problem, we impose the following constraint on the trace of  $\boldsymbol{\Sigma}$ , i.e.

$$c(\boldsymbol{\Sigma}) = \text{tr}(\boldsymbol{\Sigma}) - N = 0. \quad (23)$$

This constraint limits the parameter vector  $\phi \in \Omega$ , where  $\Omega$  is defined in (16), in a lower dimensional smooth manifold

$$\bar{\Omega} = \{\phi \in \Omega | \text{tr}(\boldsymbol{\Sigma}) = N\}, \quad (24)$$

whose dimension is  $\bar{q} = q - 1$ . Trace constraint is only an example of all the possible constrains that can be imposed on the scatter matrix to avoid scale ambiguity. For a deep and insightful analysis of the impact of the particular constraint on  $\boldsymbol{\Sigma}$  on the estimation performance, we refer to [20] and [21].

The properties of the semiparametric group model previously discussed can be exploited to derive the CSCRb for the estimation of the *constrained* parameter vector  $\phi = (\boldsymbol{\mu}^T, \text{vecs}(\boldsymbol{\Sigma})^T)^T \in \bar{\Omega}$ , where  $\bar{\Omega}$  given in (24).

<sup>2</sup>Note that the definition of the pdf of an RES distributed random vectors given here is consistent with the one proposed in [14] for CES distributed random vectors. The only difference is that, in our definition, the normalizing constant  $c_{N, g}$  introduced in [14], has been included in the density generator  $g(\cdot)$ .

#### IV. THE CONSTRAINED SEMIPARAMETRIC CRAMÉR-RAO BOUND FOR THE RES CLASS

This section is devoted to the derivation of a closed form expression of the CSCRb for the estimation of  $\phi \in \Omega$ . The theoretical foundation of the generalization of the Cramér-Rao inequality in the semiparametric framework can be found in [27, Theo. 4.1], [5, Sec. 3.4], [24] and [33]. Moreover, in [16], it is shown, in a tutorial manner, how the SCRb can be obtained as a result of a limit process of the classical CRb derived in the presence of a finite-dimensional nuisance parameter vector. Here, as mentioned before, we focus on the calculation of the SCRb for the particular case of the RES distributions.

As explained in the previous section, to avoid the scale ambiguity of the RES class, we need to put a constraint on the scatter matrix. In order to take this requirement into account, we propose in the sequel the extension of Theorem 1 in [16] to the case of constrained, finite-dimensional, parameter vector. In particular, suppose that the finite-dimensional parameter vector of interest  $\gamma_0 \in \Gamma \subseteq \mathbb{R}^q$  is required to satisfy  $k$  (with  $k < q$ ) continuously differentiable constraints ([34], [35], [36]):

$$\mathbf{c}(\gamma_0) = \mathbf{0}. \quad (25)$$

This set of constraints define a smooth manifold, of dimension  $\bar{q} = q - k$ , in the parameter space  $\Gamma$ , such that:

$$\bar{\Gamma} = \{\gamma \in \Gamma \subseteq \mathbb{R}^q | \mathbf{c}(\gamma) = \mathbf{0}\} \quad (26)$$

Moreover, suppose that the  $k \times q$  Jacobian matrix of the constraints, defined as  $[\mathbf{J}_c(\gamma)]_{i,j} \triangleq \partial c_i(\gamma) / \partial \gamma_j$  has full row rank for any  $\gamma \in \Gamma$  satisfying (25). Consequently, there exists a matrix  $\mathbf{U} \in \mathbb{R}^{q \times \bar{q}}$  whose columns form an orthonormal basis for the null space of  $\mathbf{J}_c(\gamma_0)$ , i.e.

$$\mathbf{J}_c(\gamma_0)\mathbf{U} = \mathbf{0}_{k \times \bar{q}}, \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}. \quad (27)$$

**Theorem IV.1.** *The Constrained Semiparametric Cramér-Rao Bound (CSCRb) for the estimation of the constrained finite-dimensional vector  $\gamma_0 \in \bar{\Gamma}$  in the presence of the nuisance function  $l_0 \in \mathcal{L}$  is given by:*

$$\text{CSCRb}(\gamma_0 | l_0) = \mathbf{U}(\mathbf{U}^T \bar{\mathbf{I}}(\gamma_0 | l_0) \mathbf{U})^{-1} \mathbf{U}^T, \quad (28)$$

where:

$$\bar{\mathbf{I}}(\gamma_0 | l_0) \triangleq E_0\{\bar{\mathbf{s}}_0 \bar{\mathbf{s}}_0^T\}, \quad (29)$$

is the semiparametric Fisher Information Matrix (SFIM) and  $\bar{\mathbf{s}}_0$  is the semiparametric efficient score vector defined as:

$$\bar{\mathbf{s}}_0 = \mathbf{s}_{\gamma_0} - \Pi(\mathbf{s}_{\gamma_0} | \mathcal{T}_{l_0}), \quad (30)$$

where  $\Pi(\mathbf{s}_{\gamma_0} | \mathcal{T}_{l_0})$  is the orthogonal projection of the score vector of the parameters of interest  $\mathbf{s}_{\gamma_0}$  on the semiparametric nuisance tangent space. Finally, matrix  $\mathbf{U}$  is defined in (27). Note that  $\bar{\mathbf{s}}_0$ ,  $\mathbf{s}_{\gamma_0}$  and  $\Pi(\mathbf{s}_{\gamma_0} | \mathcal{T}_{l_0})$  are  $q$ -dimensional functions of the observation vector  $\mathbf{x}$ .

The proof of the ‘‘unconstrained’’ part of this theorem can be found in [27, Theo. 4.1] and in [24], while a more abstract and general formulation can be found in [33] and in [5, Sec. 3.4]. The proof of the ‘‘constrained’’ part can be obtained through a

straightforward application of the approach discussed in [35] for the constrained CRb.

The remainder of this section is devoted to the evaluation of the CSCRb in (28) for the estimation of the mean vector  $\boldsymbol{\mu}$  and of the constrained scatter matrix  $\boldsymbol{\Sigma}$ , such that  $\text{tr}(\boldsymbol{\Sigma}) = N$ , of a RES distributed random vector  $\mathbf{x} \sim \text{RES}_N(\mathbf{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0)$  in the presence of the nuisance function  $g_0 \in \mathcal{G}$ . To this end, according to Theorem IV.1, we have to evaluate:

- the score vector of the parameters of interest  $\mathbf{s}_{\phi_0} \equiv \mathbf{s}_{\phi_0}(\mathbf{x})$ , where  $\phi_0 = (\boldsymbol{\mu}_0^T, \text{vecs}(\boldsymbol{\Sigma}_0)^T)^T \in \Omega$  given in (16),
- the projection operator  $\Pi(\mathbf{s}_{\phi_0} | \mathcal{T}_{g_0})$ , where  $\mathcal{T}_{g_0} \equiv \mathcal{T}_{g_0}(\phi_0)$  is the semiparametric nuisance tangent space for the RES distributions class evaluated at the true semiparametric vector  $(\phi_0^T, g_0)^T$ ,
- the semiparametric efficient score vector  $\bar{\mathbf{s}}_0 \equiv \bar{\mathbf{s}}_0(\mathbf{x})$  in (30) and the SFIM  $\bar{\mathbf{I}}(\phi_0 | g_0)$  in (29),
- the matrix  $\mathbf{U}$  in (27) and then the CSCRb in (28).

In what follows, we will provide this calculation step-by-step.

##### A. Evaluation of the score vector $\mathbf{s}_{\phi_0}(\mathbf{x})$

By definition, the score vector of the parameters of interest is given by:

$$\mathbf{s}_{\phi_0}(\mathbf{x}) = \nabla_{\phi} \ln p_X(\mathbf{x} | \phi_0, g_0) = \begin{pmatrix} \mathbf{s}_{\boldsymbol{\mu}_0}(\mathbf{x}) \\ \mathbf{s}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{x}) \end{pmatrix}, \quad (31)$$

where  $\phi_0 \in \Omega$  and  $\Omega$  is the *unconstrained* parameter space defined in (16) and:

$$\mathbf{s}_{\boldsymbol{\mu}_0}(\mathbf{x}) = \nabla_{\boldsymbol{\mu}} \ln p_X(\mathbf{x} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0), \quad (32)$$

$$\mathbf{s}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{x}) = \nabla_{\text{vecs}(\boldsymbol{\Sigma})} \ln p_X(\mathbf{x} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0), \quad (33)$$

and where  $\boldsymbol{\mu}_0$ ,  $\boldsymbol{\Sigma}_0$  and  $g_0$  represents the true mean vector, the true scatter matrix and the true density generator, respectively. By substituting in (32) the explicit expression of  $p_X(\mathbf{x} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0)$  given in (21) and by exploiting the differentiation rules with respect to vector and matrices provided e.g. in [37, Ch. 8], we have that:

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\mu}_0}(\mathbf{x}) &= -2\psi_0(Q_0)\boldsymbol{\Sigma}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0) \\ &= {}_d -2\sqrt{Q}\psi_0(Q)\boldsymbol{\Sigma}_0^{-1/2}\mathbf{u} \end{aligned} \quad (34)$$

where, according to (20),

$$Q_0 = \|\alpha_{\phi_0}^{-1}(\mathbf{x})\|^2 = (\mathbf{x} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0) = {}_d Q, \quad (35)$$

and the last two equalities follow directly from the Stochastic Representation of a RES vector given in (19) and

$$\psi_0(t) \triangleq \frac{1}{g_0(t)} \frac{dg_0(t)}{dt}. \quad (36)$$

Similarly, the term  $\mathbf{s}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{x})$  in (33) can be evaluated by applying the rules of the differential matrix calculus detailed in [37, Ch. 8] and by using standard properties of the Kronecker product and of the vec operator ([37, Ch. 15], [31,32]) as:

$$\begin{aligned} \mathbf{s}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{x}) &= -\mathbf{D}_N^T (2^{-1} \text{vec}(\boldsymbol{\Sigma}_0^{-1}) + \\ &\quad + \psi_0(Q_0)\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1} \text{vec}((\mathbf{x} - \boldsymbol{\mu}_0)(\mathbf{x} - \boldsymbol{\mu}_0)^T)) \\ &= {}_d -\mathbf{D}_N^T (2^{-1} \text{vec}(\boldsymbol{\Sigma}_0^{-1}) + \\ &\quad + Q\psi_0(Q)\boldsymbol{\Sigma}_0^{-1/2} \otimes \boldsymbol{\Sigma}_0^{-1/2} \text{vec}(\mathbf{u}\mathbf{u}^T)) \end{aligned} \quad (37)$$

As before, the second equality follows from the Stochastic Representation (19) while  $\mathbf{D}_N$  is the  $N^2 \times N(N+1)/2$  duplication matrix that is implicitly defined by the equality  $\mathbf{D}_N \text{vecs}(\mathbf{A}) = \text{vec}(\mathbf{A})$  for any  $N^2 \times N^2$  symmetric matrix  $\mathbf{A}$  ([31], [32]). See also [38]–[40] for similar calculation.

### B. Evaluation of the projection operator $\Pi(\mathbf{s}_{\phi_0}|\mathcal{T}_{g_0})$

In order to obtain the explicit expression of the projection operator, we will exploit the fact that the RES class is a semiparametric group model. In particular,  $\Pi(\mathbf{s}_{\phi_0}|\mathcal{T}_{g_0})$  can be obtained by specializing Proposition II.1 for the RES distributions.

For the sake of clarity, let us recall that, according to the definition of the group  $\mathcal{A}$  of the affine transformations (17), for any  $\phi_0 \in \Omega$ , we have:

$$\alpha_{\phi_0}(\mathbf{z}) \triangleq \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0^{1/2} \mathbf{z}, \quad \mathbf{z} \sim SS(g_0), \quad (38)$$

$$\alpha_{\phi_e}(\mathbf{z}) \triangleq \mathbf{0} + \mathbf{I}^{1/2} \mathbf{z} = \mathbf{z}, \quad \mathbf{z} \sim SS(g_0), \quad (39)$$

$$\alpha_{\phi_0}^{-1}(\mathbf{x}) \triangleq \boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x} - \boldsymbol{\mu}_0), \quad \mathbf{x} \sim RES_N(\mathbf{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0). \quad (40)$$

Let  $\mathcal{T}_{g_0} \equiv \mathcal{T}_{g_0}(\phi_0)$  and  $\mathcal{T}_{S_0} \equiv \mathcal{T}_{g_0}(\phi_e)$  be the semiparametric nuisance tangent spaces of the RES class evaluated at the true semiparametric vector  $(\phi_0^T, g_0)^T$  and at  $(\phi_e^T, g_0)^T$ , where  $\phi_e = (\mathbf{0}^T, \text{vecs}(\mathbf{I})^T)^T$  is the vector that characterizes the identity transformation given in (39). Note that, since it is evaluated at the identity transformation,  $\mathcal{T}_{S_0}$  can be interpreted as the tangent space of the SS distribution class evaluated at the true  $g_0$ , and this explains the chosen notation. Now, by directly applying Proposition II.1, we have that the projection of the score vector of the parameters of interest  $\mathbf{s}_{\phi_0}(\mathbf{x})$  on the tangent space  $\mathcal{T}_{g_0}$  can be expressed as:

$$\Pi(\mathbf{s}_{\phi_0}(\mathbf{x})|\mathcal{T}_{g_0})(\mathbf{x}) = (\Pi((\mathbf{s}_{\phi_0} \circ \alpha_{\phi_0})(\mathbf{z})|\mathcal{T}_{S_0}) \circ \alpha_{\phi_0}^{-1})(\mathbf{x}), \quad (41)$$

where  $\mathbf{x} \sim RES_N(\mathbf{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, g_0)$  and  $\mathbf{z} \sim SS(g_0)$ .

In what follows, we show how to obtain an explicit expression of the projection in (41).

1) *Calculation of  $(\mathbf{s}_{\phi_0} \circ \alpha_{\phi_0})(\mathbf{z})$ :* From a visual inspection of the expressions in (34) and (37), we notice that they can be rewritten as a function of the inverse transformation  $\alpha_{\phi_0}^{-1}(\mathbf{x})$ , i.e.:

$$\mathbf{s}_{\boldsymbol{\mu}_0}(\mathbf{x}) = (\tilde{\mathbf{s}}_{\boldsymbol{\mu}_0} \circ \alpha_{\phi_0}^{-1})(\mathbf{x}) = -2\psi_0(Q_0)\boldsymbol{\Sigma}_0^{-1/2}\alpha_{\phi_0}^{-1}(\mathbf{x}), \quad (42)$$

and

$$\begin{aligned} \mathbf{s}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{x}) &= (\tilde{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)} \circ \alpha_{\phi_0}^{-1})(\mathbf{x}) = -\mathbf{D}_N^T (2^{-1}\text{vec}(\boldsymbol{\Sigma}_0^{-1}) + \\ &+ \psi_0(Q_0)\boldsymbol{\Sigma}_0^{-1/2} \otimes \boldsymbol{\Sigma}_0^{-1/2}\text{vec}(\alpha_{\phi_0}^{-1}(\mathbf{x})\alpha_{\phi_0}^{-1}(\mathbf{x})^T)). \end{aligned} \quad (43)$$

Then, we have that:

$$\begin{aligned} (\mathbf{s}_{\phi_0} \circ \alpha_{\phi_0})(\mathbf{z}) &= \begin{pmatrix} (\tilde{\mathbf{s}}_{\boldsymbol{\mu}_0} \circ \alpha_{\phi_0}^{-1} \circ \alpha_{\phi_0})(\mathbf{z}) \\ (\tilde{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)} \circ \alpha_{\phi_0}^{-1} \circ \alpha_{\phi_0})(\mathbf{z}) \end{pmatrix} \\ &= \begin{pmatrix} \tilde{\mathbf{s}}_{\boldsymbol{\mu}_0}(\mathbf{z}) \\ \tilde{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{z}) \end{pmatrix}, \end{aligned} \quad (44)$$

and

$$\tilde{\mathbf{s}}_{\boldsymbol{\mu}_0}(\mathbf{z}) = -2\psi_0(Q_0)\boldsymbol{\Sigma}_0^{-1/2}\mathbf{z} =_d -2\sqrt{Q}\psi_0(Q)\boldsymbol{\Sigma}_0^{-1/2}\mathbf{u} \quad (45)$$

$$\begin{aligned} \tilde{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{z}) &= -\mathbf{D}_N^T (2^{-1}\text{vec}(\boldsymbol{\Sigma}_0^{-1}) + \\ &+ \psi_0(Q_0)\boldsymbol{\Sigma}_0^{-1/2} \otimes \boldsymbol{\Sigma}_0^{-1/2}\text{vec}(\mathbf{z}\mathbf{z}^T)) \\ &=_d -\mathbf{D}_N^T (2^{-1}\text{vec}(\boldsymbol{\Sigma}_0^{-1}) + \\ &+ Q\psi_0(Q)\boldsymbol{\Sigma}_0^{-1/2} \otimes \boldsymbol{\Sigma}_0^{-1/2}\text{vec}(\mathbf{u}\mathbf{u}^T)). \end{aligned} \quad (46)$$

2) *Derivation of  $\mathcal{T}_{S_0}$ :* The next step is the evaluation of  $\mathcal{T}_{S_0}$ , i.e. the tangent space of  $\mathcal{S}$  in (12), evaluated at the true density generator  $g_0$ . Using the procedure discussed in the Appendix A.3 of [5], it is possible to verify that  $\mathcal{T}_{S_0}$  is a  $q$ -replicating Hilber space  $\mathcal{T}_{S_0} = \mathcal{T} \times \dots \times \mathcal{T}$  (see also Appendix A) such that:

$$\mathcal{T}_{S_0} \triangleq \{l\mathbf{a} \mid \mathbf{a} \text{ is any vector in } \mathbb{R}^q, l \in \mathcal{T}\}, \quad (47)$$

where

$$\begin{aligned} \mathcal{T} &= \{l \in \mathcal{H}^1 \mid l \text{ is invariant under } \mathcal{O}\} \\ &= \{l \mid l(\mathbf{z}) \equiv l(\|\mathbf{z}\|), E_0\{l(\|\mathbf{z}\|) = 0\}\}, \end{aligned} \quad (48)$$

and  $\mathcal{O}$  is the group of orthogonal transformations defined in (8). Let us now recall that, according to Property P4 in Sec. III, the modular variate  $\mathcal{R}$  is a maximal invariant statistic for an SS distribution. Then, by using the procedure discussed in [5, Sec. 6.3, Example 1] and in accordance with the discussion provided in Appendix B, we have that the projection operator  $\Pi(\cdot|\mathcal{T}_{S_0})$  on the tangent space  $\mathcal{T}_{S_0}$  can be obtained as the expectation operator  $E_{0|\mathcal{R}}\{\cdot|\mathcal{R}\}$  with respect to the maximal invariant statistic  $\mathcal{R}$ , i.e.

$$\Pi(\cdot|\mathcal{T}_{S_0}) = E_{0|\mathcal{R}}\{\cdot|\mathcal{R}\}, \quad \forall l \in \mathcal{H}^q. \quad (49)$$

3) *Calculation of the projection  $\Pi(\mathbf{s}_{\phi_0}|\mathcal{T}_{g_0})$ :* In order to evaluate  $\Pi(\cdot|\mathcal{T}_{g_0})$ , we can use the property of the semiparametric group models given in (41). Let us start by deriving the expression for  $\Pi((\mathbf{s}_{\phi_0} \circ \alpha_{\phi_0})|\mathcal{T}_{S_0}) \equiv \Pi(\tilde{\mathbf{s}}_{\phi_0}|\mathcal{T}_{S_0})$ . By exploiting the results collected in the previous subsections,  $\Pi(\tilde{\mathbf{s}}_{\phi_0}|\mathcal{T}_{S_0})$  can be easily obtained by substituting (45) and (46) in (49) and then evaluating the expectation operator. Specifically:

$$\begin{aligned} \Pi(\tilde{\mathbf{s}}_{\boldsymbol{\mu}_0}|\mathcal{T}_{S_0}) &= E_{0|\mathcal{R}}\{\tilde{\mathbf{s}}_{\boldsymbol{\mu}_0}|\mathcal{R}\} \\ &=_d -2\sqrt{Q}\psi_0(Q)\boldsymbol{\Sigma}_0^{-1/2}E\{\mathbf{u}\} = \mathbf{0}, \end{aligned} \quad (50)$$

and

$$\begin{aligned} \Pi(\tilde{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)}|\mathcal{T}_{S_0}) &= E_{0|\mathcal{R}}\{\tilde{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{z})|\mathcal{R}\} \\ &=_d -\mathbf{D}_N^T (2^{-1}\text{vec}(\boldsymbol{\Sigma}_0^{-1}) + \\ &+ Q\psi_0(Q)\boldsymbol{\Sigma}_0^{-1/2} \otimes \boldsymbol{\Sigma}_0^{-1/2}E\{\text{vec}(\mathbf{u}\mathbf{u}^T)\}) \\ &= -\mathbf{D}_N^T \left( \frac{1}{2} + \frac{1}{N}Q\psi_0(Q) \right) \text{vec}(\boldsymbol{\Sigma}_0^{-1}), \end{aligned} \quad (51)$$

then, consequently:

$$\begin{aligned} \Pi(\tilde{\mathbf{s}}_{\phi_0}|\mathcal{T}_{S_0}) &=_d \begin{pmatrix} \mathbf{0} \\ -\mathbf{D}_N^T \left( \frac{1}{2} + \frac{1}{N}Q\psi_0(Q) \right) \text{vec}(\boldsymbol{\Sigma}_0^{-1}) \end{pmatrix} \\ &=_d \Pi(\mathbf{s}_{\phi_0}|\mathcal{T}_{g_0}), \end{aligned} \quad (52)$$

where the last equality follows from (41) and from the fact that  $\Pi(\tilde{\mathbf{s}}_{\phi_0}|\mathcal{T}_{S_0})$  does not depend on  $\mathbf{z}$ .

Now, a comment is in order. Equation (52) tells us that the score function of the mean value is orthogonal to the

nuisance tangent space  $\mathcal{T}_{g_0}$ . This means that not knowing the true density generator does not have any impact in the (asymptotic) estimation performance of the mean vector  $\boldsymbol{\mu}_0$  [17].

C. *Calculation of the semiparametric efficient score vector  $\bar{\mathbf{s}}_0(\mathbf{x})$  and of the SFIM  $\bar{\mathbf{I}}(\phi_0, g_0)$*

By collecting the previous results, the semiparametric efficient score vector in (30) can be evaluated as:

$$\bar{\mathbf{s}}_0(\mathbf{x}) = \mathbf{s}_{\phi_0}(\mathbf{x}) - \Pi(\mathbf{s}_{\phi_0}(\mathbf{x})|\mathcal{T}_{g_0})(\mathbf{x}) = \begin{pmatrix} \bar{\mathbf{s}}_{\boldsymbol{\mu}_0}(\mathbf{x}) \\ \bar{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)}(\mathbf{x}) \end{pmatrix} =_d \begin{pmatrix} -2\sqrt{Q}\psi_0(Q)\boldsymbol{\Sigma}_0^{-1/2}\mathbf{u} \\ -\mathbf{D}_N^T Q\psi_0(Q) \left( \boldsymbol{\Sigma}_0^{-1/2} \otimes \boldsymbol{\Sigma}_0^{-1/2} \text{vec}(\mathbf{u}\mathbf{u}^T) - \frac{\text{vec}(\boldsymbol{\Sigma}_0^{-1})}{N} \right) \end{pmatrix}. \quad (53)$$

Finally, through direct calculation, the SFIM  $\bar{\mathbf{I}}(\phi_0, g_0)$  can be obtained as:

$$\begin{aligned} \bar{\mathbf{I}}(\phi_0|g_0) &= E_0\{\bar{\mathbf{s}}_0(\mathbf{x})\bar{\mathbf{s}}_0(\mathbf{x})^T\} \\ &= \begin{pmatrix} \mathbf{C}_0(\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}) & \mathbf{0} \\ \mathbf{0}^T & \mathbf{C}_0(\bar{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)}) \end{pmatrix}, \end{aligned} \quad (54)$$

where  $\mathbf{C}_0(\mathbf{1}) \triangleq E_0\{\mathbf{1}\mathbf{1}^T\}$ ,  $\forall \mathbf{1} \in \mathcal{H}^q$ . Note that the off-diagonal block matrices in (54) are nil because all the third-order moments of  $\mathbf{u}$  vanish [14, Lemma 1]. From the previous results and using some algebra, we get:

$$\mathbf{C}_0(\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}) = \frac{4E\{Q\psi_0(Q)^2\}}{N}\boldsymbol{\Sigma}_0^{-1}, \quad (55)$$

and

$$\begin{aligned} \mathbf{C}_0(\bar{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)}) &= \frac{2E\{Q^2\psi_0(Q)^2\}}{N(N+2)} \times \\ &\times \mathbf{D}_N^T \left( \boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1} - \frac{1}{N} \text{vec}(\boldsymbol{\Sigma}_0^{-1})\text{vec}(\boldsymbol{\Sigma}_0^{-1})^T \right) \mathbf{D}_N. \end{aligned} \quad (56)$$

Note that we had not imposed the constraint on the scatter matrix  $\boldsymbol{\Sigma}_0$  as yet. In particular, in all the previous equations,  $\boldsymbol{\Sigma}_0$  can be considered as the *unconstrained* scatter matrix. The next subsection is then dedicated to the derivation of the SCRb for the constrained parameter vector.

D. *Evaluation of the CSCRb( $\phi_0|g_0$ )*

As showed in Theorem IV.1, as a prerequisite of the derivation of the CSCRb on the estimation of  $\phi_0 \in \bar{\Omega}$ , we have to calculate the matrix  $\mathbf{U}$  defined in (27). This can be done by using the same procedure discussed in [41]. Specifically, let us start by evaluating the gradient of the constraint in (23) as:

$$\mathbf{J}_c(\boldsymbol{\Sigma}_0) = \nabla_{\text{vecs}(\boldsymbol{\Sigma})}^T c(\boldsymbol{\Sigma}_0) = \mathbf{1}_I^T, \quad (57)$$

where  $\mathbf{1}_I$  is the  $N(N+1)/2$ -dimensional column vector defined as:

$$[\mathbf{1}_I]_i = \begin{cases} 1 & i \in I \\ 0 & \text{otherwise} \end{cases} \quad (58)$$

where  $I = \{i \mid i = 1 + N(j-1) - \frac{(j-1)(j-2)}{2}, j = 1, \dots, N\}$ .

Then,  $\mathbf{U}$  can be obtained by numerically evaluating, using singular value decomposition (SVD), the  $\bar{q} = q - 1$  orthonormal eigenvectors associated with the zero eigenvalue of  $\mathbf{1}_I$ .

Finally, the *constrained* SCRb (CSCRb) for the estimation of  $\phi_0 \in \bar{\Omega}$  in (24) can be expressed as:

$$\begin{aligned} \text{CSCRb}(\phi_0|g_0) &= \\ &= \begin{pmatrix} \frac{N}{4E\{Q\psi_0(Q)^2\}}\boldsymbol{\Sigma}_0 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{U}(\mathbf{U}^T \mathbf{C}_0(\bar{\mathbf{s}}_{\text{vecs}(\boldsymbol{\Sigma}_0)})\mathbf{U})^{-1} \mathbf{U}^T \end{pmatrix}. \end{aligned} \quad (59)$$

Note that the block-diagonal structure of  $\text{CSCRb}(\phi_0|g_0)$  implies that not knowing the mean vector  $\boldsymbol{\mu}_0$  does not have any impact on the asymptotic performance in the estimation of the scatter matrix  $\boldsymbol{\Sigma}_0$ . From a practical point of view, this also means that the unknown  $\boldsymbol{\mu}_0$  can be substituted with any consistent estimator without affecting the asymptotically optimal performance of the scatter matrix estimator.

## V. SIMULATION RESULTS FOR RES DISTRIBUTED DATA

In this section, we investigate the efficiency of three well-known scatter matrix estimators with respect to the CSCRb: the constrained Sample Covariance Matrix (CSCM) estimator, the constrained Tyler's (C-Tyler) estimator and the constrained Huber's (C-Hub) estimator. Note that none of these estimators relies on the a-priori knowledge of the true density generator  $g_0 \in \mathcal{G}$ .

Assume to have a set of  $M$ , RES-distributed, observation vectors  $\{\mathbf{x}_m\}_{m=1}^M$ . Let us define  $\{\bar{\mathbf{x}}_m\}_{m=1}^M$  as the set of  $M$  vectors such that:

$$\bar{\mathbf{x}}_m = \mathbf{x}_m - \hat{\boldsymbol{\mu}}, \quad m = 1, \dots, M \quad (60)$$

and  $\hat{\boldsymbol{\mu}}$  is the sample mean estimator, i.e.  $\hat{\boldsymbol{\mu}} \triangleq N^{-1} \sum_{m=1}^M \mathbf{x}_m$ .

The CSCM estimate can then be expressed as [42]:

$$\hat{\boldsymbol{\Sigma}}_{\text{CSCM}} \triangleq \frac{N}{\text{tr}(\boldsymbol{\Sigma}_{\text{SCM}})} \boldsymbol{\Sigma}_{\text{SCM}}, \quad \boldsymbol{\Sigma}_{\text{SCM}} \triangleq \frac{1}{M} \sum_{m=1}^M \bar{\mathbf{x}}_m \bar{\mathbf{x}}_m^T, \quad (61)$$

while C-Tyler and C-Hub estimates are the convergence points of the following iterative algorithm:

$$\begin{cases} \mathbf{S}_T^{(k+1)} = \frac{1}{M} \sum_{m=1}^M \varphi(t^{(k)}) \bar{\mathbf{x}}_m \bar{\mathbf{x}}_m^T, \\ \hat{\boldsymbol{\Sigma}}_T^{(k+1)} = N \mathbf{S}_T^{(k+1)} / \text{tr}(\mathbf{S}_T^{(k+1)}) \end{cases}, \quad (62)$$

where  $t^{(k)} = \bar{\mathbf{x}}_m^T (\hat{\boldsymbol{\Sigma}}_T^{(k)})^{-1} \bar{\mathbf{x}}_m$  and the starting point is  $\hat{\boldsymbol{\Sigma}}_T^{(0)} = \mathbf{I}$ . The weight function  $\varphi(t)$  for Tyler's estimator is defined as (see e.g. [14,43] and [15, Ch. 4] and references therein):

$$\varphi_{\text{Tyler}}(t) = N/t, \quad (63)$$

whereas the weight function for Huber's estimator is given by ([14], [44] and [15, Ch. 4]):

$$\varphi_{\text{Hub}}(t) = \begin{cases} 1/b & t \leq \delta^2 \\ \delta^2/(tb) & t > \delta^2 \end{cases}, \quad (64)$$

and  $\delta = F_{\chi_N^2}(u)$  where  $F_{\chi_N^2}(\cdot)$  indicates the distribution of a chi-squared random variable with  $N$  degrees of freedom and  $q \in (0, 1]$  is a tuning parameter. Moreover, the parameter  $b$  is

usually chosen as  $b = F_{\chi_{N+2}^2}(\delta^2) + \delta^2(1 - F_{\chi_N^2}(\delta^2))/N$  [14], [44].

To compare the three scatter matrix estimators with the CSCRb, we define the following performance index:

$$\varepsilon_\alpha \triangleq \|E\{(\text{vecs}(\hat{\Sigma}_\alpha) - \text{vecs}(\Sigma_0))(\text{vecs}(\hat{\Sigma}_\alpha) - \text{vecs}(\Sigma_0))^T\}\|_F, \quad (65)$$

where  $\alpha = \{CSCM, C - Tyler, C - Hub(u)\}$  indicates the particular estimator under test and  $\|\mathbf{A}\|_F^2 \triangleq \text{tr}(\mathbf{A}^T \mathbf{A})$  is the Frobenius norm of matrix  $\mathbf{A}$ . Similarly, for the sample mean, we have the following index:

$$\varepsilon_{\mu_0} \triangleq \|E\{(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T\}\|_F. \quad (66)$$

For the sake of comparison, we show in the following figures also the *constrained* CRB (CCRB) for the estimation of  $\phi_0 = (\boldsymbol{\mu}_0^T, \text{vecs}(\Sigma_0)^T)^T \in \bar{\Omega}$ . The classical FIM can be obtained from the score vectors previously derived in Subsection IV-A as a block matrix of the form [39]:

$$\mathbf{I}(\phi_0) = E_0\{\mathbf{s}_{\phi_0} \mathbf{s}_{\phi_0}^T\} = \begin{pmatrix} \mathbf{C}_0(\mathbf{s}_{\mu_0}) & \mathbf{0} \\ \mathbf{0}^T & \mathbf{C}_0(\mathbf{s}_{\text{vecs}(\Sigma_0)}) \end{pmatrix}. \quad (67)$$

Through direct calculation, it is easy to verify that:

$$\mathbf{C}_0(\mathbf{s}_{\mu_0}) = \mathbf{C}_0(\bar{\mathbf{s}}_{\mu_0}) = \frac{4E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\}}{N} \Sigma_0^{-1}, \quad (68)$$

and

$$\mathbf{C}_0(\mathbf{s}_{\text{vecs}(\Sigma_0)}) = \mathbf{D}_N^T (a_1 \text{vec}(\Sigma_0^{-1}) \text{vec}(\Sigma_0^{-1})^T + a_2 \Sigma_0^{-1} \otimes \Sigma_0^{-1}) \mathbf{D}_N \quad (69)$$

where:

$$a_1 \triangleq \frac{1}{4} + \frac{E\{\mathcal{Q}\psi_0(\mathcal{Q})\}}{N} + \frac{a_2}{2}, \quad (70)$$

$$a_2 \triangleq \frac{2E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\}}{N(N+2)}. \quad (71)$$

It is worth highlighting here that the difference between the classical FIM in (67) and the SFIM in (54) is due to the different expressions of the covariance matrix of the efficient score vector  $\bar{\mathbf{s}}_{\text{vecs}(\Sigma_0)}$  in (56) and of the one of the score vector  $\mathbf{s}_{\text{vecs}(\Sigma_0)}$  in (69).

Finally, the CCRB on the estimation of  $\phi_0 \in \bar{\Omega}$  can be obtained using exactly the same procedure discussed for the CSCRb in Subsection IV-D (see also [41]).

As performance bounds, the following indices are plotted:

$$\varepsilon_{CCRB, \Sigma_0} \triangleq \|[\text{CCRB}(\phi_0)]_{\Sigma_0}\|_F, \quad (72)$$

$$\varepsilon_{CSCRb, \mu_0} \triangleq \|[\text{CSCRb}(\phi_0, g_0)]_{\mu_0}\|_F, \quad (73)$$

$$\varepsilon_{CSCRb, \Sigma_0} \triangleq \|[\text{CSCRb}(\phi_0, g_0)]_{\Sigma_0}\|_F, \quad (74)$$

where  $[\cdot]_{\mu_0}$  and  $[\cdot]_{\Sigma_0}$  indicate the top-left and the bottom-right submatrices of the  $\text{CCRB}(\phi_0)$  and of the  $\text{CSCRb}(\phi_0, g_0)$ , respectively.

We analyze two different cases:

- 1) The true RES distribution is a  $t$ -distribution,
- 2) The true RES distribution is a Generalized Gaussian (GG) distribution.

The simulation parameters that are common to the two cases are:

- $[\Sigma_0]_{i,j} = \rho^{|i-j|}$ ,  $i, j = 1, \dots, N$ . Moreover,  $\rho = 0.8$  and  $N = 8$ .
- The data power is chosen to be  $\sigma_X^2 = E_{\mathcal{Q}}\{\mathcal{Q}\}/N = 4$ .
- The data mean value is chosen to be  $[\boldsymbol{\mu}_0]_i = 1$ ,  $i = 1, \dots, N$ .
- The number of the available independent and identically distributed (i.i.d.) data vectors is  $M = 3N = 24$ . Note that, since we assume to have  $M$  i.i.d. data vectors, the SFIM in (54) and the FIM in (67) have to be multiply by  $M$ .
- The tuning parameter  $u$  of Huber's estimator has been chosen as  $u = 0.9, 0.5, 0.1$ . Note that for  $u = 1$  Huber's estimator is equal to the SCM, while for  $u \rightarrow 0$  Huber's estimator tends to Tyler's estimator [14].
- The number of independent Monte Carlo runs is  $10^6$ .

#### A. Case 1: the $t$ -distribution

The density generator for the  $t$ -distribution is:

$$g_0(t) \triangleq \frac{2^{N/2} \Gamma(\frac{\lambda+N}{2})}{\pi^{N/2} \Gamma(\lambda/2)} \left(\frac{\lambda}{\eta}\right)^{\lambda/2} \left(\frac{\lambda}{\eta} + t\right)^{-\frac{\lambda+N}{2}} \quad (75)$$

and then  $\psi_0(t)$  in (36) is given by  $\psi_0(t) = -2^{-1}(\lambda + N)(\lambda/\eta + t)^{-1}$ . Consequently, from (13), we have that:

$$p_{\mathcal{Q}}(q) = \frac{\Gamma(\frac{\lambda+N}{2})}{\Gamma(\lambda/2)\Gamma(N/2)} \left(\frac{\lambda}{\eta}\right)^{\lambda/2} q^{N/2-1} \left(\frac{\lambda}{\eta} + q\right)^{-\frac{\lambda+N}{2}} \quad (76)$$

Using the integral in [45, pp. 315, n. 3.194 (3)], we have that:

$$E\{\mathcal{Q}\psi_0(\mathcal{Q})\} = -N/2 \quad (77)$$

$$E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\} = \frac{\eta N(\lambda + N)}{4(N + \lambda + 2)} \quad (78)$$

$$E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\} = \frac{N(N+2)(\lambda + N)}{4(N + \lambda + 2)} \quad (79)$$

Then, the coefficients (70) and (71) for the  $t$ -distribution are:

$$a_{1,t} \triangleq -\frac{1}{2(N + \lambda + 2)}, \quad a_{2,t} \triangleq \frac{\lambda + N}{2(N + \lambda + 2)}. \quad (80)$$

By substituting the previous results in (68), (69) and (56), we obtain the matrices  $\mathbf{C}_0(\mathbf{s}_{\mu_0})$ ,  $\mathbf{C}_0(\mathbf{s}_{\text{vecs}(\Sigma_0)})$  and  $\mathbf{C}_0(\bar{\mathbf{s}}_{\text{vecs}(\Sigma_0)})$  for the  $t$ -distribution.

Fig. 1 shows the MSE index of the sample mean compared with the CSCRb as function of the shape parameter  $\lambda$ . As we can see, when  $\lambda \rightarrow \infty$ , the sample mean tends to be an efficient estimator. This is an expected result, since, when the shape parameter  $\lambda$  goes to infinity, the  $t$ -distribution becomes a Gaussian distribution, then the sample mean is the ML estimator for  $\boldsymbol{\mu}_0$ . In Fig. 2 we can observe an interesting fact: the distance between the CSCRb and the CCRB increases as  $\lambda \rightarrow \infty$ . This means that the lack of knowledge of the particular density generator, i.e. the lack of knowledge of the particular RES distribution of the data, has an higher impact when the tails of the true distribution become lighter. This behavior has been already observed in [20]. Regarding the *constrained* scatter matrix estimators, the CSCM achieves the



CSCRB as  $\lambda \rightarrow \infty$ , i.e. as the data tends to be Gaussian distributed. Note that while it is well-known that the SCM is the ML estimator for the *unconstrained* scatter matrix, the CSCM is *not* the ML estimator for the constrained scatter matrix  $\Sigma_0$ . This is why, as  $\lambda \rightarrow \infty$ , the CSCM does not achieve the CCRB. Regarding C-Tyler's and C-Huber's estimators, from Fig. 2 we can see that C-Huber's estimator has better estimation performance for all the three values of  $u$  with respect to C-Tyler's estimator. In particular, the estimation performance of C-Huber's estimator improves as  $u$  tends to 1, i.e. when it tends to collapse to the CSCM. Both C-Tyler's and C-Huber's estimators are far from being efficient with respect to the CSCRB. However, it must be mentioned here that efficiency is not the only property that an estimator should have. Robustness is also important in the choice of an estimation algorithm. We will investigate the trade-off between efficiency and robustness in future works.

Another important question that may arise is related to the behaviour of the Maximum Likelihood estimator of the scatter matrix. To answer this question, firstly we have to note that the density generator of the  $t$ -distribution in (75) depends on two additional parameters: the shape  $\lambda$  and the scale  $\eta$ . If we assume to know perfectly *both* the functional form of  $g_0$  and the scale and shape parameters, then the ML estimator of the scatter matrix  $\Sigma$  will outperform all  $M$ -estimators and will achieve the classical CRB for  $\Sigma$ . This scenario is discussed in [39]. A more realistic situation is when only the functional form of  $g_0$  is assumed to be a priori known while the scale and shape parameters have to be jointly estimated with the scatter matrix  $\Sigma$ . A joint ML (JML) estimator for  $\Sigma$ ,  $\lambda$  and  $\eta$  generally does not exist, so we have to rely of sub-optimal strategies as the one discussed in [41]. Specifically, the recursive joint estimator of  $\Sigma$ ,  $\lambda$  and  $\eta$  proposed in [41] exploits the Method of Moments (MoM) to estimate the scale and shape parameters, while an estimation of the scatter matrix is obtained using an ML approach where the unknown parameters  $\lambda$  and  $\eta$  are replaced by their MoM estimates. Clearly, this JML algorithm will no longer achieve the classical CRB on the scatter matrix due to the lack of a priori knowledge about the scale and shape parameters. It would be interesting to investigate how the JML estimator behaves with respect to the CSCRB. As we can see from Fig. 1, the MSE index of the JML is larger than the CSCRB and this would suggest that not knowing the shape and scale parameters has the same impact of not knowing the whole functional form of the density generator. Of course, this aspect deserves further investigation and we leave it to future work.

### B. Case 2: the Generalized Gaussian distribution

The density generator relative to the Generalized Gaussian (GG) distribution is:

$$g_0(t) \triangleq \frac{2^{N/2} s \Gamma(N/2)}{\pi^{N/2} (2b)^{\frac{N}{2s}} \Gamma(N/2s)} \exp\left(-\frac{t^s}{2b}\right) \quad (81)$$

and then  $\psi_0(t)$  in (36) is given by  $\psi_0(t) = -s(2b)^{-1} t^{s-1}$ .

Consequently, from (13), we have that:

$$p_{\mathcal{Q}}(q) = \frac{s q^{N/2-1}}{(2b)^{\frac{N}{2s}} \Gamma(N/2s)} \exp\left(-\frac{q^s}{2b}\right). \quad (82)$$

Using the integral in [45, pp. 370, n. 3.478 (1)], we have that:

$$E\{\mathcal{Q}\psi_0(\mathcal{Q})\} = -N/2. \quad (83)$$

$$E\{\mathcal{Q}\psi_0(\mathcal{Q})^2\} = \frac{s^2 \Gamma(\frac{N+4s-2}{2s})}{(2b)^{1/s} \Gamma(N/2s)} \quad (84)$$

$$E\{\mathcal{Q}^2\psi_0(\mathcal{Q})^2\} = N(N+2s)/4. \quad (85)$$

Then, the coefficients (70) and (71) for the GG distribution are:

$$a_{1,GG} \triangleq \frac{s-1}{2(N+2)}, \quad a_{2,GG} \triangleq \frac{N+2s}{2(N+2)}. \quad (86)$$

As before, by substituting the previous results in (68), (69) and (56), we obtain the matrices  $\mathbf{C}_0(\mathbf{s}_{\mu_0})$ ,  $\mathbf{C}_0(\mathbf{s}_{\text{vecs}(\Sigma_0)})$  and  $\mathbf{C}_0(\bar{\mathbf{s}}_{\text{vecs}(\Sigma_0)})$  for the GG distribution.

The simulation results for the GG distributed data confirm all the previous discussions about  $t$ -distributed data:

- The sample mean estimator is an efficient estimator of  $\mu_0$  when the data is Gaussian distributed. In fact, as we can see from Fig. 3 that the MSE index  $\varepsilon_{\mu_0}$  equates the CSCRB for  $s = 1$ , i.e., when the GG distribution becomes the Gaussian one.
- The distance between the CSCRB and the CCRB increases as the tails of the true data distribution become lighter. This behavior can be observed in Fig. 4. It is worth recalling here that for  $0 < s < 1$  the GG distribution has heavier tails and for  $s > 1$  lighter tails compared to the Gaussian distribution that can be obtained for  $s = 1$ .
- The CSCM is an efficient estimator for  $\Sigma_0$  w.r.t. the CSCRB when the data is Gaussian distributed, i.e., when  $s = 1$  (see Fig. 4). However, it does not achieve the CCRB since, as discussed before, the CSCM is not the ML estimator for the constrained scatter matrix under Gaussian distributed data.

## VI. CONCLUSION

This paper is organized in two interrelated parts. The first part is devoted to place the class of RES distributions within the framework of semiparametric group models. This analysis allows to look at the well-known RES family from a different and enlightening standpoint. The main features of the semiparametric group models have been presented and discussed, paying particular attention to their implications on the family of RES distributions. In the second part of the paper, we showed how the direct application of these properties leads to derive a closed-form expression for the CSCRB for the joint estimation of the mean vector  $\mu_0$  and of the constrained scatter matrix  $\Sigma$  of a set of RES distributed random vectors.

Even if the semiparametric inference offers us a wide range of research opportunities, a huge amount of work still remains to be done. Our short-term research activity will be devoted to the extension of the CSCRB to the complex field, i.e. to the joint estimation of a *complex* mean vector and a *complex* scatter matrix in the family of Complex Elliptically Symmetric (CES) distributions. This is of great relevance in radar applications where both the data and the parameters to

be estimated are modeled as complex quantities. Regarding the long-term research activities, our efforts will be devoted to an in-depth study of the robustness property of an estimator in the framework of semiparametric models. Specifically, particular attention will be devoted to the analysis of the trade-off between robustness and semiparametric efficiency of an estimator.

## APPENDIX A THE HILBERT SPACE OF $q$ -DIMENSIONAL RANDOM FUNCTIONS

This Appendix provides some additional details on the Hilbert space  $\mathcal{H}^q$  of the zero-mean,  $q$ -dimensional functions and on the projection operator  $\Pi(\cdot|\mathcal{V})$ , where  $\mathcal{V} \subseteq \mathcal{H}^q$ . The following discussion does not claim to provide a complete mathematical characterization of these two elements, but could be useful as background material for the derivation of the CSCRb provided in this paper.

Let us introduce the underlying probability space  $(\mathcal{X}, \mathfrak{F}, P_X)$ , where  $\mathcal{X} \subseteq \mathbb{R}^N$  is the sample space,  $\mathfrak{F}$  is the Borel  $\sigma$ -algebra of events in  $\mathcal{X}$  and  $P_X$  is the probability measure. Let  $\mathbf{x} \in \mathcal{X}$  be a random vector, then  $P_X(\mathbf{a}) \triangleq P_X(x_1 \leq a_1, \dots, x_N \leq a_N)$  is its cumulative distribution function (cdf). We assume that the cdf  $P_X$  admits a relevant probability density function (pdf) (with respect to the standard Lebesgue measure), denoted as  $p_X$ , such that  $dP_X(\mathbf{a}) = p_X(\mathbf{a})d\mathbf{a}$ .

Consider now the vector space of the one-dimensional square-integrable function defined on  $(\mathcal{X}, \mathfrak{F}, P_X)$ :

$$L_2(P_X) = \left\{ h : \mathcal{X} \rightarrow \mathbb{R} \left| \int_{\mathcal{X}} |h(\mathbf{x})|^2 dP_X(\mathbf{x}) < \infty \right. \right\}, \quad (87)$$

with the inner product given by:

$$\langle h, g \rangle_X = E_X\{hg\} \triangleq \int_{\mathcal{X}} h(\mathbf{x})g(\mathbf{x})dP_X(\mathbf{x}), \quad \forall h, g \in L_2(P_X). \quad (88)$$

Let us define  $\mathcal{H}^1 \subseteq L_2(P_X)$  as the subspace of all the one-dimensional, zero-mean functions on  $(\mathcal{X}, \mathfrak{F}, P_X)$  such that:

$$\mathcal{H}^1 = \{h \in L_2(P_X) | E_X\{h\} = 0\}. \quad (89)$$

It is immediate to verify that  $\mathcal{H}^1$ , endowed with the inner product defined in (88), is an infinite-dimensional Hilbert space [27, Ch. 2].

The “ $q$ -replicating” Hilbert space  $\mathcal{H}^q$  of the zero-mean,  $q$ -dimensional functions on  $(\mathcal{X}, \mathfrak{F}, P_X)$  is defined as the Cartesian product of  $q$  copies of the Hilbert space  $\mathcal{H}^1$ , i.e.  $\mathcal{H}^q \triangleq \mathcal{H}^1 \times \mathcal{H}^1 \times \dots \times \mathcal{H}^1$  [27, Ch. 3, Def. 6], such that:

$$\mathcal{H}^q = \{\mathbf{h} = [h_1, \dots, h_q]^T | h_i \in \mathcal{H}^1, i = 1, \dots, q\}. \quad (90)$$

Due to the Cartesian product-based construction, the inner product of  $\mathcal{H}^q$  is naturally induced by the one of  $\mathcal{H}^1$ , i.e.:

$$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle_X \triangleq E_X\{\mathbf{h}_1^T \mathbf{h}_2\} = \sum_{i=1}^q E_X\{h_{1,i} h_{2,i}\}, \quad (91)$$

and consequently the norm is:

$$\|\mathbf{h}\|_X = \sqrt{\langle \mathbf{h}, \mathbf{h} \rangle_X} = \sum_{i=1}^q E_X\{h_i^2\}. \quad (92)$$

Let us now investigate the geometrical structure of  $\mathcal{H}^q$ , with a particular focus on the orthogonal projection of a generic element  $\mathbf{h} \in \mathcal{H}^q$  into a closed subspace  $\mathcal{V}$  of  $\mathcal{H}^q$ . The following theorem is a fundamental result in Hilbert spaces theory, and can be established in a very general setting (see e.g. [46, Theo. 3.9.3]). Here, we will adapt it to the particular Hilbert space  $\mathcal{H}^q$ .

**Theorem A.1** (The Projection Theorem). *Let  $\mathcal{V}$  be a closed subspace of the Hilbert space  $\mathcal{H}^q$  and let  $\mathbf{h}$  and  $\mathbf{g}$  be two  $q$ -dimensional, zero-mean functions on  $(\mathcal{X}, \mathfrak{F}, P_X)$ , such that  $\mathbf{h} \in \mathcal{H}^q$  and  $\mathbf{g} \in \mathcal{V}$ . Then, the following conditions are equivalent:*

- 1)  $\|\mathbf{h} - \mathbf{g}\|_X = \inf_{\mathbf{l} \in \mathcal{V}} \|\mathbf{h} - \mathbf{l}\|_X$ ,
- 2)  $\mathbf{h}$  can be uniquely written as

$$\mathbf{h} = \mathbf{g} + (\mathbf{h} - \mathbf{g}), \quad (93)$$

where  $\mathbf{g} \triangleq \Pi(\mathbf{h}|\mathcal{V}) \in \mathcal{V}$  and  $\mathbf{h} - \Pi(\mathbf{h}|\mathcal{V}) \in \mathcal{V}^\perp$ , where  $\mathcal{V}^\perp$  indicates the orthogonal complement of  $\mathcal{V}$  in  $\mathcal{H}^q$ ,

- 3) the element  $\mathbf{g} \triangleq \Pi(\mathbf{h}|\mathcal{V}) \in \mathcal{V}$  is then uniquely determined by the orthogonality constraint

$$\langle \mathbf{h} - \Pi(\mathbf{h}|\mathcal{V}), \mathbf{l} \rangle_X = 0, \quad \forall \mathbf{l} \in \mathcal{V}. \quad (94)$$

The operator  $\Pi(\cdot|\mathcal{V})$  defined in Theorem A.1 is called the *orthogonal projection operator* onto the closed subspace  $\mathcal{V}$ . The unique element  $\Pi(\mathbf{h}|\mathcal{V})$  is then called the *orthogonal projection* of  $\mathbf{h} \in \mathcal{H}^q$  onto  $\mathcal{V}$ . Furthermore, as a consequence of the Condition 2) in Theorem A.1, the Hilbert space  $\mathcal{H}^q$  can be written as the Cartesian product of the subspace  $\mathcal{V}$  and of its orthogonal complement  $\mathcal{V}^\perp$ , i.e.  $\mathcal{H}^q = \mathcal{V} \times \mathcal{V}^\perp$ .

## APPENDIX B PROJECTION OPERATOR AND CONDITIONAL EXPECTATION

In Appendix A we defined  $\mathcal{H}^q$  as the Hilbert space of the  $q$ -dimensional, zero-mean functions on the probability space  $(\mathcal{X}, \mathfrak{F}, P_X)$ . Let  $\mathfrak{G}(V) \subseteq \mathfrak{F}$  be the sub-sigma algebra generated by the random variable  $V$ . It can be shown (see e.g. [47, Ch. 23] and [5, Appendix 3]) that the set of all the  $q$ -dimensional, zero-mean functions on the probability space  $(\mathcal{X}, \mathfrak{G}(V), P_{X|V})$  is a closed linear subspace, say  $\mathcal{V}$ , of the Hilbert space  $\mathcal{H}^q$ .

This fact can be exploited to establish a link between the projection operator  $\Pi(\cdot|\mathcal{V})$  and the *conditional expectation*  $E_{X|V}\{\cdot|V\}$ . Firstly, let us define  $E_{X|V}\{\cdot|V\}$  as in [47, Ch. 23] and [5, Appendix 3].

**Definition B.1.** *Let  $\mathbf{h} \in \mathcal{H}^q$  and  $\mathbf{l} \in \mathcal{V} \subseteq \mathcal{H}^q$  be two  $q$ -dimensional, zero-mean functions on the probability spaces  $(\mathcal{X}, \mathfrak{F}, P_X)$  and  $(\mathcal{X}, \mathfrak{G}(V), P_{X|V})$  with  $\mathfrak{G}(V) \subseteq \mathfrak{F}$ , respectively. Then the conditional expectation  $E_{X|V}\{\mathbf{h}|V\}$  is the unique element in  $\mathcal{V}$ , such that:*

$$\langle \mathbf{h} - E_{X|V}\{\mathbf{h}|V\}, \mathbf{l} \rangle_X \triangleq E_X\{(\mathbf{h} - E_{X|V}\{\mathbf{h}|V\})^T \mathbf{l}\} = 0, \quad (95)$$

for every  $\mathbf{l} \in \mathcal{V}$ .

For a more general and formal definition we refer the reader to [47, Ch. 23] and [5, Appendix 3]. The condition (95) is

equivalent to (94) in Theorem A.1 that defines the projection operator, and consequently, we have that:

$$\Pi(\cdot|\mathcal{V}) = E_{\mathcal{X}|V}\{\cdot|V\}. \quad (96)$$

The usefulness of this relation is in the fact that, for some semiparametric models, the tangent space presents an invariance structure with respect to a group of transformations and it admits a characterization through a certain sub-sigma algebra generated by the relevant maximal invariant statistic [30]. An example of a semiparametric model that owns this property is the semiparametric group model of RES distributions discussed in Subsection III-B.

#### ACKNOWLEDGMENT

The work of Stefano Fortunati has been partially supported by the Air Force Office of Scientific Research under award number FA9550-17-1-0065.

#### REFERENCES

- [1] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 142–157, Nov 2017.
- [2] S. Fortunati, M. S. Greco, and F. Gini, "Asymptotic robustness of Kelly's GLRT and Adaptive Matched Filter detector under model misspecification," in *ISI World Statistics Congress 2017 (ISI2017)*, 2017. [Online]. Available: <https://arxiv.org/abs/1709.08667>
- [3] P. J. Bickel and J. Kwon, "Inference for semiparametric models: some questions and an answer," *Statistica Sinica*, vol. 11, pp. 863–960, 2001.
- [4] J. A. Wellner, "Semiparametric models: Progress and problems," in *Bulletin of the International Statistical Institute*, ser. 4, ISI, Ed., vol. 51, 1985.
- [5] P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, 1993.
- [6] S.-I. Amari and J. F. Cardoso, "Blind source separation-semiparametric statistical approach," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2692–2700, Nov 1997.
- [7] U. Hammes, E. Wolsztynski, and A. M. Zoubir, "Transformation-based robust semiparametric estimation," *IEEE Signal Processing Letters*, vol. 15, pp. 845–848, 2008.
- [8] 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.
- [9] C. D. Richmond, "Adaptive array signal processing and performance analysis in non-Gaussian environments," Ph.D. dissertation, Massachusetts Institute of Technology, 1996. [Online]. Available: <https://dspace.mit.edu/handle/1721.1/11005>
- [10] M. A. Chmielewski, "Elliptically symmetric distributions: A review and bibliography," *International Statistical Review / Revue Internationale de Statistique*, vol. 49, no. 1, pp. 67–74, 1981.
- [11] K. Yao, "A representation theorem and its applications to spherically-invariant random processes," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 600–608, September 1973.
- [12] S. Zozor and C. Vignat, "Some results on the denoising problem in the elliptically distributed context," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 134–150, Jan 2010.
- [13] S. Kotz and S. Nadarajah, "Some extremal type elliptical distributions," *Statistics & Probability Letters*, vol. 54, no. 2, pp. 171 – 182, 2001.
- [14] 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.
- [15] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.
- [16] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy, "A fresh look at the semiparametric Cramér-Rao bound," in *26th European Signal Processing Conference (EUSIPCO)*, 2018. [Online]. Available: <http://arxiv.org/abs/1803.00267>
- [17] P. J. Bickel, "On adaptive estimation," *The Annals of Statistics*, vol. 10, no. 3, pp. 647–671, 1982.
- [18] 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.
- [19] 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.
- [20] M. Hallin and D. Paindaveine, "Parametric and semiparametric inference for shape: the role of the scale functional," *Statistics & Decisions*, vol. 24, no. 3, pp. 327–350, 2009.
- [21] D. Paindaveine, "A canonical definition of shape," *Statistics & Probability Letters*, vol. 78, no. 14, pp. 2240 – 2247, 2008.
- [22] L. LeCam and G. L. Yang, *Asymptotics in Statistics: Some Basic Concepts (second edition)*. Springer series in statistics, 2000.
- [23] M. Hallin and B. J. M. Werker, "Semi-parametric efficiency, distribution-freeness and invariance," *Bernoulli*, vol. 9, no. 1, pp. 137–165, 2003.
- [24] W. K. Newey, "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, vol. 5, no. 2, pp. 99–135, 1990.
- [25] —, "The asymptotic variance of semiparametric estimators," *Econometrica*, vol. 62, no. 6, pp. 1349–1382, 1994.
- [26] C. A. J. Klaassen, "Consistent estimation of the influence function of locally asymptotically linear estimators," *Ann. Statist.*, vol. 15, no. 4, pp. 1548–1562, 12 1987.
- [27] A. Tsiatis, *Semiparametric Theory and Missing Data*. Springer series in statistics, 2006.
- [28] H. Rieder, *Robust Asymptotic Statistics*. Springer series in statistics, 1994.
- [29] K.-T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*. Monographs on Statistics and Applied Probability, Springer US, 1990.
- [30] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer Texts in Statistics, 2004.
- [31] J. R. Magnus and H. Neudecker, "The commutation matrix: Some properties and applications," *The Annals of Statistics*, vol. 7, no. 2, pp. 381–394, 03 1979.
- [32] —, "The elimination matrix: Some lemmas and applications," *SIAM Journal on Algebraic Discrete Methods*, vol. 1, no. 4, pp. 422–449, 1980.
- [33] J. M. Begun, W. J. Hall, W.-M. Huang, and J. A. Wellner, "Information and asymptotic efficiency in parametric-nonparametric models," *The Annals of Statistics*, vol. 11, no. 2, pp. 432–452, 1983.
- [34] P. Stoica and B. C. Ng, "On the Cramér-Rao Bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, July 1998.
- [35] T. J. Moore, R. J. Kozick, and B. M. Sadler, "The constrained Cramér-Rao bound from the perspective of fitting a model," *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 564–567, Aug 2007.
- [36] S. Fortunati, F. Gini, and M. S. Greco, "The Constrained Misspecified Cramér - Rao Bound," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 718–721, May 2016.
- [37] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed., 1999.
- [38] M. Greco, S. Fortunati, and F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions," *Signal Processing*, vol. 104, pp. 381–386, 2014.
- [39] M. Greco and F. Gini, "Cramér-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6401–6409, 2013.
- [40] S. Fortunati, F. Gini, and M. Greco, "The Misspecified Cramér-Rao bound and its application to scatter matrix estimation in complex elliptically symmetric distributions," *IEEE Transactions on Signal Processing*, vol. 64, no. 9, pp. 2387 – 2399, 2016.
- [41] S. Fortunati, F. Gini, and M. S. Greco, "Matched, mismatched, and robust scatter matrix estimation and hypothesis testing in complex t-distributed data," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 123, 2016. [Online]. Available: <https://doi.org/10.1186/s13634-016-0417-0>
- [42] F. Pascal, P. Forster, J. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2206–2217, June 2008.
- [43] Y. Chitour and F. Pascal, "Exact maximum likelihood estimates for sirv covariance matrix: Existence and algorithm analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4563–4573, Oct 2008.
- [44] E. Ollila, I. Solovychik, D. E. Tyler, and A. Wiesel, "Simultaneous penalized M-estimation of covariance matrices using geodesically convex optimization," *Submitted to Journal of Multivariate Analysis*. [Online]. Available: <https://arxiv.org/abs/1608.08126>

- [45] I. S. Gradshteyn and M. Ryzhik, *Tables of Integrals, Series, and Products (7th edition)*. Academic Press, Orlando, Florida, 2007.
- [46] L. Debnath and P. Mikusiński, *Introduction to Hilbert Spaces with Applications (Second Edition)*. Academic Press, 1999.
- [47] J. Jacod and P. Protter, *Probability Essentials*. Springer series in statistics, 2004.

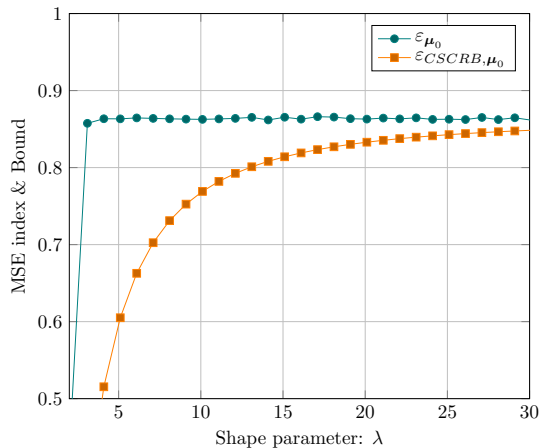


Fig. 1: MSE index for  $\hat{\mu}$  and the related CSCRb as functions of the shape parameter  $\lambda$  for  $t$ -distributed data ( $M = 3N$ ).

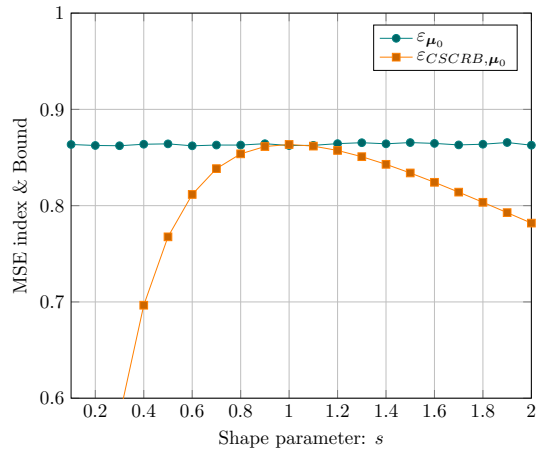


Fig. 3: MSE index for  $\hat{\mu}$  and the related CSCRb as functions of the shape parameter  $s$  for GG distributed data ( $M = 3N$ ).

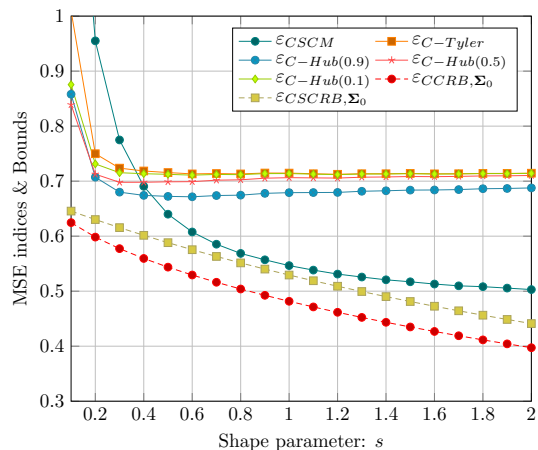


Fig. 4: MSE indices for the constrained scatter matrix estimators and the related CCRb and CSCRb as functions of the shape parameter  $s$  for GG distributed data ( $M = 3N$ ).

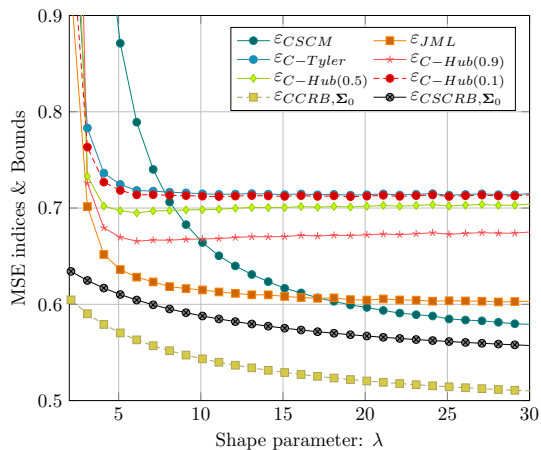


Fig. 2: MSE indices for the constrained scatter matrix estimators and the related CCRb and CSCRb as functions of the shape parameter  $\lambda$  for  $t$ -distributed data ( $M = 3N$ ).