Matched and Mismatched Estimation of Kronecker Product of Linearly Structured Scatter Matrices under Elliptical Distributions

Bruno Mériaux, Chengfang Ren, *Member, IEEE*, Arnaud Breloy, *Member, IEEE*,
Mohammed Nabil El Korso, *Member, IEEE* and Philippe Forster, *Member, IEEE*

Abstract—The estimation of covariance matrices is a core problem in many modern adaptive signal processing applications. For matrix- and array-valued data, e.g., MIMO communication, EEG/MEG (time versus channel), the covariance matrix of vectorized data may belong to the non-convex set of Kronecker product structure. In addition, the Kronecker factors can also exhibit an additional linear structure. Taking this prior knowledge into account during the estimation process drastically reduces the amount of unknown parameters, and then improves the estimation accuracy. On the other hand, the broad class of complex elliptically symmetric distributions, as well as the related complex angular elliptical distribution, are particularly suited to model heavy-tailed multivariate data. In this context, we first establish novel robust estimators of scatter and shape matrices (both related to a covariance matrix), having a Kronecker product structure with linearly structured Kronecker factors. Then, we conduct a theoretical analysis of their asymptotic performance (i.e., consistency, asymptotic distribution and efficiency), in matched and mismatched scenarios, i.e., when misspecifications between the *true* and *assumed* models occur. Finally, numerical results illustrate the theoretical analysis and assess the usefulness of the proposed estimators.

Index Terms—Kronecker product, scatter/shape matrix, structured and robust estimation, M-estimators, complex elliptically symmetric distributions, misspecified models.

I. INTRODUCTION

COVARIANCE matrices estimation plays a central role in many adaptive statistical signal processing methods [1]. Besides their Hermitian positive definite property, the latter usually possess a refined structure depending on the considered application. For instance, the Kronecker Product (KP) structure can arise for separable statistical models [2], or in diverse scenarios such as MIMO communication [3] or analysis of MEG/EEG data [4], [5]. Furthermore, the Kronecker factors may also have their own structure, e.g., a Toeplitz structure in MIMO communications with uniform linear array at the receiver or transmitter side [6]. Taking this prior knowledge into account in the estimation scheme, i.e., reducing the degree of freedom in the estimation problem, leads to a better accuracy and thus better performance of the processing chain. The problem of covariance estimation

B. Mériaux and C. Ren are with SONDRA, CentraleSupélec, France. M.N. El Korso and A. Breloy are with LEME, Paris-Nanterre University, France. P. Forster is with SATIE, Paris-Nanterre University, France.

with KP structure and potential linearly constrained factors has been widely explored in the Gaussian framework [6]–[10], notably by using the Extended Invariance Principle (EXIP) [11]. Furthermore, other approaches, dealing with high dimensional data, have been proposed in [12], [13].

However, assuming a Gaussian model may lead to a performance degradation in practical applications, since the latter is not suited for heavy-tailed distributions, nor in the presence of outliers. This is why, the class of circular Complex Elliptically Symmetric distributions (CES) has attracted research interest [14]–[17]. Indeed, CES distributions encompass the Gaussian model, but above all, a large number of non-Gaussian distributions such as Generalized Gaussian, compound Gaussian, t-distribution, K-distribution, etc [18]. In the zero-mean case, this broad class of distributions is related to the Complex Angular Elliptical (CAE) distribution [18], which is obtained by normalizing any centered CES random vectors. In the scope of robust estimation, the study of model misspecifications within this family and their impact on the performance has also attracted attention [19]–[22]. Such misspecifications are often unavoidable in practice, e.g., due to imperfect knowledge of the *true* data model.

In the context of robust and structured covariance estimation, several methods have been recently proposed [8], [22]–[26]. In [22], efficient robust estimators of convexly structured scatter matrices of CES distributions are studied, with an analysis of possible misspecifications on the statistical data model. The counterpart for the shape matrix of CAE distribution is addressed in [23]. However, these works are inadequate for the considered case since the KP does not constitute a convex structure, even for linearly constrained Kronecker factors. In [8], the Maximum Likelihood (ML) estimator of KP structured matrices for CAE distributed samples is studied and can be characterized thanks to the geodesic convexity of the likelihood function [27]. Similarly [26] proposes a ML-type KP structured scatter matrix estimator suited for compound-Gaussian distribution. Nevertheless, this result cannot be transposed to linearly structured Kronecker factors, which do not preserve the convex geodesic property. In [24] and [25], estima-

tors have been proposed as minimizers of Tyler's cost function, i.e., the likelihood related to CAE distributed samples [28], for various structures and computed with Majorization-Minimization algorithms. For a KP of linearly constrained factors, several approaches of [24] can be stacked, yielding an imbrication of convex problems, which can become computationally costly. This algorithmic approach yields good results in practice but cannot be characterized by statistical properties.

The contributions of this paper are the following:

- we propose and study the statistical properties of new robust estimators of the scatter/shape matrix with a KP structure and potential linearly constrained Kronecker factors. The proposed method consists in two main steps, namely an unstructured M-estimation of the matrix of interest, followed by an appropriate weighted least-squares problem. The last stage has a closed-form solution. In addition, recursion of the last step, which consists in iteratively refining the metric, is empirically exhibiting better performance at low sample support, by achieving faster the Cramér-Rao Bound.

- we conduct a theoretical study of the asymptotic performance of the proposed estimators in several cases. Under CES distributed samples, we consider possible misspecifications on the density generator and we show that the estimate is consistent, in the sense of the misspecified framework [29], and is asymptotically Gaussian distributed. The latter can also be asymptotically efficient, depending on the chosen first step. Under CAE distributed observations (i.e., assuming normalized samples drawn from an arbitrary zero-mean CES distribution), we show that the proposed procedure yields a consistent, asymptotically efficient and Gaussian distributed estimate.

- we exhibit simplified expressions of the Cramér-Rao Bounds, related to the different considered models, throughout the asymptotic analysis.

This paper is organized as follows. In section II, a brief review on CES distributions and its related Cramér-Rao bounds is presented. Section III focuses on the proposed estimator. The theoretical performance analysis, specifically weak consistency and asymptotic Gaussianity, is treated in Section IV. The statistical efficiency of the proposed estimate is addressed in Section V. Section VI tackles the particular case of robust estimation of the shape matrix. Some simulations results in Section VII illustrate the previous theoretical analysis before a conclusion in Section VIII.

In the following, convergence in distribution and in probability are, respectively, denoted by $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{\mathcal{P}}$. The notation $\overset{d}{=}$ indicates "has the same distribution as". For a matrix $\mathbf{X}$, the operator $\mathbf{X}^\dagger$ denotes the Moore-Penrose inverse. $|\mathbf{X}|$ and $\mathrm{Tr}\,(\mathbf{X})$ denote the determinant and the trace of $\mathbf{X}$. $\mathbf{X}^T$ (respectively $\mathbf{X}^H$ and $\mathbf{X}^*$) stands for the transpose (respectively conjugate transpose and

conjugate) matrix. The vec-operator $\mathrm{vec}(\mathbf{X})$ stacks all columns of $\mathbf{X}$ into a vector and $\mathbf{X} \succeq \mathbf{Y}$ implies that $\mathbf{X} - \mathbf{Y}$ is positive semi-definite. The operator $\otimes$ refers to Kronecker matrix product. The matrix $\mathbf{K}_{k,\ell}$ denotes the commutation matrix, satisfying $\mathbf{K}_{k,\ell}\mathrm{vec}\,(\mathbf{X}) = \mathrm{vec}\left(\mathbf{X}^T\right)$ for all $\mathbf{X} \in \mathbb{C}^{k\times\ell}$ [30]. $\mathbf{I}_m$ is the identity matrix of size m. The notation $\mathcal{GCN}\,(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ refers to the zero mean noncircular complex Gaussian distribution, where $\boldsymbol{\Sigma}$ (respectively $\boldsymbol{\Omega}$) denotes the covariance matrix (respectively the pseudo-covariance matrix) [31], [32]. The operator $\mathbb{E}_p\,[\cdot]$ refers to the expectation operator w.r.t. the probability density function (p.d.f.) p, which will be omitted if there is no ambiguity. Finally, the set of non-negative real numbers is denoted by $\mathbb{R}^+$.

II. Background and Problem Setup

In this section, we recall key elements for the class of CES distributions, as well as some lower bounds on the error covariance matrix of estimators, which will be useful for the following analysis. The reader is referred to [18] for a complete overview on CES distributions.

A. Complex Elliptical Distributions

A random vector (r.v.), $\mathbf{y} \in \mathbb{C}^m$ is said to be CES distributed if and only if it admits the following stochastic representation [18]:

$$\mathbf{y} \overset{d}{=} \mathbf{m} + \sqrt{Q}\mathbf{Su}, \tag{1}$$

where $\mathbf{m} \in \mathbb{C}^m$ denotes the location parameter and the matrix $\mathbf{S} \in \mathbb{C}^{m\times k}$ has $\mathrm{rank}(\mathbf{S}) = k$. The 2nd-order modular variate Q is a non-negative real-valued random variable statistically independent of the complex r.v. $\mathbf{u}$. The latter r.v. is uniformly distributed on the unit k-sphere $\mathbb{C}S^k \triangleq \left\{\mathbf{z} \in \mathbb{C}^k \mid \|\mathbf{z}\| = 1\right\}$ with $k \leq m$. Moreover, under existence condition, the mean of $\mathbf{y}$ is equal to $\mathbf{m}$ and the covariance matrix of $\mathbf{y}$ is proportional to the scatter matrix, $\mathbf{R} = \mathbf{SS}^H$, i.e., $\mathbb{E}\left[(\mathbf{y} - \mathbf{m})(\mathbf{y} - \mathbf{m})^H\right] = \frac{1}{k}\mathbb{E}\,[Q]\,\mathbf{R}$. In the following, we assume that $\mathbf{m}$ is known and equal to zero, without loss of generality and that r.v. $\mathbf{y}$ is absolutely continuous, so $k = m$ and $\mathrm{rank}(\mathbf{R}) = m$. In this case, the p.d.f. of such a vector exists and is expressed by [18]:

$$p_{\mathbf{Y}}(\mathbf{y}; \mathbf{R}, g) = C_{m,g}|\mathbf{R}|^{-1}g\left(\mathbf{y}^H\mathbf{R}^{-1}\mathbf{y}\right), \tag{2}$$

in which the function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is called the density generator satisfying $\delta_{m,g} \triangleq \int\limits_0^{+\infty} t^{m-1}g(t)\mathrm{d}t < \infty$ and $C_{m,g} = \dfrac{\Gamma(m)}{\pi^m}\delta_{m,g}^{-1}$ is the normalizing constant. In this case and thereafter, we shorten such r.v. by the notation $\mathbf{y} \sim \mathcal{CES}_m\,(\mathbf{0}, \mathbf{R}, g)$. As noted above, various non-Gaussian distributions (e.g., Generalized Gaussian, t-distribution, K-distribution, etc.) fall into this large class. Expressions of density generator function can be found in [22, Table 1] for commonly used CES distributions.

B. Cramér-Rao Bounds

In the context of deterministic parameter estimation, the derivation of lower bounds on the error covariance matrix is widespread in order to characterize the ultimate performance that an estimator can achieve. Among them, the so-called Cramér-Rao Bound (CRB) and its variants are commonly used. The latter is related to the Fisher Information Matrix (FIM), of which a closed-form expression is given in [33] for CES distributions. Notably for a centered circular complex CES distribution with scatter matrix $\mathbf{R}$ parameterized by $\boldsymbol{\mu} \in \mathbb{R}^n$, i.e., $\mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}(\boldsymbol{\mu}), g\right)$, the FIM expression is given by:

$$\mathbf{F}\left(\boldsymbol{\mu}\right) = \left(\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^T}\right)^H \mathbf{Y} \frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^T} \qquad (3)$$

where $\mathbf{Y} = \sigma_{1,\text{ML}}^{-1} \mathbf{W}_{\mathbf{R}}^{-1} + (\sigma_{1,\text{ML}}^{-1} - 1)\text{vec}\left(\mathbf{R}^{-1}\right)\text{vec}\left(\mathbf{R}^{-1}\right)^H$ in which $\mathbf{W}_{\mathbf{R}} = \mathbf{R}^T \otimes \mathbf{R}$, $\sigma_{1,\text{ML}} = \dfrac{m(m+1)}{\mathbb{E}\left[\psi_{\text{ML}}^2\left(|\mathbf{t}|^2\right)\right]} \in \left]0, \dfrac{m+1}{m}\right[$, with $\mathbf{t} \sim \mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{I}, g\right)$ and $\psi_{\text{ML}}(s) = -\dfrac{sg'(s)}{g(s)}$. The matrix $\dfrac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^T}$ refers to the Jacobian matrix of $\mathbf{r}(\boldsymbol{\mu}) = \text{vec}\left(\mathbf{R}(\boldsymbol{\mu})\right)$. For the particular case of the Gaussian distribution, we retrieve the Slepian-Bang's formula.

For unconstrained and identifiable models, the FIM is not singular. Thus, the Mean Square Error (MSE) of any unbiased estimator $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$ can be lower bounded by the inverse of the FIM, which is then referred to as the CRB. Specifically, we have

$$\mathbb{E}\left[\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)^T\right] \succeq \mathbf{F}^{-1}\left(\boldsymbol{\mu}\right) \triangleq \mathbf{CRB}_{\boldsymbol{\mu}}. \qquad (4)$$

In case of equality in (4) in the (in)finite-sample regime, the estimator $\hat{\boldsymbol{\mu}}$ is then considered to be (asymptotically) efficient.

In some cases, the estimator $\hat{\boldsymbol{\mu}}$ must satisfy a set of $k < n$ continuously differentiable non redundant constraints [34], [35],

$$\mathbf{h}\left(\hat{\boldsymbol{\mu}}\right) = \mathbf{0} \qquad (5)$$

which must be consistent, i.e., the set $\left\{\boldsymbol{\mu} \,|\, \mathbf{h}\left(\boldsymbol{\mu}\right) = \mathbf{0}\right\}$ is nonempty [35]. Then, we introduce a matrix $\mathbf{U} \in \mathbb{R}^{n \times (n-k)}$, whose columns form an orthonormal basis for the nullspace of the Jacobian matrix of $\mathbf{h}\left(\boldsymbol{\mu}\right)$, i.e.,

$$\frac{\partial \mathbf{h}\left(\boldsymbol{\mu}\right)}{\partial \boldsymbol{\mu}^T} \mathbf{U} = \mathbf{0} \quad \text{and} \quad \mathbf{U}^T \mathbf{U} = \mathbf{I}_{n-k}.$$

In this context, a Constrained version of the CRB (CCRB) has been introduced in [35] for any unbiased estimate of $\boldsymbol{\mu}$ satisfying (5), given by

$$\mathbb{E}\left[\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right)^T\right] \succeq \mathbf{U}\left(\mathbf{U}^T \mathbf{F}\left(\boldsymbol{\mu}\right)\mathbf{U}\right)^{-1}\mathbf{U}^T \triangleq \mathbf{CCRB}_{\boldsymbol{\mu}} \qquad (6)$$

In most estimation problems, the perfect knowledge of the data model generally turns out to be a key assumption. However, model misspecifications are often unavoidable in practice, due to incomplete knowledge of the *true* data model for instance [19]. The p.d.f. related to the *true* (respectively *assumed*) model is then denoted by $p(\cdot)$ (respectively $f_{\boldsymbol{\mu}}(\cdot)$). An extension of the classical CRB has been derived in [20], [29], [36], leading to the Misspecified Cramér-Rao Bound (MCRB). Before deriving the MCRB, the so-called *pseudo-true* parameter vector, $\boldsymbol{\mu}_0$, is classically introduced for an asymptotic analysis [19], [29], [36]. The latter is defined as a minimizer of the Kullback-Leibler divergence (KLD) between the *true* and the *assumed* models, specifically:

$$\boldsymbol{\mu}_0 = \arg\min_{\boldsymbol{\mu}} \mathcal{D}\left(p\|f_{\boldsymbol{\mu}}\right) = \arg\max_{\boldsymbol{\mu}} \mathbb{E}_p\left[\log f_{\boldsymbol{\mu}}\left(\mathbf{z}; \boldsymbol{\mu}\right)\right], \quad (7)$$

where $\mathcal{D}\left(p\|f_{\boldsymbol{\mu}}\right) \triangleq \mathbb{E}_p\left[\log \dfrac{p\left(\mathbf{z}; \boldsymbol{\mu}_e\right)}{f_{\boldsymbol{\mu}}\left(\mathbf{z}; \boldsymbol{\mu}\right)}\right]$. In the following, we always assume the existence and the uniqueness of this *pseudo-true* parameter vector (the reader is referred to [36] for necessary and sufficient conditions). From [29], the related MCRB and the error covariance matrix of an unconstrained and unbiased estimate $\hat{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}_0$ are connected through

$$\mathbb{E}\left[\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\right)\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\right)^T\right] \succeq \boldsymbol{\Gamma}_{\boldsymbol{\mu}_0}^{-1} \boldsymbol{\Upsilon}_{\boldsymbol{\mu}_0} \boldsymbol{\Gamma}_{\boldsymbol{\mu}_0}^{-1} \triangleq \mathbf{MCRB}_{\boldsymbol{\mu}_0} \qquad (8)$$

where $\boldsymbol{\Gamma}_{\boldsymbol{\mu}} = \mathbb{E}_p\left[\dfrac{\partial^2 \log f_{\boldsymbol{\mu}}}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T}\right]$ and $\boldsymbol{\Upsilon}_{\boldsymbol{\mu}} = \mathbb{E}_p\left[\dfrac{\partial \log f_{\boldsymbol{\mu}}}{\partial \boldsymbol{\mu}} \dfrac{\partial \log f_{\boldsymbol{\mu}}}{\partial \boldsymbol{\mu}^T}\right]$. For misspecifications on the density generator function of CES distributions, $\mathbf{MCRB}_{\boldsymbol{\mu}_0}$ of (8) has been explicitly derived in [22], [29]. Likewise in case of equality in (8) in the (in)finite-sample regime, the estimator $\hat{\boldsymbol{\mu}}$ is then called to be (asymptotically) m-efficient.

Finally, in a similar manner as the CCRB, a Constrained version of the MCRB (CMCRB) has been introduced in [21] for any unbiased estimate of $\boldsymbol{\mu}_0$ satisfying (5), expressed by

$$\mathbb{E}\left[\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\right)\left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\right)^T\right] \succeq \mathbf{U}\left(\mathbf{U}^T \boldsymbol{\Gamma}_{\boldsymbol{\mu}_0} \mathbf{U}\right)^{-1} \mathbf{U}^T \boldsymbol{\Upsilon}_{\boldsymbol{\mu}_0} \mathbf{U}$$
$$\times \left(\mathbf{U}^T \boldsymbol{\Gamma}_{\boldsymbol{\mu}_0} \mathbf{U}\right)^{-1} \mathbf{U}^T \triangleq \mathbf{CMCRB}_{\boldsymbol{\mu}_0} \qquad (9)$$

C. Problem Setup

Let us consider K i.i.d. zero mean CES distributed observations, denoted by $\mathbf{y}_k \sim \mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}_e, g\right)$, $k = 1, \ldots, K$. The *true* p.d.f. is denoted by $p_{\mathbf{Y}}\left(\mathbf{y}_k; \mathbf{R}_e\right)$. Note that the function $g(\cdot)$ is not assumed to be known and is not leveraged in the following algorithm. In other words, we only assume that data are drawn from an unspecified centered CES distribution $\mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}, g_{\text{mod}}\right)$, where the *assumed* density generator $g_{\text{mod}}(t)$ may be different from $g(t)$ for all $t \in \mathbb{R}^+$. The related p.d.f. is then denoted by $f_{\mathbf{Y}}\left(\mathbf{y}_k; \mathbf{R}\right)$.

The scatter matrix $\mathbf{R}$ is assumed to be structured as a KP, i.e.,

$$\mathbf{R}_e = \mathbf{A}_e \otimes \mathbf{B}_e \qquad (10)$$

where $\mathbf{A}_e \in \mathbb{C}^{n \times n}$ and $\mathbf{B}_e \in \mathbb{C}^{p \times p}$, also called Kronecker factors, are positive definite Hermitian matrices. We assume that the latter are linearly parameterized by $\boldsymbol{\theta}_A \in \mathbb{R}^{n_A}$ and $\boldsymbol{\theta}_B \in \mathbb{R}^{n_B}$ respectively, such that

$$\text{vec}\,(\mathbf{A}) \triangleq \mathbf{a} = \mathbf{P}_A \boldsymbol{\theta}_A \quad \text{and} \quad \text{vec}\,(\mathbf{B}) \triangleq \mathbf{b} = \mathbf{P}_B \boldsymbol{\theta}_B \tag{11}$$

with $n_A \leq n^2$, $n_B \leq p^2$ and where $\mathbf{P}_A$ and $\mathbf{P}_B$ are known full column rank matrices. This parameterization allows a potential refined structure on the Kronecker factors, such as persymmetric, Toeplitz, banded, etc. We denote the concatenated parameter vector by $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_A^T, \boldsymbol{\theta}_B^T \right]^T$, with exact value $\boldsymbol{\theta}_e$ and $\mathbf{R}_e = \mathbf{R}\,(\boldsymbol{\theta}_e) \triangleq \mathbf{A}\,(\boldsymbol{\theta}_{A_e}) \otimes \mathbf{B}\,(\boldsymbol{\theta}_{B_e})$. The latter is the parameter vector of interest, that one aims to estimate. The dependance on $\boldsymbol{\theta}$ will be omitted in the following, if there is no ambiguity. It is worth noting that the KP structure is by nature scale ambiguous, because

$$(\gamma \mathbf{A}) \otimes \left(\gamma^{-1}\mathbf{B}\right) = \mathbf{A} \otimes \mathbf{B} \quad \text{for any } \gamma \neq 0.$$

Since we aim to study estimators of the structured scatter matrix and not of the parameter vector, this ambiguity is irrelevant. However in the following analysis, a constraint ensuring the model identifiability w.r.t. $\boldsymbol{\theta}$, and thus the bijectivity of $\boldsymbol{\theta} \mapsto \mathbf{R}\,(\boldsymbol{\theta})$, will be added when required. In the below derivations, we will consider the following constraint

(C) $\quad c\,(\boldsymbol{\theta}) \triangleq \dfrac{1}{n}\text{vec}\,(\mathbf{I}_{n_A})^T \mathbf{P}_A \boldsymbol{\theta}_A - 1 = 0 \quad \Leftrightarrow \quad \text{Tr}\,(\mathbf{A}) = n.$

Note that the considered misspecification only concerns the density generator function and not the scatter matrix structure, since the latter is usually derived from physical considerations on the measurement system.

The minimization of the *assumed* likelihood function w.r.t. $\boldsymbol{\theta}$ is generally a complex and time-consuming task. To bypass this issue, we propose a new and tractable estimation procedure in the next section. Furthermore, our approach is still usable when the access to raw data is impossible, which is especially relevant in applications such as radio-astronomy [37], [38], where the data is too large to be handled and where practitioners have only access to intermediate estimates $\widehat{\mathbf{R}}$.

III. PROPOSED ALGORITHM

In this section, we introduce a structured estimator of the scatter matrix $\mathbf{R}_e$, which consists of a procedure with two main steps. The latter takes into account the underlying KP structure and the potential refined structure of the Kronecker factors. Specifically, we first compute an unstructured estimate of $\mathbf{R}_e$, denoted by $\widehat{\mathbf{R}}$. Then, we impose the structure by solving an appropriate weighted covariance fitting, inspired from [7]. For notational accommodation, we skip the dependence on K for the estimators obtained from K samples when there is no ambiguity.

A. Step 1: unstructured M-estimation of R_e

From a set of K i.i.d. $\mathbf{y}_k \sim \mathbb{C}\mathcal{ES}_m\,(\mathbf{0}, \mathbf{R}_e, g)\,, k = 1, \ldots, K$ with $K > m \triangleq np$, an unstructured M-estimator of the scatter matrix is given by the solution of the following fixed-point equation [18]:

$$\widehat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^{K} u\left(\mathbf{y}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}_k \right) \mathbf{y}_k \mathbf{y}_k^H \triangleq \mathcal{H}_K(\widehat{\mathbf{R}}) \tag{12}$$

where the choice of the function $u(\cdot)$ is quite loose, while some constraints are met [18], [39]. Specifically, the function $u(\cdot)$ must be non-negative, continuous on $[0, \infty)$, non-increasing such that $\psi(s) = su(s)$ is increasing, upper bounded, say, by $K_\psi < \infty$ and strictly increasing on the interval where $\psi(s) < K_\psi$. Existence and uniqueness of the solution of (12) have been studied in [18]. In addition, the latter can be easily computed with the iterative algorithm $\mathbf{R}_{i+1} = \mathcal{H}_K(\mathbf{R}_i)$ which converges to $\widehat{\mathbf{R}}$ whatever the initialization [18]. Furthermore, the asymptotic behavior of $\widehat{\mathbf{R}}$ is well-known [18, Theorem 9], specifically:

$$\widehat{\mathbf{R}} \xrightarrow{\mathcal{P}} \sigma^{-1}\mathbf{R}_e \text{ and } \sqrt{K}\left(\hat{\mathbf{r}} - \sigma^{-1}\mathbf{r}_e\right) \xrightarrow{\mathcal{L}} \mathcal{GCN}\,(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}\mathbf{K}_{m,m}) \tag{13}$$

where $\hat{\mathbf{r}} = \text{vec}\left(\widehat{\mathbf{R}}\right)$, $\mathbf{r}_e = \text{vec}\,(\mathbf{R}_e)$, $\mathbf{W}_{\mathbf{R}_e} = \mathbf{R}_e^T \otimes \mathbf{R}_e$, $\boldsymbol{\Sigma} = \sigma^{-2}\left(\sigma_1 \mathbf{W}_{\mathbf{R}_e} + \sigma_2 \mathbf{r}_e\,\mathbf{r}_e^H\right)$ and

$$\begin{cases} \sigma_1 = \dfrac{a_1(m+1)^2}{(a_2+m)^2}, \ \sigma_2 = \dfrac{a_1 - 1}{a_2^2} - a_1(a_2 - 1)\dfrac{m + (m+2)a_2}{a_2^2(a_2+m)^2}, \\ a_1 = \dfrac{\mathbb{E}\left[\psi^2\left(\sigma|\mathbf{t}|^2\right)\right]}{m(m+1)}, \ a_2 = \dfrac{1}{m}\mathbb{E}\left[\sigma|\mathbf{t}|^2\psi'\left(\sigma|\mathbf{t}|^2\right)\right] > 0 \end{cases}$$

in which σ is the solution of $\mathbb{E}\left[\psi(\sigma|\mathbf{t}|^2)\right] = m$ with $\mathbf{t} \sim \mathbb{C}\mathcal{ES}_m\,(\mathbf{0}, \mathbf{I}, g)$ and $\psi(s) = su(s)$. For the particular case of unstructured ML estimation, obtained when $u(s) = -g'(s)/g(s) \triangleq u_{\text{ML}}(s)$ in (12) and thus $\psi(\cdot) = \psi_{\text{ML}}(\cdot)$, a subscript $_{\text{ML}}$ is added for $\sigma_{2,\text{ML}}$, which can also be expressed as

$$\sigma_{2,\text{ML}} = -\frac{\sigma_{1,\text{ML}}\left(1 - \sigma_{1,\text{ML}}\right)}{1 + m\left(1 - \sigma_{1,\text{ML}}\right)}.$$

For common CES distributions, explicit expressions of $\sigma_{1,\text{ML}}$ and $\sigma_{2,\text{ML}}$ are given in [22, Table 1].

B. Step 2: covariance fitting

For the second step, we enforce the KP structure with linearly parameterized factors by minimizing the following covariance matching

$$\mathcal{V}\,(\boldsymbol{\theta}) = \left\|\mathbf{Q}^{1/2}\text{vec}\left(\widehat{\mathbf{R}} - \mathbf{A}\,(\boldsymbol{\theta}_A) \otimes \mathbf{B}\,(\boldsymbol{\theta}_B)\right)\right\|_2^2 \tag{14}$$

As detailed later in Section V, the obtained estimate may be statistically efficient, depending on the choice of the matrix $\mathbf{Q}$. The minimization of (14) can be achieved through an alternating optimization scheme. However, with an appropriate choice of $\mathbf{Q}$, a minimizer of (14) can be efficiently found using the Singular Value Decomposition (SVD) [40].

First, let us introduce the following rearrangement of a matrix $\mathbf{R}$ [40]:

$$\mathcal{R}(\mathbf{R}) = \left[\text{vec}\left(\mathbf{R}^{11}\right) \ldots \text{vec}\left(\mathbf{R}^{n1}\right) \text{vec}\left(\mathbf{R}^{12}\right) \ldots \text{vec}\left(\mathbf{R}^{nn}\right)\right]^T$$

where $\mathbf{R}^{k\ell}$ is $(k, \ell)^{\text{th}}$ block of size $p \times p$ of $\mathbf{R}$ satisfying the following properties

$$\mathcal{R}(\mathbf{A} \otimes \mathbf{B}) = \text{vec}(\mathbf{A})\,\text{vec}(\mathbf{B})^T \quad \text{and} \quad \text{vec}(\mathbf{R}) = \mathcal{P}_R \text{vec}(\mathcal{R}(\mathbf{R}))$$

with $\mathcal{P}_R$ a permutation matrix.

Now, we assume that the matrix $\mathbf{Q}$ can be written as

$$\mathbf{Q} = \mathbf{R}^{-T} \otimes \mathbf{R}^{-1} \qquad (15)$$

where $\mathbf{R}$ also possesses a KP structure, i.e. $\mathbf{R} = \mathbf{A} \otimes \mathbf{B}$. A method providing such a matrix $\mathbf{Q}$ is proposed just after. With this particular structure of $\mathbf{Q}$, the function (14) can be equivalently reformulated as a rank-one approximation problem (See [7, Section V.])

$$\mathcal{V}_1(\theta) = \left\|\mathbf{Q}_A^H \mathcal{R}\left(\check{\mathbf{R}}\right)\mathbf{Q}_B^* - \mathbf{T}_A \theta_A \theta_B^T \mathbf{T}_B^T\right\|_F^2 \qquad (16)$$

where

$$\begin{cases} \check{\mathbf{R}} = \left(\mathbf{A}^{-1/2} \otimes \mathbf{B}^{-1/2}\right)\widehat{\mathbf{R}}\left(\mathbf{A}^{-1/2} \otimes \mathbf{B}^{-1/2}\right), \\ \mathbf{Q}_A \mathbf{T}_A = \left(\mathbf{A}^{-T/2} \otimes \mathbf{A}^{-1/2}\right)\mathbf{P}_A, \; \mathbf{Q}_B \mathbf{T}_B = \left(\mathbf{B}^{-T/2} \otimes \mathbf{B}^{-1/2}\right)\mathbf{P}_B \end{cases}$$

in which $\mathbf{Q}_A \in \mathbb{C}^{n^2 \times n_A}$ and $\mathbf{Q}_B \in \mathbb{C}^{p^2 \times n_B}$ have orthonormal columns vectors and $\mathbf{T}_A$ and $\mathbf{T}_B$ are invertible matrices. Since minimizing (16) is a rank-one approximation problem, this can be efficiently achieved using an SVD of the matrix $\mathbf{Q}_A^H \mathcal{R}\left(\check{\mathbf{R}}\right)\mathbf{Q}_B^*$ [40].

To obtain the desired matrix $\mathbf{Q}$ as in (15), it is simply enough to minimize, a first time, (14) or equivalently (16) with $\mathbf{Q} = \mathbf{I}_{m^2}$. To resume, Step 2 is divided into two sub-steps:

1) Minimize (14) with $\mathbf{Q} = \mathbf{I}_{m^2}$, which yields a minimizer $\widehat{\theta}_{\text{int}} = \left[\widehat{\theta}_{A,1}^T, \widehat{\theta}_{B,1}^T\right]^T$ and build $\widehat{\mathbf{Q}} = \widehat{\mathbf{R}}_{\text{int}}^{-T} \otimes \widehat{\mathbf{R}}_{\text{int}}^{-1}$ where $\widehat{\mathbf{R}}_{\text{int}} = \mathbf{R}\left(\widehat{\theta}_{\text{int}}\right) \triangleq \mathbf{A}\left(\widehat{\theta}_{A,1}\right) \otimes \mathbf{B}\left(\widehat{\theta}_{B,1}\right)$

2) Minimize (14) with $\mathbf{Q} = \widehat{\mathbf{Q}}$, which provides $\widehat{\theta} = \left[\widehat{\theta}_A^T, \widehat{\theta}_B^T\right]^T$ and therefore the final structured estimate $\widehat{\mathbf{R}}_{KPr} = \mathbf{R}\left(\widehat{\theta}\right) \triangleq \mathbf{A}\left(\widehat{\theta}_A\right) \otimes \mathbf{B}\left(\widehat{\theta}_B\right)$.

Remark. *It is worth noting that the minimized function (16) does not coincide with the EXIP approach, since the weighted matrix $\mathbf{Q}$ is not the inverse of the (asymptotic) covariance matrix of $\text{vec}\left(\widehat{\mathbf{R}} - \mathbf{R}_e\right)$ given by (13). An EXIP-based procedure could be applicable, but the equivalence between (14) and (16) would be lost and would require an alternating optimization scheme to solve the corresponding weighted covariance fitting. Similarly, it is not a particular case of [22], since the KP is not a convex structure.*

Remark. *We will show, in the next sections, that the intermediary estimate $\widehat{\mathbf{R}}_{int}$ is only consistent, whereas $\widehat{\mathbf{R}}_{Kpr}$ may also be asymptotically efficient.*

Algorithm 1 Algorithm for the scatter matrix

Input: K i.i.d. data, $\mathbf{y}_k \sim \mathcal{CES}_m\left(\mathbf{0}, \mathbf{R}_e, g\right)$ with $K > m$
1: Compute $\widehat{\mathbf{R}}$ from $\mathbf{y}_1, \ldots, \mathbf{y}_K$ with (12).
2: Find a minimizer $\widehat{\theta}_{\text{int}} = \left[\widehat{\theta}_{A,1}^T, \widehat{\theta}_{B,1}^T\right]^T$ of (14) with $\mathbf{Q} = \mathbf{I}_{m^2}$.
   Build $\widehat{\mathbf{Q}} = \widehat{\mathbf{R}}_{\text{int}}^{-T} \otimes \widehat{\mathbf{R}}_{\text{int}}^{-1}$ where $\widehat{\mathbf{R}}_{\text{int}} = \mathbf{R}\left(\widehat{\theta}_{\text{int}}\right) \triangleq \mathbf{A}\left(\widehat{\theta}_{A,1}\right) \otimes \mathbf{B}\left(\widehat{\theta}_{B,1}\right)$.
   Find a minimizer $\widehat{\theta} = \left[\widehat{\theta}_A^T, \widehat{\theta}_B^T\right]^T$ of (14) with $\mathbf{Q} = \widehat{\mathbf{Q}}$.
3: (Optional) Compute iteratively the solution of (17) N_{it} times.
4: **Return:** $\widehat{\mathbf{R}}_{(R)-KPr} = \mathbf{R}\left(\widehat{\theta}\right) = \mathbf{A}\left(\widehat{\theta}_A\right) \otimes \mathbf{B}\left(\widehat{\theta}_B\right)$

C. Step 3 (optional): iteration of Step 2

In this subsection, we propose an additional optional step for the estimator, which is based on an iterative refinement of the cost function $\mathcal{V}(\theta)$. As set out futher in Section IV, this extra step provides an estimator, which preserves the asymptotic behavior but empirically improves the performance in the pre-asymptotic regime.

In the cost function $\mathcal{V}(\theta)$ in (14), the weighted matrix $\mathbf{Q}$ can be viewed as a metric specification. Consequently, better the estimate $\mathbf{Q}$ is, the better a solution $\widehat{\theta}$ minimizing (14) should be. Moreover, to keep benefiting from the equivalence between the minimization problems in (14) and (16), the matrix $\mathbf{Q}$ has to be doubled KP structured as in (15). Thus, we proposed to iteratively substitute $\mathbf{R}\left(\widehat{\theta}_{\text{int}}\right)$ in $\widehat{\mathbf{Q}}$ by $\mathbf{R}\left(\widehat{\theta}\right)$ obtained at the previous iteration. Specifically, for a finite number N_{it} of iterations, we solve, at the ℓ-th iteration, the problem

$$\widehat{\theta}_{\ell+1} = \arg\min_\theta \left\|\widehat{\mathbf{Q}}_\ell^{1/2} \text{vec}\left(\widehat{\mathbf{R}} - \mathbf{A}(\theta_A) \otimes \mathbf{B}(\theta_B)\right)\right\|_2^2 \qquad (17)$$

where $\widehat{\mathbf{Q}}_\ell = \mathbf{R}\left(\widehat{\theta}_\ell\right)^{-T} \otimes \mathbf{R}\left(\widehat{\theta}_\ell\right)^{-1}$ and with $\widehat{\mathbf{Q}}_0 = \widehat{\mathbf{Q}}$. This yields the final structured estimate $\widehat{\mathbf{R}}_{R-KPr} = \mathbf{R}\left(\widehat{\theta}_{N_{\text{it}}}\right)$. Since we assume $N_{\text{it}} < \infty$, the existence of $\widehat{\theta}_{N_{\text{it}}}$ is always ensured. In practice, we can use a more flexible implementation rather than imposing a fixed number of iterations. For example, the stopping criterion can combine a maximal number of iterations, N_{it}, and a relative gap between the estimates of two successive iterations below a defined threshold, ε_{tol}, i.e., $\left\|\widehat{\theta}_{\ell+1} - \widehat{\theta}_\ell\right\| \leq \varepsilon_{\text{tol}} \left\|\widehat{\theta}_\ell\right\|$. The proposed algorithm is summarized in the box **Algorithm 1**.

Remark. *The proposed algorithm differs from the one given in [7] in several points. First, we leverage the entire family of M-estimators, instead of the sample covariance matrix, which brings robustness. Second, the intermediary estimate $\widehat{\mathbf{R}}_{int}$ possesses the full structure (KP + linearly parameterized factors) and not just the KP one. Last, we propose an optional recursive step, which provides better results at low sample*

support.

IV. ASYMPTOTIC BEHAVIOR OF THE PROPOSED METHOD

This section provides a statistical analysis of the proposed estimator $\widehat{\mathbf{R}}_{\text{KPr}} = \mathbf{R}\left(\widehat{\boldsymbol{\theta}}_K\right)$, where $\widehat{\boldsymbol{\theta}}_K$ is a minimizer of (14) with $\mathbf{Q} = \widehat{\mathbf{Q}} = \widehat{\mathbf{R}}_{\text{int}}^{-T} \otimes \widehat{\mathbf{R}}_{\text{int}}^{-1}$. Specifically, we study the consistency and the asymptotic distribution of the estimate.

Theorem 1. Under the considered data model, the estimator $\widehat{\mathbf{R}}_{\text{KPr}}$ is consistent w.r.t. $\sigma^{-1}\mathbf{R}_{\text{e}}$. In addition, if $\widehat{\boldsymbol{\theta}}_K$ satisfies the constraint (C), then $\widehat{\boldsymbol{\theta}}_K \overset{\mathcal{P}}{\to} \boldsymbol{\theta}_c$ such that $\mathbf{R}_c \triangleq \mathbf{R}(\boldsymbol{\theta}_c) = \sigma^{-1}\mathbf{R}_{\text{e}}$ and $\boldsymbol{\theta}_c$ meet the constraint (C). Recall that σ is the solution of $\mathbb{E}\left[\psi(\sigma|\mathbf{t}|^2)\right] = m$ with $\mathbf{t} \sim \mathcal{CES}_m(\mathbf{0}, \mathbf{I}, g)$ and $\psi(s) = su(s)$.

Proof. First, the consistency of $\widehat{\mathbf{R}}$ leads straightforwardly to the one of $\widehat{\mathbf{R}}_{\text{int}}$ for $\sigma^{-1}\mathbf{R}_{\text{e}}$ and thus $\widehat{\mathbf{Q}} \overset{\mathcal{P}}{\to} \sigma^2\mathbf{W}_{\mathbf{R}_{\text{e}}}^{-1}$, which is a full-rank matrix. Then, in a similar manner, we obtain $\widehat{\mathbf{R}}_{\text{KPr}} \overset{\mathcal{P}}{\to} \sigma^{-1}\mathbf{R}_{\text{e}}$. Moreover, if $\widehat{\boldsymbol{\theta}}_K$ satisfies the constraint (C), then the continuous mapping yields $\widehat{\boldsymbol{\theta}}_K \overset{\mathcal{P}}{\to} \boldsymbol{\theta}_c$. ∎

Remark. *Since the parameterization is linear w.r.t. $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_B$, this easily leads to $\boldsymbol{\theta}_c = \begin{bmatrix} \boldsymbol{\theta}_{A_c} \\ \boldsymbol{\theta}_{B_c} \end{bmatrix} = \begin{pmatrix} \mathbf{I}_{n_A} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma}\mathbf{I}_{n_B} \end{pmatrix} \begin{bmatrix} \boldsymbol{\theta}_{A_e} \\ \boldsymbol{\theta}_{B_e} \end{bmatrix}$, where $\boldsymbol{\theta}_e$ is assumed to meet the constraint (C).*

Theorem 2. Under the considered data model, the asymptotic distribution of $\widehat{\mathbf{R}}_{\text{KPr}}$ is given by

$$\sqrt{K}\text{vec}\left(\widehat{\mathbf{R}}_{\text{KPr}} - \mathbf{R}_c\right) \overset{\mathcal{L}}{\to} \mathcal{GCN}\left(\mathbf{0}, \boldsymbol{\Xi}, \boldsymbol{\Xi}\mathbf{K}_{np,np}\right) \text{ with}$$

$$\boldsymbol{\Xi} = \sigma^{-2}\mathcal{P}_{\mathbf{R}}\left(\frac{\sigma_1}{p}\left(\mathbf{b}_{\text{e}}\mathbf{b}_{\text{e}}^H\right) \otimes \mathbf{P}_A\mathbf{F}_A^{-1}\mathbf{P}_A^H + \frac{\sigma_1}{n}\mathbf{P}_B\mathbf{F}_B^{-1}\mathbf{P}_B^H \otimes \left(\mathbf{a}_{\text{e}}\mathbf{a}_{\text{e}}^H\right)\right.$$

$$\left. + \left(\sigma_2 - \frac{\sigma_1}{np}\right)\left(\mathbf{b}_{\text{e}}\mathbf{b}_{\text{e}}^H\right) \otimes \left(\mathbf{a}_{\text{e}}\mathbf{a}_{\text{e}}^H\right)\right)\mathcal{P}_{\mathbf{R}}^H \qquad (18)$$

where $\mathbf{a}_{\text{e}} = \text{vec}(\mathbf{A}_{\text{e}})$, $\mathbf{b}_{\text{e}} = \text{vec}(\mathbf{B}_{\text{e}})$, $\mathbf{F}_A = \mathbf{P}_A^H\mathbf{W}_{\mathbf{A}_{\text{e}}}^{-1}\mathbf{P}_A$, $\mathbf{F}_B = \mathbf{P}_B^H\mathbf{W}_{\mathbf{B}_{\text{e}}}^{-1}\mathbf{P}_B$ in which $\mathbf{W}_{\mathbf{A}_{\text{e}}} = \mathbf{A}_{\text{e}}^T \otimes \mathbf{A}_{\text{e}}$ and $\mathbf{W}_{\mathbf{B}_{\text{e}}} = \mathbf{B}_{\text{e}}^T \otimes \mathbf{B}_{\text{e}}$.

Proof. The proof of Theorem 2 is quite standard, however a few subtleties should be carrefully addressed. Here, we describe the main steps for the considered case. In order to apply the Delta method [41], we perform a Taylor expansion around $\boldsymbol{\theta}_c$ of the gradient of the cost function $\mathcal{V}(\boldsymbol{\theta})$ evaluated in $\widehat{\boldsymbol{\theta}}_K$, which satisfies the constraint (C):

$$0 = \left.\frac{\partial\mathcal{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\widehat{\boldsymbol{\theta}}_K} = \left.\frac{\partial\mathcal{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_c} + \left.\frac{\partial^2\mathcal{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\right|_{\xi_K}\left(\widehat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_c\right)$$

with ξ_K on the line segment connecting $\boldsymbol{\theta}_c$ and $\widehat{\boldsymbol{\theta}}_K$, i.e., $\exists c \in]0,1[$ such that $\xi_K = c\boldsymbol{\theta}_c + (1-c)\widehat{\boldsymbol{\theta}}_K$ [42, Theorem 5.4.8]. Thus the consistency of $\widehat{\boldsymbol{\theta}}_K$ implies $\xi_K \overset{\mathcal{P}}{\to} \boldsymbol{\theta}_c$. On one hand, the consistency of $\widehat{\mathbf{R}}$ yields

$$\mathbf{H}(\xi_K) = \left.\frac{\partial^2\mathcal{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\xi_K} \overset{\mathcal{P}}{\to} 2\boldsymbol{\Gamma}_c^H\mathbf{P}^H\mathbf{W}_{\mathbf{R}_c}^{-1}\mathbf{P}\boldsymbol{\Gamma}_c \triangleq \mathbf{H}_\infty$$

with $\mathbf{P} = \mathcal{P}_{\mathbf{R}}(\mathbf{P}_B \otimes \mathbf{P}_A)$, $\boldsymbol{\Gamma}_c \triangleq \boldsymbol{\Gamma}(\boldsymbol{\theta}_c) = \left[\boldsymbol{\theta}_{B_c} \otimes \mathbf{I}_{n_A}, \ \mathbf{I}_{n_B} \otimes \boldsymbol{\theta}_{A_c}\right]$ and $\mathbf{W}_{\mathbf{R}_c} = \mathbf{R}_c^T \otimes \mathbf{R}_c$. Note that the matrix $\mathbf{H}_\infty$ is singular and real-valued. On the other hand, we obtain similarly as in [22, Section 4.2]

$$\begin{cases} \sqrt{K}\mathbf{g}_K(\boldsymbol{\theta}_c) = \sqrt{K}\left.\dfrac{\partial\mathcal{V}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_c} \overset{\mathcal{L}}{\to} \mathcal{N}\left(\mathbf{0}, \mathbf{R}_\infty\right) \text{ with} \\ \mathbf{R}_\infty = 4\boldsymbol{\Gamma}_c^H\mathbf{P}^H\left[\sigma_1\mathbf{W}_{\mathbf{R}_c}^{-1} + \sigma_2\text{vec}\left(\mathbf{R}_c^{-1}\right)\text{vec}\left(\mathbf{R}_c^{-1}\right)^H\right]\mathbf{P}\boldsymbol{\Gamma}_c \end{cases}$$

Furthermore, we can write, as in [7, Section VII]

$$\sqrt{K}\mathbf{P}\boldsymbol{\Gamma}_c\left(\widehat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_c\right) = \mathbf{P}\boldsymbol{\Gamma}_c\mathbf{H}^\dagger(\xi_K)\,\sqrt{K}\mathbf{g}_K(\boldsymbol{\theta}_c)$$

which leads, by the Delta-method derived for complex-valued parameters [43], to

$$\sqrt{K}\mathbf{P}\boldsymbol{\Gamma}_c\left(\widehat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_c\right) \overset{\mathcal{L}}{\to} \mathcal{GCN}\left(\mathbf{0}, \boldsymbol{\Xi}, \boldsymbol{\Xi}\mathbf{K}_{np,np}\right) \qquad (19)$$

where $\boldsymbol{\Xi} = \mathbf{P}\boldsymbol{\Gamma}_c\mathbf{H}_\infty^\dagger\mathbf{R}_\infty\mathbf{H}_\infty^\dagger\boldsymbol{\Gamma}_c^H\mathbf{P}^H$, can be simplified as requested, by using Lemma 1 in Appendix A with $\alpha_1 = 1$ and $\alpha_2 = 0$. The proof is concluded by performing a Taylor expansion around $\boldsymbol{\theta}_c$ of $\text{vec}(\mathbf{R}(\boldsymbol{\theta})) = \mathbf{P}\text{vec}\left(\boldsymbol{\theta}_A\boldsymbol{\theta}_B^T\right)$ evaluated in $\widehat{\boldsymbol{\theta}}_K$, which yields

$$\text{vec}\left(\widehat{\mathbf{R}}_{\text{KPr}} - \mathbf{R}_c\right) = \mathbf{P}\left.\frac{\partial\boldsymbol{\theta}_B \otimes \boldsymbol{\theta}_A}{\partial\boldsymbol{\theta}^T}\right|_{\xi_K}\left(\widehat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_c\right)$$

with $\left.\dfrac{\partial\boldsymbol{\theta}_B \otimes \boldsymbol{\theta}_A}{\partial\boldsymbol{\theta}^T}\right|_{\xi_K} \overset{\mathcal{P}}{\to} \boldsymbol{\Gamma}_c$ and thus, by Slutsky's lemma,

$$\sqrt{K}\text{vec}\left(\widehat{\mathbf{R}}_{\text{KPr}} - \mathbf{R}_c\right) \overset{\mathcal{L}}{\to} \mathcal{GCN}\left(\mathbf{0}, \boldsymbol{\Xi}, \boldsymbol{\Xi}\mathbf{K}_{np,np}\right). \qquad (20)$$
∎

Remark. *The exhibited multivariate non-circular complex Gaussian distribution is degenerate w.r.t the Lebesgue measure, since the matrix $\boldsymbol{\Xi}$ is not full rank. This is due to the fact that $\mathbf{R}$ is parameterized by $\boldsymbol{\theta}$ and then considering all the elements of $\mathbf{R}$ brings redundancies.*

Corollary 1. The estimator $\widehat{\mathbf{R}}_{\text{R-KPr}}$ obtained with the additional Step 3 is consistent and follows the same asymptotic Gaussian distribution as in (20).

Proof. Theorems 1 and 2 are valid at each iteration, which concludes the proof. ∎

Remark. *While a covariance/scatter matrix is known to be positive semi-definite (p.s.d.), this constraint is not enforced on the estimate provided by* **Algorithm 1***. However, due to the consistency of $\widehat{\mathbf{R}}_{\text{KPr}}$, the latter will be p.s.d. for sufficiently large K with a probability arbitrarily close to one.*

V. STATISTICAL EFFICIENCY OF THE METHOD

In this section, we examine how the proposed estimator is doing compared to the ultimate achievable performance. Specifically, we compare the asymptotic error covariance matrix $\boldsymbol{\Xi}$ exhibited in the previous section with the CRB depending on the *assumed* model and the choice of the function $u(\cdot)$ in (12).

Proposition 1. Let us consider K i.i.d. centered CES distributed observations, denoted by $\mathbf{y}_k \sim \mathcal{CES}_m(\mathbf{0}, \mathbf{R}_{\text{e}}, g)$,

$k = 1, \ldots, K$. The CRB for $\mathrm{vec}\,(\mathbf{R_e})$, denoted by $\mathbf{CRB_{r_e}}$, is given by:

$$\mathbf{CRB_{r_e}} = \frac{1}{K}\mathcal{P}_R\left(\frac{\sigma_{1,\mathrm{ML}}}{p}\left(\mathbf{b_e}\mathbf{b_e}^H\right)\otimes\mathbf{P}_A\mathbf{F}_A^{-1}\mathbf{P}_A^H + \frac{\sigma_{1,\mathrm{ML}}}{n}\mathbf{P}_B\mathbf{F}_B^{-1}\mathbf{P}_B^H\otimes\left(\mathbf{a_e}\mathbf{a_e}^H\right)\right.$$
$$\left. + \left(\sigma_{2,\mathrm{ML}} - \frac{\sigma_{1,\mathrm{ML}}}{np}\right)\left(\mathbf{b_e}\mathbf{b_e}^H\right)\otimes\left(\mathbf{a_e}\mathbf{a_e}^H\right)\right)\mathcal{P}_R^H \qquad (21)$$

Proof. From (3), the unconstrained FIM is given by

$$\mathbf{F}\left(\boldsymbol{\theta}_\mathrm{e}\right) = K\boldsymbol{\Gamma}_\mathrm{e}^H\mathbf{P}^H\left(\sigma_{1,\mathrm{ML}}^{-1}\mathbf{W}_{\mathbf{R}_\mathrm{e}}^{-1} + (\sigma_{1,\mathrm{ML}}^{-1} - 1)\mathrm{vec}\left(\mathbf{R}_\mathrm{e}^{-1}\right)\mathrm{vec}\left(\mathbf{R}_\mathrm{e}^{-1}\right)^H\right)\mathbf{P}\boldsymbol{\Gamma}_\mathrm{e}$$

where $\boldsymbol{\Gamma}_\mathrm{e} = \boldsymbol{\Gamma}\left(\boldsymbol{\theta}_\mathrm{e}\right)$. The matrix $\mathbf{F}\left(\boldsymbol{\theta}_\mathrm{e}\right)$ is singular since the mapping between $\boldsymbol{\theta}$ and $\mathbf{R}$ is many-to-one due to the scale ambiguity. In order to obtain the CRB for estimators of $\mathrm{vec}\,(\mathbf{R})$, we use the results of [35] with the CCRB, which yields

$$\mathbf{CRB_{r_e}} = \mathbf{P}\boldsymbol{\Gamma}_\mathrm{e}\mathbf{CCRB}_{\boldsymbol{\theta}_\mathrm{e}}\boldsymbol{\Gamma}_\mathrm{e}^H\mathbf{P}^H = \mathbf{P}\boldsymbol{\Gamma}_\mathrm{e}\mathbf{U}\left(\mathbf{U}^T\mathbf{F}\left(\boldsymbol{\theta}_\mathrm{e}\right)\mathbf{U}\right)^{-1}\mathbf{U}^T\boldsymbol{\Gamma}_\mathrm{e}^H\mathbf{P}^H$$

where $\mathbf{U} \in \mathbb{R}^{(n_A+n_B)\times(n_A+n_B-1)}$ is an orthonormal basis for the nullspace of the gradient of the constraint (C) [35], specifically

$$\mathbf{U}^T\mathbf{U} = \mathbf{I}_{n_A+n_B-1} \quad \text{and} \quad \frac{\partial c\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\mathbf{U} = \mathbf{0}.$$

Without loss of generality, the matrix $\mathbf{U}$ can be chosen structured as

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_A & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_B} \end{bmatrix} \quad \text{with } \mathbf{U}_A \in \mathbb{R}^{n_A\times(n_A-1)} \qquad (22)$$

such that $\mathbf{U}_A^T\mathbf{U}_A = \mathbf{I}_{n_A-1}$. By using Lemma 2 in Appendix B with $\alpha_1 = \sigma_{1,\mathrm{ML}}^{-1} > 0$ and $\alpha_2 = \sigma_{1,\mathrm{ML}}^{-1} - 1 > -\alpha_1/(np)$, we can rewrite $\mathbf{CRB_{re}}$ as stated. ∎

To recall, we considered the case, where there exists a potential misspecification on the density generator function of the statistical data model. Notably, the samples are truly drawn from an *unknown* centered CES distribution $\mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}_\mathrm{e}, g\right)$, but the *assumed* model is provided by $\mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}, g_\mathrm{mod}\right)$. Therefore, we introduce the following mismatched analysis tools.

Proposition 2. With the considered misspecification on the density generator function, the so-called *pseudo-true* parameter vector $\boldsymbol{\theta}_0$, which minimizes the KLD between the *true* and the *assumed* models $\mathcal{D}\left(p_Y\|f_{\boldsymbol{\theta}}\right)$, is defined by

$$\mathbf{R}\left(\boldsymbol{\theta}_0\right) = \sigma^{-1}\mathbf{R}_\mathrm{e} = \mathbf{A}_\mathrm{e}\otimes\left(\sigma^{-1}\mathbf{B}_\mathrm{e}\right) = \mathbf{R}_c \qquad (23)$$

where σ is the solution of $\mathbb{E}\left[\psi_\mathrm{mod}(\sigma|\mathbf{t}|^2)\right] = m$ with $\mathbf{t} \sim \mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{I}, g\right)$ and $\psi_\mathrm{mod}(s) = -sg'_\mathrm{mod}(s)/g_\mathrm{mod}(s)$. To ensure the uniqueness of $\boldsymbol{\theta}_0$, we assume that $\boldsymbol{\theta}_0$ has to satisfy the constraint (C). If the weight function $u(\cdot)$ of (12) is of the form $u(s) = -g'_\mathrm{mod}(s)/g_\mathrm{mod}(s)$, thus $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_c$

Proof. The derivation to obtain (23) is identical to [22, Corollary 1]. ∎

Proposition 3. For the considered misspecification on the data model, the MCRB for $\mathrm{vec}\,(\mathbf{R}_c)$, denoted by $\mathbf{MCRB_{r_c}}$, is given by:

$$\mathbf{MCRB_{r_c}} = \frac{\sigma^{-2}}{K}\mathcal{P}_R\left(\frac{\sigma_1}{p}\left(\mathbf{b_e}\mathbf{b_e}^H\right)\otimes\mathbf{P}_A\mathbf{F}_A^{-1}\mathbf{P}_A^H + \frac{\sigma_1}{n}\mathbf{P}_B\mathbf{F}_B^{-1}\mathbf{P}_B^H\otimes\left(\mathbf{a_e}\mathbf{a_e}^H\right)\right.$$
$$\left. + \left(\sigma_2 - \frac{\sigma_1}{np}\right)\left(\mathbf{b_e}\mathbf{b_e}^H\right)\otimes\left(\mathbf{a_e}\mathbf{a_e}^H\right)\right)\mathcal{P}_R^H \qquad (24)$$

Proof. We obtain the MCRB for $\mathrm{vec}\,(\mathbf{R}_c)$ from the CMCRB for $\boldsymbol{\theta}_c$ by using the following relation

$$\mathbf{MCRB_{r_c}} = \mathbf{P}\boldsymbol{\Gamma}_c\mathbf{CMCRB}_{\boldsymbol{\theta}_c}\boldsymbol{\Gamma}_c^H\mathbf{P}^H$$

where $\mathbf{CMCRB}_{\boldsymbol{\theta}_c}$ can be obtained from (8) and (9)

$$\mathbf{CMCRB}_{\boldsymbol{\theta}_c} = \frac{1}{K}\mathbf{U}\left(\mathbf{U}^T\mathbf{C}_1\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{C}_2\mathbf{U}\left(\mathbf{U}^T\mathbf{C}_1\mathbf{U}\right)^{-1}\mathbf{U}^T$$

in which

$$\begin{cases} \mathbf{C}_1 = \boldsymbol{\Gamma}_c^H\mathbf{P}^H\left(-\frac{a_2+m}{m+1}\mathbf{W}_{\mathbf{R}_c}^{-1} + \frac{1-a_2}{m+1}\mathrm{vec}\left(\mathbf{R}_c^{-1}\right)\mathrm{vec}\left(\mathbf{R}_c^{-1}\right)^H\right)\mathbf{P}\boldsymbol{\Gamma}_c \\ \mathbf{C}_2 = \boldsymbol{\Gamma}_c^H\mathbf{P}^H\left(a_1\mathbf{W}_{\mathbf{R}_c}^{-1} + (a_1-1)\mathrm{vec}\left(\mathbf{R}_c^{-1}\right)\mathrm{vec}\left(\mathbf{R}_c^{-1}\right)^H\right)\mathbf{P}\boldsymbol{\Gamma}_c \end{cases}$$

Similarly, by using Lemma 2 in Appendix B with $\alpha_1 = \frac{a_2+m}{m+1} > 0$ and $\alpha_2 = \frac{a_2-1}{m+1} > -\alpha_1/m$, we simplify $\mathbf{U}\left(\mathbf{U}^T\mathbf{C}_1\mathbf{U}\right)^{-1}\mathbf{U}^T$ and then obtain the intended expression of $\mathbf{MCRB_{r_c}}$. ∎

Theorem 3. Let us consider K i.i.d. centered CES distributed samples, denoted by $\mathbf{y}_k \sim \mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}_\mathrm{e}, g\right)$, $k = 1, \ldots, K$. These observations are assumed to follow a $\mathbb{C}\mathcal{ES}_m\left(\mathbf{0}, \mathbf{R}, g_\mathrm{mod}\right)$ in the estimation process. According to the choice of the function $u(\cdot)$ in (12), we obtain the following statements :

(i) if $g_\mathrm{mod}(\cdot) = g(\cdot)$ and $u(s) = -\frac{g'(s)}{g(s)}$, then $\widehat{\mathbf{R}}_\mathrm{KPr}$ and $\widehat{\mathbf{R}}_\mathrm{R\text{-}KPr}$ are asymptotically efficient.

(ii) if $g_\mathrm{mod}(\cdot) \neq g(\cdot)$ and $u(s) = -\frac{g'_\mathrm{mod}(s)}{g_\mathrm{mod}(s)}$, then $\widehat{\mathbf{R}}_\mathrm{KPr}$ and $\widehat{\mathbf{R}}_\mathrm{R\text{-}KPr}$ are asymptotically m-efficient.

(iii) if $u(\cdot)$ cannot be written as one of the above forms (e.g., Huber's M-estimator [18]), nothing more than Theorem 1 and 2 can be elaborated on the asymptotic behavior of $\widehat{\mathbf{R}}_\mathrm{KPr}$.

Proof. For the case (i), this particular choice of function $u(\cdot)$ coincides with the unstructured ML estimate as step 1. Thus, the coefficient σ is equal to $\sigma = 1$ and $\boldsymbol{\theta}_c = \boldsymbol{\theta}_\mathrm{e}$. Consequently, the asymptotic error covariance in (18) can be rewritten by substituting σ_1 and σ_2 by $\sigma_{1,\mathrm{ML}}$ and $\sigma_{2,\mathrm{ML}}$ and finally coincides with $\mathbf{CRB_{r_e}}$ in (21). For the case (ii), we note that the asymptotic error covariance in (18) coincides with the $\mathbf{MCRB_{r_c}}$ in (24). For the case (iii), the function $u(\cdot)$ is not related to a p.d.f. of a CES distribution. Thus, the mismatched framework can not be applied; however the chosen M-estimator is well-defined so the statistical characterization reported in (18) holds.

The parameter vector $\mathrm{vec}\,(\mathbf{R})$ being complex-valued, the intended results are easily achievable by working on the vector concatenating the real and imaginary parts of $\mathrm{vec}\,(\mathbf{R})$ and noting that $\boldsymbol{\Xi}^T = \mathbf{K}_{np,np}\boldsymbol{\Xi}\mathbf{K}_{np,np}$. ∎

Remark. *Note that the case (i) of Theorem 3 is unrealistic, and has a purely theoretical interest.*

VI. Robust Structured Estimator of the Shape Matrix

In this section, we consider the particular case of KP structured shape matrix robust estimation. Indeed, one may only require an estimate of the shape of the scatter matrix, i.e., to capture the intrinsic structure of the matrix without needing its scale. This occurs for scale-invariant processings, i.e., involving a zero degree homogeneous function of the scatter matrix. For instance, we can cite the adaptive normalized matched filter detector [44] or the MUSIC method. The proposed method for this case is quite similar to the one introduced in Section III. Likewise, we conduct a theoretical asymptotic analysis of the proposed estimator.

A. Reminders on Robust Statistical Model

The class of CES distributions covers various statistical models, due to the diversity of density generator functions. However as studied previously, this may lead to misspecifications if the density generator is improperly chosen. Nevertheless, any CES distributions are related to a common CAE distribution [18]. It is worth noting that this interesting property is only valid for centered CES distributions. Let be $\mathbf{y} \sim \mathbb{CES}_m(\mathbf{0}, \mathbf{R}, g)$, then the normalized vector

$$\mathbf{z} = \frac{\mathbf{y}}{\|\mathbf{y}\|}, \mathbf{y} \neq \mathbf{0},$$

follows a CAE distribution, denoted by $\mathbf{z} \sim \mathbb{CAE}_m(\mathbf{R})$. The r.v. $\mathbf{z}$ has the following p.d.f. w.r.t. spherical measure which is the natural Borel measure on the unit m-sphere $\mathbb{CS}^m$ [45]

$$\mathrm{p}(\mathbf{z} \mid \mathbf{R}) \propto |\mathbf{R}|^{-1} \left(\mathbf{z}^H \mathbf{R}^{-1} \mathbf{z} \right)^{-m} \quad (25)$$

Note that, in this context of CAE distribution, the matrix $\mathbf{R}$ is scale ambiguous, i.e., defined up to an arbitrary scale factor. To avoid scaling ambiguity, we introduce the matrix $\mathbf{V} = m\mathbf{R}/\mathrm{Tr}(\mathbf{R})$, which is referred to as the shape matrix of $\mathbf{z}$.

In addition, for CAE distribution with shape matrix $\mathbf{V}$ parameterized by $\boldsymbol{\mu} \in \mathbb{R}^n$, i.e., $\mathbb{CAE}_m(\mathbf{V}(\boldsymbol{\mu}))$, the FIM can be written as

$$\mathbf{F}_{\mathrm{CAE}}(\boldsymbol{\mu}) = \left(\frac{\partial \, \mathrm{vec}(\mathbf{V}(\boldsymbol{\mu}))}{\partial \boldsymbol{\mu}^T} \right)^H \mathbf{Y}_{\mathrm{CAE}} \frac{\partial \, \mathrm{vec}(\mathbf{V}(\boldsymbol{\mu}))}{\partial \boldsymbol{\mu}^T} \quad (26)$$

where $\mathbf{Y}_{\mathrm{CAE}} = \frac{m}{m+1} \mathbf{W}_{\mathbf{V}}^{-1} - \frac{1}{m+1} \mathrm{vec}(\mathbf{V}^{-1}) \mathrm{vec}(\mathbf{V}^{-1})^H$, which is rank-deficient and $\mathbf{W}_{\mathbf{V}} = \mathbf{V}^T \otimes \mathbf{V}$.

Formally, we consider a shape matrix, exhibiting the same structure as $\mathbf{R}_{\mathrm{e}}$

$$\mathbf{V}_{\mathrm{e}} = \frac{m}{\mathrm{Tr}(\mathbf{R}_{\mathrm{e}})} \mathbf{R}_{\mathrm{e}} = \left(\frac{n\mathbf{A}_{\mathrm{e}}}{\mathrm{Tr}(\mathbf{A}_{\mathrm{e}})} \right) \otimes \left(\frac{p\mathbf{B}_{\mathrm{e}}}{\mathrm{Tr}(\mathbf{B}_{\mathrm{e}})} \right) \triangleq \tilde{\mathbf{A}}_{\mathrm{e}} \otimes \tilde{\mathbf{B}}_{\mathrm{e}} \quad (27)$$

where the parameterizations in (10) and (11) are still relevant for $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$, if the following constraint are met

$$(\mathrm{C1}) \quad \mathbf{c}_1(\boldsymbol{\theta}) \triangleq \begin{bmatrix} \frac{1}{n} \mathrm{vec}(\mathbf{I}_{n_A})^T \mathbf{P}_A \boldsymbol{\theta}_A - 1 \\ \frac{1}{p} \mathrm{vec}(\mathbf{I}_{n_B})^T \mathbf{P}_B \boldsymbol{\theta}_B - 1 \end{bmatrix} = \mathbf{0} \Leftrightarrow \begin{cases} \mathrm{Tr}(\mathbf{A}) = n \\ \mathrm{Tr}(\mathbf{B}) = p \end{cases}$$

which ensures that $\mathrm{Tr}(\mathbf{V}(\boldsymbol{\theta})) = \mathrm{Tr}(\mathbf{A})\mathrm{Tr}(\mathbf{B}) = np = m$.

B. Algorithm

From a set of K i.i.d. $\mathbf{z}_k \sim \mathbb{CAE}_m(\mathbf{V}_{\mathrm{e}}), k = 1, \ldots, K$ with $K > m$, we first compute a ML-type estimate of the shape matrix, for which the full structure $\mathbf{V}(\boldsymbol{\theta})$ is not taken into account. For this, two solutions can be deployed:

- the unstructured ML estimate, coinciding with Tyler's M-estimate [28], which is given by the solution of the following fixed-point equation [18]:

$$\widehat{\mathbf{V}} = \frac{m}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \widehat{\mathbf{V}}^{-1} \mathbf{z}_k} \quad \text{such that} \quad \mathrm{Tr}\left(\widehat{\mathbf{V}}\right) = m \quad (28)$$

  The above solution can be obtained by an iterative algorithm, in a similar manner as (12).
- the KP structured ML estimate, given in [8] and denoted by $\widehat{\mathbf{V}}_{\mathrm{KP}}$. The latter is obtained using an alternating optimization scheme, which converges towards the global minimum of the related likelihood function [8].

In the literature, Tyler's M-estimate is said to be distribution-free, since equation (28) does not involve any data-dependant function g. The latter can be used as shape estimator for any centered multivariate distribution, but it is optimal, in the ML sense, for CAE distribution.

Remark. *Since $\widehat{V}_{KP}$ already possesses the KP structure, it seems more appealing than the unstructured $\widehat{V}$. In addition, the constraint on the number of samples, i.e., $K > m$, can be strongly relaxed for $\widehat{V}_{KP}$ [8]. However, the computational cost of the latter is higher than the one of $\widehat{V}$.*

Then, we apply Step 2 of Section III to obtain an unconstrained but structured $\widehat{\mathbf{R}}_{KPr}$, for which we replace $\widehat{\mathbf{R}}$ by $\widehat{\mathbf{V}}$ or $\widehat{\mathbf{V}}_{\mathrm{KP}}$. Similarly as **Algorithm 1**, the additional/optional iteration of Step 3 is still valid. Finally, we compute $\widehat{\mathbf{V}}_{KPr} = m\widehat{\mathbf{R}}_{KPr}/\mathrm{Tr}\left(\widehat{\mathbf{R}}_{KPr}\right)$. The algorithm can be summarized in a very similar manner as the box **Algorithm 1**.

C. Asymptotic Analysis

Similarly as Section IV, we perform an analysis of the asymptotic performance of $\widehat{\mathbf{V}}_{(R)-KPr}$.

Proposition 4. Let us consider K i.i.d. centered CAE distributed observations, denoted by $\mathbf{z}_k \sim \mathbb{CAE}_m(\mathbf{V}_{\mathrm{e}})$,

$k = 1, \ldots, K$. The CRB for $\mathrm{vec}\,(\mathbf{V}_{\mathrm{e}})$, denoted by $\mathbf{CRB}_{\mathbf{v}_{\mathrm{e}}}$, is given by:

$$\mathbf{CRB}_{\mathbf{v}_{\mathrm{e}}} = \frac{m+1}{mK}\mathcal{P}_R\left(\frac{1}{p}\left(\tilde{\mathbf{b}}_{\mathrm{e}}\tilde{\mathbf{b}}_{\mathrm{e}}^H\right)\otimes\left[\mathbf{P}_A\mathbf{F}_{\tilde{A}}^{-1}\mathbf{P}_A^H + \eta\tilde{\mathbf{a}}_{\mathrm{e}}\tilde{\mathbf{a}}_{\mathrm{e}}^H - \chi_A\right]\right.$$
$$\left. + \frac{1}{n}\left[\mathbf{P}_B\mathbf{F}_{\tilde{B}}^{-1}\mathbf{P}_B^H + \eta_1\tilde{\mathbf{b}}_{\mathrm{e}}\tilde{\mathbf{b}}_{\mathrm{e}}^H - \chi_B\right]\otimes\left(\tilde{\mathbf{a}}_{\mathrm{e}}\tilde{\mathbf{a}}_{\mathrm{e}}^H\right)\right)\mathcal{P}_R^H \quad (29)$$

where $\tilde{\mathbf{a}}_{\mathrm{e}} = \mathrm{vec}\left(\tilde{\mathbf{A}}_{\mathrm{e}}\right)$, $\tilde{\mathbf{b}}_{\mathrm{e}} = \mathrm{vec}\left(\tilde{\mathbf{B}}_{\mathrm{e}}\right)$, $\mathbf{F}_{\tilde{A}} = \mathbf{P}_A^H\mathbf{W}_{\tilde{\mathbf{A}}_{\mathrm{e}}}^{-1}\mathbf{P}_A$, $\mathbf{F}_{\tilde{B}} = \mathbf{P}_B^H\mathbf{W}_{\tilde{\mathbf{B}}_{\mathrm{e}}}^{-1}\mathbf{P}_B$, $\mathbf{W}_{\tilde{\mathbf{A}}_{\mathrm{e}}} = \tilde{\mathbf{A}}_{\mathrm{e}}^T\otimes\tilde{\mathbf{A}}_{\mathrm{e}}$, $\mathbf{W}_{\tilde{\mathbf{B}}_{\mathrm{e}}} = \tilde{\mathbf{B}}_{\mathrm{e}}^T\otimes\tilde{\mathbf{B}}_{\mathrm{e}}$, $\eta = \mathbf{u}_A^T\mathbf{F}_{\tilde{A}}^{-1}\mathbf{u}_A$, $\eta_1 = \mathbf{u}_B^T\mathbf{F}_{\tilde{B}}^{-1}\mathbf{u}_B$, in which $\mathbf{u}_A^T = \mathrm{vec}\,(\mathbf{I}_n)^T\,\mathbf{P}_A/n$, $\mathbf{u}_B^T = \mathrm{vec}\left(\mathbf{I}_p\right)^T\mathbf{P}_B/p$ and with

$$\begin{cases} \chi_A = \mathbf{P}_A\mathbf{F}_{\tilde{A}}^{-1}\mathbf{u}_A\tilde{\mathbf{a}}_{\mathrm{e}}^H + \tilde{\mathbf{a}}_{\mathrm{e}}\mathbf{u}_A^T\mathbf{F}_{\tilde{A}}^{-1}\mathbf{P}_A^H \\ \chi_B = \mathbf{P}_B\mathbf{F}_{\tilde{B}}^{-1}\mathbf{u}_B\tilde{\mathbf{b}}_{\mathrm{e}}^H + \tilde{\mathbf{b}}_{\mathrm{e}}\mathbf{u}_B^T\mathbf{F}_{\tilde{B}}^{-1}\mathbf{P}_B^H \end{cases}.$$

Proof. We obtain the CRB for $\mathrm{vec}\,(\mathbf{V})$ by using the following relation

$$\mathbf{CRB}_{\mathbf{v}_{\mathrm{e}}} = \mathbf{P}\boldsymbol{\Gamma}_{\mathrm{e}}\mathbf{U}_1\left(\mathbf{U}_1^T\mathbf{F}_{\mathrm{CAE}}\left(\boldsymbol{\theta}_{\mathrm{e}}\right)\mathbf{U}_1\right)^{-1}\mathbf{U}_1^T\boldsymbol{\Gamma}_{\mathrm{e}}^H\mathbf{P}^H$$

where $\mathbf{U}_1 \in \mathbb{R}^{(n_A+n_B)\times(n_A+n_B-2)}$ is an orthonormal basis for the nullspace of the gradient of the constraint (C1) [35], specifically

$$\mathbf{U}_1^T\mathbf{U}_1 = \mathbf{I}_{n_A+n_B-2} \quad \text{and} \quad \frac{\partial\mathbf{c}_1\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\mathbf{U}_1 = \begin{bmatrix}\mathbf{u}_A^T & \mathbf{0} \\ \mathbf{0} & \mathbf{u}_B^T\end{bmatrix}\mathbf{U}_1 = \mathbf{0}.$$

Without loss of generality, the matrix $\mathbf{U}_1$ can be chosen structured as

$$\mathbf{U}_1 = \begin{bmatrix}\mathbf{U}_A & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_B\end{bmatrix} \text{ with } \mathbf{U}_A \in \mathbb{R}^{n_A\times(n_A-1)}, \mathbf{U}_B \in \mathbb{R}^{n_B\times(n_B-1)} \quad (30)$$

such that $\mathbf{U}_A^T\mathbf{U}_A = \mathbf{I}_{n_A-1}$ and $\mathbf{U}_B^T\mathbf{U}_B = \mathbf{I}_{n_B-1}$. By using Lemma 3 in Appendix C, we can rewrite $\mathbf{CRB}_{\mathbf{ve}}$ as announced in (29). ∎

Theorem 4. Under the considered CAE model, the proposed shape matrix estimator $\widehat{\mathbf{V}}_{(R)-KPr}$ is a consistent, asymptotically efficient and Gaussian distributed estimate of $\mathbf{V}_{\mathrm{e}}$, for both $\widehat{\mathbf{V}}$ or $\widehat{\mathbf{V}}_{KP}$ as Step 1.

Proof. The consistency is a direct result of Theorem 1 with $\sigma = \mathrm{Tr}\,(\mathbf{R}_{\mathrm{e}})/m$ since $\widehat{\mathbf{V}} \overset{\mathcal{P}}{\to} \mathbf{V}_{\mathrm{e}}$ [18] and $\widehat{\mathbf{V}}_{KP} \overset{\mathcal{P}}{\to} \mathbf{V}_{\mathrm{e}}$ [8], respectively. In order to obtain the asymptotic distribution of $\mathrm{vec}\left(\widehat{\mathbf{V}}_{KPr}\right)$, we reuse Theorem 2 and its proof, giving the asymptotic distribution of $\mathrm{vec}\left(\widehat{\mathbf{R}}_{KPr}\right)$. The derivations are all the same, except the expression of $\mathbf{R}_\infty$, which is given, in this case, by

$$\mathbf{R}_\infty = 4\boldsymbol{\Gamma}_c^H\mathbf{P}^H\mathbf{W}_{\mathbf{V}_e}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}\mathbf{W}_{\mathbf{V}_e}^{-1}\mathbf{P}\boldsymbol{\Gamma}_c.$$

where $\mathbf{W}_{\mathbf{V}_{\mathrm{e}}} = \mathbf{V}_{\mathrm{e}}^T\otimes\mathbf{V}_{\mathrm{e}}$ and $\boldsymbol{\Sigma}_{\mathbf{V}}$ denotes the asymptotic covariance matrix, up to a factor K, of either $\widehat{\mathbf{V}}$ or $\widehat{\mathbf{V}}_{KP}$. For the first case, the expression of $\boldsymbol{\Sigma}_{\mathbf{V}}$ can be found in [18, Section VI.C.]. For the second one, the matrix $\boldsymbol{\Sigma}_{\mathbf{V}}$ is given by (29), derived for the KP structure only, i.e., the

Kronecker factors are simply Hermitian. We recall that

$$\sqrt{K}\mathrm{vec}\left(\widehat{\mathbf{R}}_{KPr} - \sigma^{-1}\mathbf{R}_{\mathrm{e}}\right) \overset{\mathcal{L}}{\to} \mathcal{GCN}\left(\mathbf{0}, \boldsymbol{\Xi}, \boldsymbol{\Xi}\mathbf{K}_{np,np}\right)$$

where $\boldsymbol{\Xi} = \mathbf{P}\boldsymbol{\Gamma}_c\mathbf{H}_\infty^\dagger\mathbf{R}_\infty\mathbf{H}_\infty^\dagger\boldsymbol{\Gamma}_c^H\mathbf{P}^H$ with $\mathbf{H}_\infty = 2\boldsymbol{\Gamma}_c^H\mathbf{P}^H\mathbf{W}_{\mathbf{V}_e}^{-1}\mathbf{P}\boldsymbol{\Gamma}_c$.

Finally, the Delta-method on $\widehat{\mathbf{V}}_{KPr} = m\widehat{\mathbf{R}}_{KPr}/\mathrm{Tr}\left(\widehat{\mathbf{R}}_{KPr}\right)$, leads to

$$\sqrt{K}\mathrm{vec}\left(\widehat{\mathbf{V}}_{KPr} - \mathbf{V}_{\mathrm{e}}\right) \overset{\mathcal{L}}{\to} \mathcal{GCN}\left(\mathbf{0}, \boldsymbol{\Psi}\boldsymbol{\Xi}\boldsymbol{\Psi}^H, \boldsymbol{\Psi}\boldsymbol{\Xi}\boldsymbol{\Psi}^H\mathbf{K}_{np,np}\right).$$

with $\boldsymbol{\Psi} = \mathbf{I} - \frac{1}{m}\mathbf{v}_{\mathrm{e}}\mathrm{vec}\,(\mathbf{I})^T$. After some calculus, we can show that $\boldsymbol{\Psi}\boldsymbol{\Xi}\boldsymbol{\Psi}^H = K\mathbf{CRB}_{\mathbf{v}_{\mathrm{e}}}$ for both $\widehat{\mathbf{V}}$ or $\widehat{\mathbf{V}}_{KP}$ as Step 1, which concludes the proof. With the optional Step 3, the above results are valid at each iteration, which leads to the desired outcome on the performances of $\widehat{\mathbf{V}}_{R-KPr}$. ∎

VII. Simulations

In this section, we illustrate the results of the previous theoretical analysis for a KP structured scatter matrix, whose Kronecker factors are Hermitian Toeplitz. Specifically, we set its first row according to $\{\mathbf{A}_{\mathrm{e}}\}_{1,\ell} = \rho^{|\ell-1|}, \ell = 1, \ldots, n$ with $\rho = 0.8 + 0.3i$ and $\{\mathbf{B}_{\mathrm{e}}\}_{1,\ell} = \rho_1^{|\ell-1|}, \ell = 1, \ldots, p$ with $\rho_1 = 0.5 + 0.3i$. The minimal parameterization for the Toeplitz structure consists in stacking the real and imaginary parts of the first row of the matrix.

A. Matched scenario study with CES distributions

We generate 5000 sets of K i.i.d. t-distributed samples $\mathbf{y}_k \sim \mathbb{C}t_{m,d}\left(\mathbf{0}, \mathbf{R}_{\mathrm{e}}\right), k = 1, \ldots, K$ with $d = 3$ degrees of freedom and $n = p = 3$. We consider the matched scenario, i.e., *assumed* model is the correct one with $g_{\mathrm{mod}}(\cdot) = g(\cdot)$. We compare the performance of the proposed estimators $\widehat{\mathbf{R}}_{(R)-KPr}$ with the related CRB. For $\widehat{\mathbf{R}}_{R-KPr}$ (i.e., with optional Step 3), the implementation uses as stopping criterion a combination of $N_{\mathrm{it}} = 10$ iterations at most and a threshold of $\varepsilon_{\mathrm{tol}} = 10^{-4}$ for the relative gap between the estimates of two successive iterations. We also display the intermediate estimate $\widehat{\mathbf{R}}_{\mathrm{int}}$ in Step 2, as well as the estimate, obtained with $\widehat{\mathbf{Q}} = \mathbf{R}_e^{-T}\otimes\mathbf{R}_e^{-1}$ for a theoretical purpose.

In Fig. 1, the asymptotic covariance of the proposed structured estimate reaches the CRB, i.e., the asymptotic efficiency of the algorithm is empirically verifying Theorem 3. The unbiasedness, as well as the consistency, can also be indirectly observed in Fig. 1. Furthermore, for a very low number of samples (around the limit of existence, i.e., $K \approx m + 1$), the use of the additional Step 3 sometimes seems to underperform the case without Step 3. The latter is probably due to the missing p.s.d. constraint on the estimate. Aside from this small effect, we notice an improvement brought by Step 3 in Fig. 1. We can also note that doing recursions to refine $\widehat{\mathbf{Q}}$ behaves almost the same as using $\widehat{\mathbf{Q}} = \mathbf{R}_e^{-T}\otimes\mathbf{R}_e^{-1}$, where $\mathbf{R}_e$ refers to the *true* scatter matrix. Finally, we see the importance of the weighting matrix $\mathbf{Q}$ for achieving statistical efficiency. Indeed, the intermediate estimate
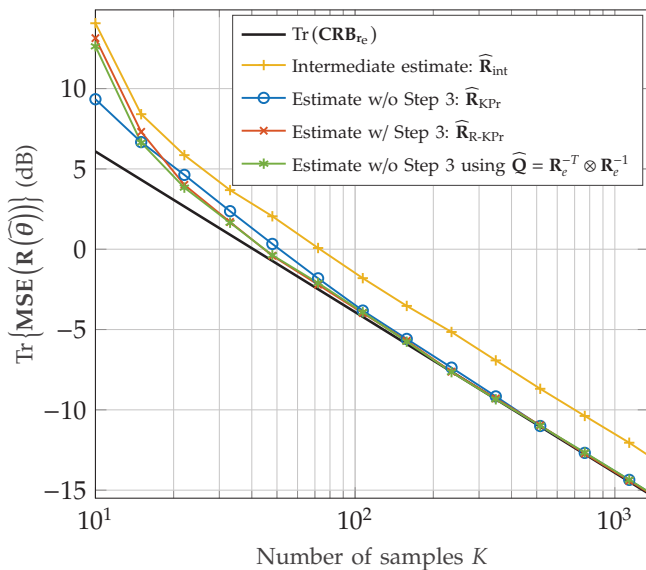
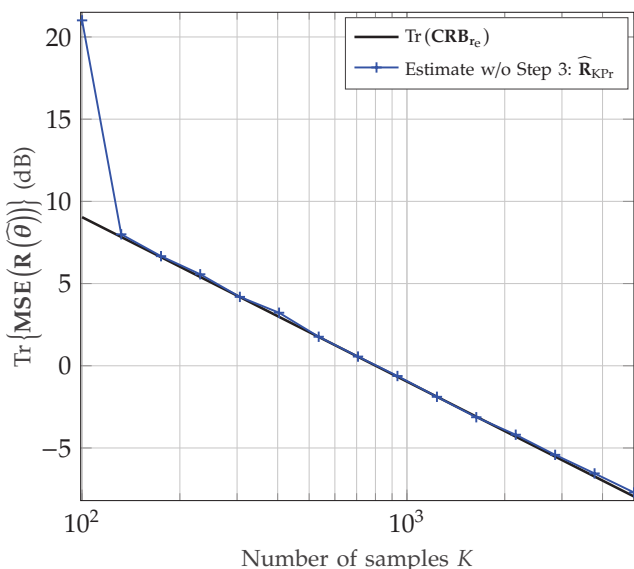Fig. 1: Efficiency simulation for scatter estimate
$(n = p = 3)$

Fig. 3: Efficiency simulation for scatter estimate in mismatched scenarios $(n = p = 3)$

Fig. 2: Efficiency simulation for scatter estimate
$(n = 11, p = 9)$

$\widehat{\mathbf{R}}_{\text{int}}$ obtained with $\mathbf{Q} = \mathbf{I}_{m^2}$ is only consistent but not asymptotically efficient.

To illustrate the interest of our method in high dimension context, we consider the same simulation as before but with $n = 11$ and $p = 9$, leading to a scatter matrix of size around 100×100. In Fig. 2, we can see the same behavior as previously, i.e., the asymptotic efficiency of the method. Furthermore, the asymptotic covariance of $\widehat{\mathbf{R}}_{\text{Kpr}}$ is almost equal to the CRB at only 140 samples. This comes from the KP structure, which strongly reduces the effective dimensions of the estimation problem compared to the apparent dimensions. Thus, for $K > m$ samples (limit of existence), the asymptotic regime can

be rapidly achieved.

B. Mismatched study with CES distributions

In this part, we consider model misspecifications. We generate samples as in the first simulation. Two scenarios have been considered for the simulations.

(S1) the *assumed* p.d.f. is a *t*-distribution but with a different degree of freedom, $d_{\text{mod}} = 20$.

(S2) the *assumed* model is a Gaussian distribution, which would be the most common hypothesis.

For the second scenario, we also display the performance of the estimator given by [7], since the latter addresses the Gaussian case. To draw the comparison, we define the Pseudo Mean Square Error (PMSE) w.r.t $\mathbf{R}_c$ by

$$\mathbf{PMSE}\left(\widehat{\mathbf{R}}_{\text{KPr}}\right) = \mathbb{E}\left[\text{vec}\left(\widehat{\mathbf{R}}_{\text{KPr}} - \mathbf{R}_c\right) \text{vec}\left(\widehat{\mathbf{R}}_{\text{KPr}} - \mathbf{R}_c\right)^H\right].$$

In Fig. 3, we can notice the asymptotic m-efficiency of the different estimators, since the asymptotic covariance of the proposed estimate under mismatched models reaches the corresponding CRB. In addition, the estimator of [7] is outperformed by our method, since the intermediate estimate $\widehat{\mathbf{R}}_{\text{int}}$ possesses the full structure (KP structure + Toeplitz factors) and not only the KP one as in [7].

C. Numerical results for CAE distribution

We generate 5000 sets of K i.i.d. CAE distributed observations $\mathbf{z}_k \sim \mathbb{C}\mathcal{A}\mathcal{E}_m(\mathbf{V}_e)$, $k = 1, \ldots, K$, where $\mathbf{V}_e = m\mathbf{R}_e/\text{Tr}(\mathbf{R}_e)$.

We compare the performance of proposed estimators for both $\widehat{\mathbf{V}}$ or $\widehat{\mathbf{V}}_{\text{KP}}$ as Step 1 with the corresponding CRB and the constrained Tyler's estimate proposed in [24]
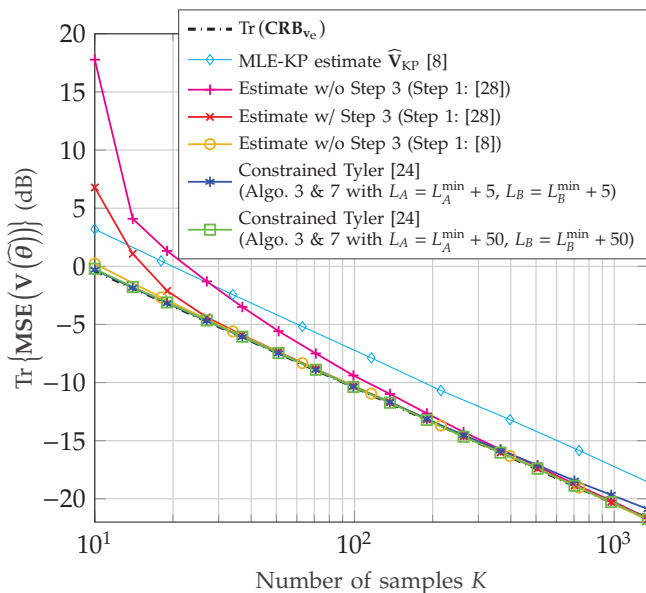
Fig. 4: Efficiency simulation for shape estimate

and in order to obtain a KP structured estimate, where the Kronecker factors are also Toeplitz structured, we intertwine Algorithm 3 into Algorithm 7 of [24]. Note that the Hermitian Toeplitz constraint on the factors is taken into account through a circulant embedding. The size of this circulant matrix is a tunable parameter, with the constraint $L_A \geq 2n - 1 \triangleq L_A^{\min}$ for the factor $\mathbf{A}$ (respectively $L_B \geq 2p - 1 \triangleq L_B^{\min}$ for $\mathbf{B}$).

In Fig. 4, the asymptotic covariance matrix of all proposed estimates reaches the corresponding CRB. The unbiasedness, as well as the consistency, can also be indirectly observed in Fig. 4. The benefit of using a KP structured Step 1 with $\widehat{\mathbf{V}}_{KP}$ is significant, notably at low sample support. This was to be expected, since Step 1 estimate, $\widehat{\mathbf{V}}_{KP}$, is already KP structured. In addition, we can note that constrained Tyler's estimate almost equally performs with respect to the proposed estimate, which uses $\widehat{\mathbf{V}}_{KP}$. For sufficiently large L_A and L_B, the latter also seems to achieve the CRB, while no theoretical guarantee can be easily drawn from this algorithmic approach. For low L_A and L_B, this estimator seems to be asymptotically biased.

Table I summarizes the average calculation time of the different algorithms. It results that the proposed algorithms are an interesting and time-efficient alternative to Constrained Tyler's estimates, while guaranteeing asymptotic performances.

D. MIMO channel covariance estimation

We consider the problem of MIMO channel covariance estimation from training data. In this case, the channel covariance matrix exhibits the KP structure [3], [46]

$$\mathbf{R}_{ch} = \mathbf{A}_e \otimes \mathbf{B}_e \tag{31}$$

where $\mathbf{A}_e \in \mathbb{C}^{5\times5}$ (respectively $\mathbf{B}_e \in \mathbb{C}^{8\times8}$) refers to the transmit (receive) covariance matrix. In addition, we assume that the transmit and receive arrays are linear and uniform, such that $\mathbf{A}$ and $\mathbf{B}$ are Toeplitz structured [47]. We consider a setup similar as the one in [47], where matrices $\mathbf{A}$ and $\mathbf{B}$ are given by:

$$\begin{cases} [\mathbf{A}_e]_{kl} = 0.9^{|k-l|} \exp\left(\frac{j\pi}{4}(k-l)\right) \\ [\mathbf{B}_e]_{kl} = 0.7^{|k-l|} \exp\left(\frac{j\pi}{3}(k-l)\right) \end{cases} \tag{32}$$

We set $n_{\text{train}} = 6$ training symbols per frame, which are assumed to be transmitted at one antenna at a time, the other ones being quiet. The K samples are drawn for a CAE distribution with a shape matrix given by $\mathbf{R}_{ch}$. This could also be interpreted as preprocessing (normalization) of CES distributed data. For the comparison, we consider the unstructured estimators, such as the SCM and Tyler's estimate [28]; the KP-structured Tyler-based estimate $\widehat{\mathbf{V}}_{KP}$ [8] and the full structured estimators such as the SCM-based one [7], Constrained Tyler [24] and finally our estimator $\widehat{\mathbf{V}}_{Kpr}$. In this filtering problem, we use the SNR-loss [48] to evaluate the performance of the different methods. We recall that the latter is defined by:

$$\rho \triangleq \rho\left(\widehat{\mathbf{R}}\right) = \frac{\left|\mathbf{t}^H \widehat{\mathbf{R}}^{-1} \mathbf{t}\right|^2}{\left(\mathbf{t}^H \mathbf{R}_e^{-1}\mathbf{t}\right)\left(\mathbf{t}^H \widehat{\mathbf{R}}^{-1} \mathbf{R}_e \widehat{\mathbf{R}}^{-1}\mathbf{t}\right)} \tag{33}$$

in which the steering vector $\mathbf{t}$ is chosen such that $\mathbf{t} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^m$.
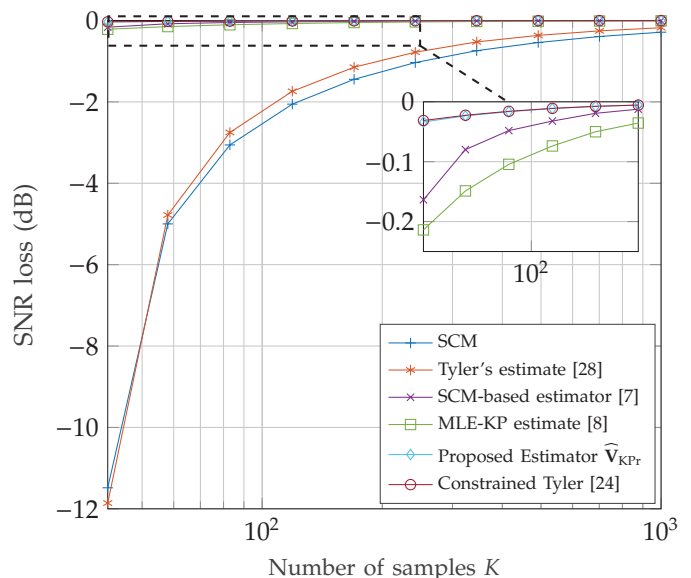

Fig. 5: SNR loss vs number of samples

On Fig. 5, we can see that the unstructured estimates are strongly outperformed by the ones, which take the structure into account. Taking the full structure into account provides better results compared with

K	Algo w/o Step 3 (Step 1: [28])	Algo. w/ Step 3 (Step 1: [28])	Algo w/o Step 3 (Step 1: [8])	Constr. Tyl $L^{min}+5$	Constr. Tyl $L^{min}+50$
50	0.0097s	0.040s	0.062s	0.0462s	0.17s
100	0.012s	0.029s	0.096s	0.061s	0.23s
500	0.014s	0.030s	0.33s	0.21s	0.67s
1000	0.016s	0.024s	0.58s	0.40s	0.89s

TABLE I: Average calculation time
(Matlab R2016b, CPU i5-6440HQ@2.60GHz, 16GB RAM)

KP structure only, even with a Gaussian-based model. Finally, our method yields the same performance as Constrained Tyler, while being an interesting trade-off between asymptotic performances and computational cost.

VIII. Conclusion

In this paper, we addressed robust estimation of KP structured scatter and shape matrices, with potential linearly structured Kronecker factors. Firstly, we have derived a two-step estimation procedure, yielding a structured estimate of scatter matrices. Besides, with a slight modification of the algorithm (iterations of the last step), we have noticed a faster convergence towards the asymptotic regime. Secondly, we have conducted a theoretical asymptotic analysis of the proposed method in terms of consistency and asymptotic Gaussianity. Thirdly, we have analyzed the statistical efficiency of the method for CES distributions depending on the choice made in the first step of the algorithm. This analysis also covers possible model misspecifications. Fourthly, we have studied the counterpart for shape matrices, where we have shown that the method is consistent, asymptotically Gaussian distributed, and efficient. Finally, numerical results corroborated the theoretical analysis and assessed the interest of the proposed algorithms.

Appendix A
Simplification of the Pseudo-inverse

This appendix introduces a lemma, which is required for the proof of Theorem 2.

Lemma 1. *Let α_1 and α_2 be two scalars such that $\alpha_1 > 0$ and $\alpha_2 > -\alpha_1/(np)$. For any admissible θ_1 such that $R_1 = R(\theta_1) = A_1 \otimes B_1$, let us introduce the matrices $\Gamma_1 = \begin{bmatrix} \theta_{B_1} \otimes I_{n_A} & I_{n_B} \otimes \theta_{A_1} \end{bmatrix}$ and $W_{R_1} = R_1^T \otimes R_1$. We have the following equality*

$$P\Gamma_1 \left(\Gamma_1^H P^H \left[\alpha_1 W_{R_1}^{-1} + \alpha_2 \text{vec}\left(R_1^{-1}\right) \text{vec}\left(R_1^{-1}\right)^H \right] P\Gamma_1 \right)^\dagger \Gamma_1^H P^H =$$

$$\mathcal{P}_R \left(\frac{1}{p\alpha_1}\left(b_1 b_1^H\right) \otimes P_A F_A^{-1} P_A^H + \frac{1}{n\alpha_1} P_B F_B^{-1} P_B^H \otimes \left(a_1 a_1^H\right) \right.$$

$$\left. - \left(\frac{\alpha_1 + 2np\alpha_2}{np\alpha_1\left(\alpha_1 + np\alpha_2\right)}\right)\left(b_1 b_1^H\right) \otimes \left(a_1 a_1^H\right) \right) \mathcal{P}_R^H$$

where $a_1 = \text{vec}(A_1)$ and $b_1 = \text{vec}(B_1)$.

Proof. From $\mathcal{P}_R^H W_{R_1}^{-1} \mathcal{P}_R = W_{B_1}^{-1} \otimes W_{A_1}^{-1}$ with $W_{A_1} = A_1^T \otimes A_1$ and $W_{B_1} = B_1^T \otimes B_1$, we can obtain

$$\Gamma_1^H P^H \left[\alpha_1 W_{R_1}^{-1} + \alpha_2 \text{vec}\left(R_1^{-1}\right)\text{vec}\left(R_1^{-1}\right)^H \right] P\Gamma_1 = G + \gamma_B \gamma_A^H$$

where

$$G = \begin{bmatrix} p\alpha_1 F_A + p^2\alpha_2 F_A \theta_{A_1}\theta_{A_1}^H F_A^H & \lambda F_A \theta_{A_1}\theta_{B_1}^H F_B^H \\ 0 & n\alpha_1 F_B + n^2\alpha_2 F_B \theta_{B_1}\theta_{B_1}^H F_B^H \end{bmatrix}$$

$$\triangleq \begin{bmatrix} G_{11} & G_{12} \\ 0 & G_{22} \end{bmatrix},$$

$$\gamma_B = \begin{pmatrix} 0 \\ \sqrt{\lambda} F_B \theta_{B_1} \end{pmatrix}, \quad \gamma_A = \begin{pmatrix} \sqrt{\lambda} F_A \theta_{A_1} \\ 0 \end{pmatrix} \text{ and } \lambda = (\alpha_1 + np\alpha_2).$$

The matrix G is invertible, such that

$$G^{-1} = \begin{bmatrix} G_{11}^{-1} & -G_{11}^{-1}G_{12}G_{22}^{-1} \\ 0 & G_{22}^{-1} \end{bmatrix}$$

in which G_{11}^{-1} and G_{22}^{-1} are given by the Sherman-Morrison formula

$$G_{11}^{-1} = \frac{1}{p}\left(\alpha_1 F_A + p\alpha_2 F_A \theta_{A_1}\theta_{A_1}^H F_A^H\right)^{-1} = \frac{1}{p\alpha_1}F_A^{-1} - \frac{\alpha_2}{\lambda\alpha_1}\theta_{A_1}\theta_{A_1}^H$$

$$G_{22}^{-1} = \frac{1}{n}\left(\alpha_1 F_B + n\alpha_2 F_B \theta_{B_1}\theta_{B_1}^H F_B^H\right)^{-1} = \frac{1}{n\alpha_1}F_B^{-1} - \frac{\alpha_2}{\lambda\alpha_1}\theta_{B_1}\theta_{B_1}^H$$

Indeed, since $\alpha_2 \neq -\alpha_1/(np)$, the matrices G_{11} and G_{22} are full-rank. Furthermore, we verify that $1 + \gamma_A^H G^{-1}\gamma_B = 0$, we can then apply [49, Theorem 6], which yields

$$\left(G + \gamma_{B_1}\gamma_{A_1}^H\right)^\dagger = G^{-1} - kk^\dagger G^{-1} - G^{-1}h^\dagger h + \kappa kh \quad (34)$$

with $k = \frac{1}{n\sqrt{\lambda}}\begin{bmatrix} -\theta_{A_1} \\ \theta_{B_1} \end{bmatrix}$, $h = \frac{1}{p\sqrt{\lambda}}\begin{bmatrix} \theta_{A_1} \\ -\theta_{B_1} \end{bmatrix}^T$, $\kappa = k^\dagger G^{-1}h^\dagger$

and where in the vector case $x^\dagger = \frac{x^H}{\|x\|^2}$ [49]. The intended result is obtained after simplification of $P\Gamma_1\left(G + \gamma_{B_1}\gamma_{A_1}^H\right)^\dagger \Gamma_1^H P^H$ with (34). ∎

Appendix B
Simplification for the CCRB of CES distributions

This appendix introduces a lemma, which is required for the proof of propositions 1 and 3.

Lemma 2. *Let α_1 and α_2 be two scalars such that $\alpha_1 > 0$ and $\alpha_2 > -\alpha_1/(np)$. For any admissible θ_1 satisfying the*

constraint (C) *and such that* $R_1 = R(\theta_1) = A_1 \otimes B_1$, *let us introduce* $\Gamma_1 = \begin{bmatrix} \theta_{B_1} \otimes I_{n_A} & I_{n_B} \otimes \theta_{A_1} \end{bmatrix}$ *and the matrix*

$$\mathbf{C} = \alpha_1 \mathbf{W}_{\mathbf{R}_1}^{-1} + \alpha_2 \text{vec}\left(\mathbf{R}_1^{-1}\right)\text{vec}\left(\mathbf{R}_1^{-1}\right)^H$$

We have the following equality

$$U\left(U^T\Gamma_1^H P^H C P \Gamma_1 U\right)^{-1} U^T = \begin{bmatrix} L_{11} & L_{12} \\ L_{12}^H & L_{22} \end{bmatrix} \text{ where}$$

$$\begin{cases} L_{11} = \dfrac{1}{p\alpha_1}\left(F_{A_1}^{-1} + \kappa \theta_{A_1}\theta_{A_1}^T - F_{A_1}^{-1}u_A\theta_{A_1}^T - \theta_{A_1}u_A^T F_{A_1}^{-1}\right) \\ L_{12} = -\dfrac{\kappa}{p\alpha_1}\left(\theta_{A_1}\theta_{B_1}^T - \dfrac{1}{\kappa}F_{A_1}^{-1}u_A\theta_{B_1}^T\right) \\ L_{22} = \dfrac{1}{n\alpha_1}F_{B_1}^{-1} + \left(\dfrac{-\alpha_2}{\alpha_1\lambda} + \dfrac{n\kappa-1}{np\alpha_1}\right)\theta_{B_1}\theta_{B_1}^T \end{cases}$$

with $\mathbf{u}_A = \dfrac{\partial c(\theta)}{\partial \theta_A} = \dfrac{1}{n}\mathbf{P}_A\text{vec}(\mathbf{I}_{n_A})$, $\kappa = u_A^T F_{A_1}^{-1} u_A$ *and* U *given in* (22).

Proof. Starting from the block decomposition of the matrix $\Gamma_1^H P^H C P \Gamma_1$ exhibited in the proof of Lemma 1, we obtain the following non singular matrix

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix} = \mathbf{U}^T\Gamma_1^H\mathbf{P}^H\mathbf{C}\mathbf{P}\Gamma_1\mathbf{U},$$

whose the inverse can be computed by the Schur complement

$$\mathbf{J}^{-1} = \begin{bmatrix} \mathcal{S}_J^{-1} & -\mathcal{S}_J^{-1}\mathbf{J}_{12}\mathbf{J}_{22}^{-1} \\ -\mathbf{J}_{22}^{-1}\mathbf{J}_{21}\mathcal{S}_J^{-1} & \mathbf{J}_{22}^{-1} + \mathbf{J}_{22}^{-1}\mathbf{J}_{21}\mathcal{S}_J^{-1}\mathbf{J}_{12}\mathbf{J}_{22}^{-1} \end{bmatrix}$$

with

$$\begin{cases} \mathbf{J}_{22}^{-1} = \mathbf{G}_{22}^{-1} = \dfrac{1}{n\alpha_1}\mathbf{F}_{B_1}^{-1} - \dfrac{\alpha_2}{\lambda\alpha_1}\theta_{B_1}\theta_{B_1}^H \\ \mathcal{S}_J \triangleq \mathbf{J}_{11} - \mathbf{J}_{12}\mathbf{J}_{22}^{-1}\mathbf{J}_{21} = p\alpha_1\mathbf{U}_A^T\left(\mathbf{F}_{A_1} - \dfrac{1}{n}\mathbf{F}_{A_1}\theta_A\theta_A^T\mathbf{F}_{A_1}^H\right)\mathbf{U}_A \end{cases}$$

By noting $\mathbf{Z}_A = \mathbf{U}_A^T\mathbf{F}_{A_1}\mathbf{U}_A$ and using the Sherman-Morrison formula, we obtain

$$\mathcal{S}_J^{-1} = \dfrac{1}{p\alpha_1}\left(\mathbf{Z}_A^{-1} + \kappa\mathbf{Z}_A^{-1}\mathbf{U}_A^T\mathbf{F}_{A_1}\theta_A\theta_A^T\mathbf{F}_{A_1}^H\mathbf{U}_A\mathbf{Z}_A^{-1}\right)$$

Finally, the proof is concluded by exploiting the following relation [35, Corollary 1]

$$\mathbf{U}_A\left(\mathbf{U}_A^T\mathbf{F}_{A_1}\mathbf{U}_A\right)^{-1}\mathbf{U}_A^T = \mathbf{F}_{A_1}^{-1} - \dfrac{1}{\kappa}\mathbf{F}_{A_1}^{-1}\mathbf{u}_A\mathbf{u}_A^T\mathbf{F}_{A_1}^{-1}.$$

∎

Appendix C
Simplification for the CCRB of CAE distribution

This appendix introduces a lemma, which is required for the proof of Theorem 4.

Lemma 3. *For any admissible* θ_1 *satisfying the constraint* (C1) *and such that* $R_1 = R(\theta_1) = A_1 \otimes B_1$, *let us introduce* $\Gamma_1 = \begin{bmatrix} \theta_{B_1} \otimes I_{n_A} & I_{n_B} \otimes \theta_{A_1} \end{bmatrix}$ *and the matrix*

$$\mathbf{C} = \mathbf{W}_{\mathbf{R}_1}^{-1} - \dfrac{1}{np}\text{vec}\left(\mathbf{R}_1^{-1}\right)\text{vec}\left(\mathbf{R}_1^{-1}\right)^H$$

We have the following equality

$$U_1\left(U_1^T\Gamma_1^H P^H C P \Gamma_1 U_1\right)^{-1} U_1^T = \begin{bmatrix} L_{11} & 0 \\ 0 & L_{22} \end{bmatrix} \text{ where}$$

$$\begin{cases} L_{11} = \dfrac{1}{p}\left(F_{A_1}^{-1} + \kappa\theta_{A_1}\theta_{A_1}^T - F_{A_1}^{-1}u_A\theta_{A_1}^T - \theta_{A_1}u_A^T F_{A_1}^{-1}\right) \\ L_{22} = \dfrac{1}{n}\left(F_{B_1}^{-1} + \kappa_1\theta_{B_1}\theta_{B_1}^T - F_{B_1}^{-1}u_B\theta_{B_1}^T - \theta_{B_1}u_B^T F_{B_1}^{-1}\right) \end{cases}$$

with $\mathbf{u}_A^T = \text{vec}(\mathbf{I}_n)^T\mathbf{P}_A/n$, $\mathbf{u}_B^T = \text{vec}(\mathbf{I}_p)^T\mathbf{P}_B/p$, $\kappa = u_A^T F_{A_1}^{-1}u_A$, $\kappa_1 = u_B^T F_{B_1}^{-1}u_B$ *and* U_1 *given in* (30).

Proof. Starting from the block decomposition of the matrix $\Gamma_1^H P^H C P \Gamma_1$ exhibited in the proof of Lemma 1 with $\lambda = 0$, we obtain the following non singular matrix

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & 0 \\ 0 & \mathbf{J}_{22} \end{bmatrix} = \mathbf{U}^T\Gamma_1^H\mathbf{P}^H\mathbf{C}\mathbf{P}\Gamma_1\mathbf{U},$$

whose the inverse can be computed by

$$\mathbf{J}^{-1} = \begin{bmatrix} \mathbf{J}_{11}^{-1} & 0 \\ 0 & \mathbf{J}_{22}^{-1} \end{bmatrix}$$

with

$$\begin{cases} \mathbf{J}_{11}^{-1} &= \dfrac{1}{p}\left(\mathbf{Z}_A - \dfrac{1}{n}\mathbf{U}_A^T\mathbf{F}_{A_1}\theta_{A_1}\theta_{A_1}^T\mathbf{F}_{A_1}^H\mathbf{U}_A\right)^{-1} \\ &= \dfrac{1}{p}\left(\mathbf{Z}_A^{-1} + \kappa\mathbf{Z}_A^{-1}\mathbf{U}_A^T\mathbf{F}_{A_1}\theta_{A_1}\theta_{A_1}^T\mathbf{F}_{A_1}^H\mathbf{U}_A\mathbf{Z}_A^{-1}\right) \\ \mathbf{J}_{22}^{-1} &= \dfrac{1}{n}\left(\mathbf{Z}_B - \dfrac{1}{p}\mathbf{U}_B^T\mathbf{F}_{B_1}\theta_{B_1}\theta_{B_1}^T\mathbf{F}_{B_1}^H\mathbf{U}_B\right)^{-1} \\ &= \dfrac{1}{n}\left(\mathbf{Z}_B^{-1} + \kappa_1\mathbf{Z}_B^{-1}\mathbf{U}_B^T\mathbf{F}_{B_1}\theta_{B_1}\theta_{B_1}^T\mathbf{F}_{B_1}^H\mathbf{U}_B\mathbf{Z}_B^{-1}\right) \end{cases}$$

where $\mathbf{Z}_A = \mathbf{U}_A^T\mathbf{F}_{A_1}\mathbf{U}_A$ and $\mathbf{Z}_B = \mathbf{U}_B^T\mathbf{F}_{B_1}\mathbf{U}_B$. Finally, the proof is concluded by exploiting the following relations [35, Corollary 1]

$$\begin{cases} \mathbf{U}_A\left(\mathbf{U}_A^T\mathbf{F}_{A_1}\mathbf{U}_A\right)^{-1}\mathbf{U}_A^T = \mathbf{F}_{A_1}^{-1} - \dfrac{1}{\kappa}\mathbf{F}_{A_1}^{-1}\mathbf{u}_A\mathbf{u}_A^T\mathbf{F}_{A_1}^{-1} \\ \mathbf{U}_B\left(\mathbf{U}_B^T\mathbf{F}_{B_1}\mathbf{U}_B\right)^{-1}\mathbf{U}_B^T = \mathbf{F}_{B_1}^{-1} - \dfrac{1}{\kappa_1}\mathbf{F}_{B_1}^{-1}\mathbf{u}_B\mathbf{u}_B^T\mathbf{F}_{B_1}^{-1} \end{cases}.$$

∎

References

[1] A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, *Array and Statistical Signal Processing*, 1st ed., ser. Academic Press Library in Signal Processing. Elsevier, 2014, vol. 3.
[2] N. Lu and D. L. Zimmerman, "The likelihood ratio test for a separable covariance matrix," *Statistics & probability letters*, vol. 73, no. 4, pp. 449–457, May 2004.
[3] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 655–665, 2004.

[4] J. C. De Munck, H. M. Huizenga, L. J. Waldorp, and R. Heethaar, "Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1565–1572, 2002.

[5] L. Beltrachini, N. von Ellenrieder, and C. H. Muravchik, "Shrinkage approach for spatiotemporal EEG covariance matrix estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1797–1808, 2013.

[6] P. Wirfält and M. Jansson, "On kronecker and linearly structured covariance matrix estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1536–1547, Mar. 2014.

[7] K. Werner, M. Jansson, and P. Stoica, "On estimation of covariance matrices with Kronecker product structure," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 478–491, Feb. 2008.

[8] I. Soloveychik and D. Trushin, "Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness," *Journal of Multivariate Analysis*, vol. 149, pp. 92–113, 2016.

[9] C. Leng and G. Pan, "Covariance estimation via sparse Kronecker structures," *Bernoulli*, vol. 24, no. 4B, pp. 3833–3863, 2018.

[10] G. Sun, Z. He, J. Tong, and X. Zhang, "Knowledge-aided covariance matrix estimation via Kronecker product expansions for airborne STAP," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 527–531, Apr. 2018.

[11] P. Stoica and T. Söderström, "On reparametrization of loss functions used in estimation and the invariance principle," *Elsevier Signal Processing*, vol. 17, no. 4, pp. 383–387, 1989.

[12] T. Tsiligkaridis, A. O. Hero, and S. Zhou, "On convergence of Kronecker graphical lasso algorithms," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 5347–5360, 2013.

[13] T. Tsiligkaridis and A. O. Hero, "Covariance estimation in high dimensions via Kronecker product expansions," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 1743–1755, 2013.

[14] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Elsevier Signal Processing*, vol. 82, no. 12, pp. 1847–1859, 2002.

[15] O. Besson and Y. I. Abramovich, "Invariance properties of the likelihood ratio for covariance matrix estimation in some complex elliptically contoured distributions," *Journal of Multivariate Analysis*, vol. 124, pp. 237–246, 2014.

[16] 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, Nov. 2016.

[17] H. Abeida and J.-P. Delmas, "Robustness of subspace-based algorithms with respect to the distribution of the noise: Application to DOA estimation," *Elsevier Signal Processing*, vol. 164, pp. 313–319, 2019.

[18] 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, Nov. 2012.

[19] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, no. 1, pp. 1–25, 1980.

[20] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2263–2278, May 2015.

[21] 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.

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

[23] B. Mériaux, C. Ren, A. Breloy, M. N. El Korso, P. Forster, and J.-P. Ovarlez, "On the recursions of Robust COMET algorithm for convexly structured shape matrix," in *Proc. of European Signal Processing Conference (EUSIPCO)*, 2019.

[24] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Transactions on Signal Processing*, vol. 14, no. 64, pp. 3576–3590, Jul. 2016.

[25] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, and D. P. Palomar, "Robust rank constrained kronecker covariance matrix estimation," in *Proc. of Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2016, pp. 810–814.

[26] Y. Wang, W. Xia, Z. He, H. Li, and A. P. Petropulu, "Polarimetric detection in compound Gaussian clutter with Kronecker structured covariance matrix," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4562–4576, 2017.

[27] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6182–6189, Dec. 2012.

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

[29] A. Mennad, S. Fortunati, M. N. El Korso, A. Younsi, A. M. Zoubir, and A. Renaux, "Slepian-Bangs-type formulas and the related misspecified Cramér-Rao bounds for complex elliptically symmetric distributions," *Elsevier Signal Processing*, vol. 142, pp. 320–329, Jul. 2017.

[30] J. R. Magnus and H. Neudecker, "The commutation matrix: some properties and applications," *The Annals of Statistics*, vol. 7, no. 2, pp. 381–394, Mar. 1979.

[31] A. Van den Bos, "The multivariate complex normal distribution - A generalization," *IEEE Transactions on Information Theory*, vol. 41, no. 2, pp. 537–539, Mar. 1995.

[32] J.-P. Delmas, "Performance bounds and statistical analysis of doa estimation," in *Array and statistical signal processing*, ser. Academic Press Library in Signal Processing. Elsevier, Dec. 2014, vol. 3, ch. 16, pp. 719–764.

[33] O. Besson and Y. I. Abramovich, "On the fisher information matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1130–1133, Nov. 2013.

[34] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.

[35] 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, Jul. 1998.

[36] S. Fortunati, F. Gini, and M. S. 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, May 2016.

[37] M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows *et al.*, "LOFAR: The low-frequency array," *Astronomy & astrophysics*, vol. 556, p. A2, 2013.

[38] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, "Robust calibration of radio interferometers in non-gaussian environment," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5649–5660, Nov. 2017.

[39] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The Annals of Statistics*, vol. 4, no. 1, pp. 51–67, Jan. 1976.

[40] C. F. Van Loan and N. Pitsianis, "Approximation with Kronecker products," in *Linear algebra for large scale and real-time applications*. Springer, 1993, pp. 293–314.

[41] A. W. Van der Vaart, *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, Jun. 2000, vol. 3.

[42] W. F. Trench, *Introduction to Real Analysis*. Prentice Hall, 2002.

[43] J.-P. Delmas and H. Abeida, "Survey and some new results on performance analysis of complex-valued parameter estimators," *Elsevier Signal Processing*, vol. 111, pp. 210–221, 2015.

[44] A. De Maio and M. Greco, *Modern radar detection theory*. The Institution of Engineering and Technology, 2016.

[45] I. Soloveychik and A. Wiesel, "Tyler's covariance matrix estimator in elliptical models with convex structure," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5251–5259, Oct. 2014.

[46] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, 2002.

[47] K. Werner and M. Jansson, "Estimating MIMO channel covariances from training data under the Kronecker model," *Elsevier Signal Processing*, vol. 89, no. 1, pp. 1–13, 2009.

[48] J. Ward, "Space-time adaptive processing for airborne radar," *IET Conference Proceedings*, pp. 2–2(1), Jan. 1998.

[49] C. D. Meyer, Jr, "Generalized inversion of modified matrices," *SIAM Journal on Applied Mathematics*, vol. 24, no. 3, pp. 315–323, 1973.