

Efficient Estimation of Kronecker Product of Linear Structured Scatter Matrices under t -distribution

Bruno Mériaux*, Chengfang Ren*, Arnaud Breloy[†], Mohammed Nabil El Korso[†] and Philippe Forster[‡]

*SONDRA, CentraleSupélec, Paris-Saclay University, F-91190 Gif-sur-Yvette, France

[†]LEME EA 4416, Paris-Nanterre University, F-92410 Ville d'Avray, France

[‡]SATIE, Paris-Nanterre University, F-94230 Cachan, France

Abstract—This paper addresses structured scatter matrix estimation within the non convex set of Kronecker product structure. The latter model usually involves two matrices, which can be themselves linearly constrained, and arises in many applications, such as MIMO communication, MEG/EEG data analysis. Taking this prior knowledge into account generally improves estimation accuracy. In the framework of robust estimation, the t -distribution is particularly suited to model heavy-tailed data. In this context, we introduce an estimator of the scatter matrix, having a Kronecker product structure and potential linear structured factors. In addition, we show that the proposed method yields a consistent and efficient estimate.

Index Terms—Structured scatter matrix, Kronecker product, t -distribution, M -estimators.

I. INTRODUCTION

Many adaptive statistical signal processing methods require the estimation of covariance matrices [1]. In addition to the Hermitian positive definite property, the latter usually own a refined structure depending on the considered application. For example, the Kronecker Product (KP) structure can arise for separable statistical models [2], or in various scenarios such as MIMO communication [3] and analysis of MEG/EEG data [4]. 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 the transmitter side [5]. Exploiting this prior knowledge in the estimation scheme leads to a better accuracy since the degree of freedom in the estimation problem decreases. The problem of covariance estimation with KP structure and potential linearly constrained factors has been investigated in the Gaussian framework [5]–[7], notably by using the Extended Invariance Principle (EXIP) [8]. However, the Gaussian case is not suited under heavy tailed observations, nor in the presence of outliers. A convenient extension occurs with the t -distribution, which offers an attractive flexibility thanks to its extra-parameter, referred to as the degree of freedom [9].

The work of B. Mériaux is partially funded by the Direction Générale de l'Armement/Agence Innovation Défense. This work is also supported by the ANR ASTRID referenced ANR-17-ASTR-0015.

In the context of robust and structured covariance estimation, several methods have been recently proposed [7], [10]–[12]. In [10], efficient robust estimators of convexly structured scatter matrices are studied. However, the latter are inadequate for the considered case, since the KP does not constitute a convex structure. In [7], the Maximum Likelihood (ML) estimator of KP structured matrices, from normalized data, is studied and can be easily obtained thanks to the geodesic convexity of the likelihood function [13]. Nevertheless, this result is not extendable for linear structured Kronecker factors. In [11], [12], estimators have been proposed minimizing Tyler's cost function under structure constraints with iterative Majorization-Minimization algorithms. For a KP of linearly constrained factors, it yields an imbrication of several convex problems, which can become computationally costly. Therefore in this paper, we propose a consistent and efficient estimator of the scatter matrix of t -distributed observations with a KP structure and possibly linear constrained Kronecker factors. The proposed method involves three steps, namely an unstructured ML-estimation of the scatter matrix, then an unweighted least squares problem and finally an appropriate weighted least squares regression. The last two stages have a closed-form solution. In addition, an extension of the proposed method empirically leads to better performance at low sample support.

In the following, convergence in distribution and in probability are, respectively, denoted by $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{\mathcal{P}}$. For a matrix \mathbf{X} , the operator \mathbf{X}^\dagger denotes the Moore-Penrose inverse. \mathbf{X}^T (respectively \mathbf{X}^H and \mathbf{X}^*) stands for the transpose (respectively conjugate transpose and conjugate) matrix. The vec-operator $\text{vec}(\mathbf{X})$, shortened by \mathbf{x} , stacks all columns of any \mathbf{X} into a vector. The operator \otimes refers to Kronecker matrix product. The matrix $\mathbf{K}_{k,\ell}$ denotes the commutation matrix, satisfying $\mathbf{K}_{k,\ell}\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}^T)$ for $\mathbf{X} \in \mathbb{C}^{k \times \ell}$ [14]. We introduce the matrix $\mathbf{W}_{\mathbf{X}}$ defined by $\mathbf{W}_{\mathbf{X}} = \mathbf{X}^T \otimes \mathbf{X}$ for any matrix \mathbf{X} .

This paper is organized as follows. In section (II), a brief review on t -distribution and the Fisher Information Matrix is presented. Section (III) focuses on the proposed estimator. The performance analysis is treated in Section

(IV). Some simulations results in Section (V) illustrate the theoretical analysis.

II. BACKGROUND AND PROBLEM SETUP

A. Background on the complex t -distribution

A random vector (r.v.), $\mathbf{y} \in \mathbb{C}^m$ follows a centered circular complex t -distribution with $d \in \mathbb{R}_+^*$ degrees of freedom, denoted by $\mathbf{y} \sim \mathcal{C}t_{m,d}(\mathbf{0}, \mathbf{R})$, if it has the following probability density function (p.d.f.) [15]:

$$p(\mathbf{y}; \mathbf{R}, d) = \frac{\Gamma(d+m)}{\pi^m d^m \Gamma(d)} \det(\mathbf{R})^{-1} g(\mathbf{y}^H \mathbf{R}^{-1} \mathbf{y}) \quad (1)$$

where the function $g(\cdot)$, called the density generator function, is given by $g(s) = (1+s/d)^{-(d+m)}$ and \mathbf{R} is the scatter matrix. This distribution has heavier tail than the Gaussian distribution, which is a limit case obtained when $d \rightarrow \infty$. It has finite 2nd-order moment for $d > 1$ and in this case the covariance matrix of the r.v. \mathbf{y} is given by $\mathbb{E}[\mathbf{y}\mathbf{y}^H] = \frac{d}{d-1} \mathbf{R}$, $d \neq 1$.

B. Fisher Information Matrix

The Fisher Information Matrix (FIM) plays a fundamental role in the analysis of statistical estimators. In [16], a closed-form of the FIM has been obtained for complex elliptical distributions. Notably for a centered circular complex t -distribution with scatter matrix \mathbf{R} parameterized by $\boldsymbol{\mu}$, it yields:

$$\mathbf{F}(\boldsymbol{\mu}) = \frac{\partial \mathbf{r}(\boldsymbol{\mu})^H}{\partial \boldsymbol{\mu}} \mathbf{Y} \frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \quad (2)$$

where $\mathbf{Y} = \sigma_1^{-1} \mathbf{W}_{\mathbf{R}}^{-1} + (\sigma_1^{-1} - 1) \text{vec}(\mathbf{R}^{-1}) \text{vec}(\mathbf{R}^{-1})^H$ in which $\sigma_1 = \frac{d+m+1}{d+m}$ and $\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$ refers to the Jacobian matrix of $\mathbf{r}(\boldsymbol{\mu}) = \text{vec}(\mathbf{R}(\boldsymbol{\mu}))$. We retrieve the Slepian-Bang's formula in the Gaussian case for $d \rightarrow \infty$. For regular and identifiable model, the FIM is not singular. Thus, the Mean Square Error (MSE) of any unbiased estimator can be lower bounded by the inverse of the FIM, which is then referred to as the Cramér-Rao bound (CRB).

C. Problem setup

Let us consider K i.i.d. zero mean t -distributed observations, $\mathbf{y}_k \sim \mathcal{C}t_{m,d}(\mathbf{0}, \mathbf{R}_e)$, $k = 1, \dots, K$, where d is assumed to be known. The invertible scatter matrix is assumed to be structured as a KP, i.e.,

$$\mathbf{R}_e = \mathbf{A}_e \otimes \mathbf{B}_e \quad (3)$$

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 at least 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

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

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 Toeplitz, banded, etc. We denote the concatenated parameter vector by $\boldsymbol{\theta} = [\boldsymbol{\theta}_A^T, \boldsymbol{\theta}_B^T]^T$, 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 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, since

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

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. $\boldsymbol{\theta}$ will be added when required.

III. PROPOSED ALGORITHM

In this section, we propose a three-step procedure to obtain an estimator of \mathbf{R}_e , taking into account the underlying KP structure and the potential refined structure of the Kronecker factors. For notational convenience, we omit the dependence on K for the estimators based on K observations when there is no ambiguity.

A. Step 1: unstructured ML-estimation of \mathbf{R}_e

From a set of K i.i.d. $\mathbf{y}_k \sim \mathcal{C}t_{m,d}(\mathbf{0}, \mathbf{R}_e)$, $k = 1, \dots, K$ with $K > m \triangleq np$, the unstructured ML estimator of the scatter matrix is given by the solution of the following fixed-point equation [15]:

$$\widehat{\mathbf{R}} = \frac{d+m}{K} \sum_{k=1}^K \frac{\mathbf{y}_k \mathbf{y}_k^H}{d + \mathbf{y}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}_k} \triangleq \mathcal{H}_K(\widehat{\mathbf{R}}) \quad (4)$$

Existence and uniqueness of the solution of (4) have been studied in [15]. 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 point [15]. Furthermore, the asymptotic behavior of $\widehat{\mathbf{R}}$ is well-known [15], specifically:

$$\widehat{\mathbf{R}} \xrightarrow{\mathcal{P}} \mathbf{R}_e \quad \text{and} \quad \sqrt{K} \text{vec}(\widehat{\mathbf{R}} - \mathbf{R}_e) \xrightarrow{\mathcal{L}} \mathcal{GCN}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma} \mathbf{K}_{m,m}) \quad (5)$$

with $\boldsymbol{\Sigma} = \sigma_1 \mathbf{W}_{\mathbf{R}_e} + \sigma_2 \mathbf{r}_e \mathbf{r}_e^H$ where $\sigma_2 = \frac{d+m+1}{d(d+m)}$ [17].

Remark 1. In the case of unknown degree of freedom of the t -distribution, step 1 could be replaced by a joint estimation of the unstructured scatter matrix and the degree of freedom [17].

B. Step 2: unweighted covariance fitting

For the second step, a KP structured estimate of \mathbf{R}_e , denoted $\widehat{\mathbf{R}}_{\text{KP}}$, is obtained by solving the standard unweighted covariance fitting:

$$(\widehat{\mathbf{A}}_1, \widehat{\mathbf{B}}_1) = \arg \min_{\mathbf{A}, \mathbf{B}} \|\widehat{\mathbf{R}} - \mathbf{A} \otimes \mathbf{B}\|_F^2 \Rightarrow \widehat{\mathbf{R}}_{\text{KP}} = \widehat{\mathbf{A}}_1 \otimes \widehat{\mathbf{B}}_1 \quad (6)$$

By introducing the following rearrangement of the matrix \mathbf{R} [18]:

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

where $\mathbf{R}^{k\ell}$ is (k, ℓ) th block of size $p \times p$ of \mathbf{R} and which satisfies the following properties

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

with \mathcal{P}_R a permutation matrix, the problem (6) can be rewritten as

$$(\widehat{\mathbf{A}}_1, \widehat{\mathbf{B}}_1) = \arg \min_{\mathbf{A}, \mathbf{B}} \left\| \mathcal{R}(\widehat{\mathbf{R}}) - \text{vec}(\mathbf{A}) \text{vec}(\mathbf{B})^T \right\|_F^2. \quad (7)$$

Note that (7) is a rank-one approximation problem, which can be efficiently solved using the Singular Value Decomposition (SVD) [18].

C. Step 3: adequate weighted covariance fitting

From the KP structured estimate $\widehat{\mathbf{R}}_{\text{KP}}$ given by (6), we introduce the weighted matrix $\widehat{\mathbf{Q}} = \widehat{\mathbf{R}}_{\text{KP}}^{-T} \otimes \widehat{\mathbf{R}}_{\text{KP}}^{-1}$ and we consider the minimization of the following weighted covariance fitting problem:

$$V(\boldsymbol{\theta}) = \left\| \widehat{\mathbf{Q}}^{-1/2} \text{vec}(\widehat{\mathbf{R}} - \mathbf{A}(\boldsymbol{\theta}_A) \otimes \mathbf{B}(\boldsymbol{\theta}_B)) \right\|_2^2 \quad (8)$$

As detailed in [6], the function (8) can be reformulated as a rank-one approximation problem

$$V_1(\boldsymbol{\theta}) = \left\| \mathbf{Q}_A^H \mathcal{R}(\check{\mathbf{R}}) \mathbf{Q}_B^* - \mathbf{T}_A \boldsymbol{\theta}_A \boldsymbol{\theta}_B^T \mathbf{T}_B^T \right\|_F^2 \quad (9)$$

where

$$\begin{cases} \check{\mathbf{R}} = \left(\widehat{\mathbf{A}}_1^{-1/2} \otimes \widehat{\mathbf{B}}_1^{-1/2} \right) \widehat{\mathbf{R}} \left(\widehat{\mathbf{A}}_1^{-1/2} \otimes \widehat{\mathbf{B}}_1^{-1/2} \right), \\ \mathbf{Q}_A \mathbf{T}_A = \left(\widehat{\mathbf{A}}_1^{-T/2} \otimes \widehat{\mathbf{A}}_1^{-1/2} \right) \mathbf{P}_A, \quad \mathbf{Q}_B \mathbf{T}_B = \left(\widehat{\mathbf{B}}_1^{-T/2} \otimes \widehat{\mathbf{B}}_1^{-1/2} \right) \mathbf{P}_B \end{cases}$$

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 orthogonal columns and \mathbf{T}_A and \mathbf{T}_B are invertible matrices. Again, (9) can be easily solved using a SVD. This yields the final structured estimate $\widehat{\mathbf{R}}_{\text{KP}r} = \mathbf{R}(\widehat{\boldsymbol{\theta}}) \triangleq \mathbf{A}(\widehat{\boldsymbol{\theta}}_A) \otimes \mathbf{B}(\widehat{\boldsymbol{\theta}}_B)$, where $\widehat{\boldsymbol{\theta}} = \left[\widehat{\boldsymbol{\theta}}_A^T, \widehat{\boldsymbol{\theta}}_B^T \right]^T$ is a minimizer of (8) or (9) equivalently.

Remark 2. It is worth noting that the minimized function (9) does not coincide with the EXIP approach, since the weighted matrix $\widehat{\mathbf{Q}}$ is not the inverse of the (asymptotic) covariance matrix of $\text{vec}(\widehat{\mathbf{R}} - \mathbf{R}_e)$ given by (5). Similarly, it is not a particular case of [10], since the KP is not a convex structure.

Remark 3. The double KP structure of $\widehat{\mathbf{Q}}$ is needed to have the equivalence between equations (8) and (9), having a closed-form solution. This weighted matrix $\widehat{\mathbf{Q}}$ is also necessary to yield the efficiency of the estimate after Step 3. In this context, Step 2 offers one possibility to easily obtain $\widehat{\mathbf{Q}}$ with the desired properties.

Remark 4. Step 3 can be iteratively repeated N_{it} times, by substituting in $\widehat{\mathbf{Q}}$ the estimate $\widehat{\mathbf{R}}_{\text{KP}}$ by $\widehat{\mathbf{R}}_{\text{KP}r}$ obtained at the previous iteration. We will see, in the next section, that the asymptotic behavior is identical but the performance at low sample support is empirically improved.

IV. ASYMPTOTIC ANALYSIS

This section provides a statistical analysis of the proposed estimator $\widehat{\mathbf{R}}_{\text{KP}r}$. We show that the latter is a consistent, asymptotic efficient and Gaussian distributed estimator of \mathbf{R}_e .

Theorem 1. Under the considered data model, the estimator $\widehat{\mathbf{R}}_{\text{KP}r}$ is consistent for \mathbf{R}_e .

Proof. First, the consistency of $\widehat{\mathbf{R}}$ leads straightforwardly to the one of $\widehat{\mathbf{R}}_{\text{KP}}$ for \mathbf{R}_e and thus $\widehat{\mathbf{Q}} \xrightarrow{\mathcal{P}} \mathbf{R}_e^{-T} \otimes \mathbf{R}_e^{-1}$, which is a full-rank matrix. Then, in a similar manner, we obtain $\widehat{\mathbf{R}}_{\text{KP}r} \xrightarrow{\mathcal{P}} \mathbf{R}_e$. ■

Remark 5. Note that the consistency of a minimizer of (9), denoted by $\widehat{\boldsymbol{\theta}}$, w.r.t. $\boldsymbol{\theta}_e$ is not ensured if no constraint, certifying the model identifiability, is imposed.

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

$$\begin{aligned} \sqrt{K} \text{vec}(\widehat{\mathbf{R}}_{\text{KP}r} - \mathbf{R}_e) &\stackrel{\mathcal{L}}{\rightarrow} \mathcal{GCN}(\mathbf{0}, \boldsymbol{\Xi}, \boldsymbol{\Xi} \mathbf{K}_{np, np}) \text{ with} \\ \boldsymbol{\Xi} &= \mathcal{P}_R \left(\frac{\sigma_1}{p} (\mathbf{b}_e \mathbf{b}_e^H) \otimes \mathbf{P}_A \mathbf{F}_A^{-1} \mathbf{P}_A^H + \frac{\sigma_1}{n} \mathbf{P}_B \mathbf{F}_B^{-1} \mathbf{P}_B^H \otimes (\mathbf{a}_e \mathbf{a}_e^H) \right. \\ &\quad \left. + \left(\sigma_2 - \frac{\sigma_1}{np} \right) (\mathbf{b}_e \mathbf{b}_e^H) \otimes (\mathbf{a}_e \mathbf{a}_e^H) \right) \mathcal{P}_R^H \end{aligned}$$

where $\mathbf{F}_A = \mathbf{P}_A^H \mathbf{W}_{A_e}^{-1} \mathbf{P}_A$ and $\mathbf{F}_B = \mathbf{P}_B^H \mathbf{W}_{B_e}^{-1} \mathbf{P}_B$.

Proof. Theorem 2 can be proved following the same methodology as in [6, Theorem 4]. Here, we describe the main steps and subtleties for the considered case. In order to apply the Delta method [19], by performing a Taylor expansion of the cost function in (8) around $\boldsymbol{\theta}_e$, a constraint has to be considered, to get rid off the scale ambiguity of the model, e.g. $[\boldsymbol{\theta}]_1 = 1$. Hence, Theorem 1 and the continuous mapping yield the consistency on $\widehat{\boldsymbol{\theta}}_K$ w.r.t. $\boldsymbol{\theta}_e$, with $\widehat{\boldsymbol{\theta}}_K$ a minimizer of (8). Then, we obtain similarly as in [6, Section VII] or [10, Section 4.2]

$$\sqrt{K} \mathbf{P} \boldsymbol{\Gamma}_e (\widehat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_e) = \mathbf{P} \boldsymbol{\Gamma}_e \mathbf{H}^+ (\boldsymbol{\xi}_K) \sqrt{K} \mathbf{g}_K(\boldsymbol{\theta}_e)$$

with $\mathbf{P} = \mathcal{P}_R (\mathbf{P}_B \otimes \mathbf{P}_A)$, $\boldsymbol{\Gamma}_e = \boldsymbol{\Gamma}(\boldsymbol{\theta}_e) = \begin{bmatrix} \boldsymbol{\theta}_{B_e} \otimes \mathbf{I}_{n_A} & \mathbf{I}_{n_B} \otimes \boldsymbol{\theta}_{A_e} \end{bmatrix}$ and

$$\begin{cases} \boldsymbol{\xi}_K \text{ such that } |\boldsymbol{\xi}_K - \boldsymbol{\theta}_e| \leq |\widehat{\boldsymbol{\theta}}_K - \boldsymbol{\theta}_e|, \text{ i.e., } \boldsymbol{\xi}_K \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_e \\ \mathbf{H}(\boldsymbol{\xi}_K) = \left. \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\xi}_K} \xrightarrow{\mathcal{P}} 2 \boldsymbol{\Gamma}_e^H \mathbf{P}^H (\mathbf{R}_e^{-T} \otimes \mathbf{R}_e^{-1}) \mathbf{P} \boldsymbol{\Gamma}_e \triangleq \mathbf{H}_\infty \\ \sqrt{K} \mathbf{g}_K(\boldsymbol{\theta}_e) = \sqrt{K} \left. \frac{\partial V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_e} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{R}_\infty) \\ \mathbf{R}_\infty = 4 \boldsymbol{\Gamma}_e^H \mathbf{P}^H \left[\sigma_1 \mathbf{W}_{R_e}^{-1} + \sigma_2 \text{vec}(\mathbf{R}_e^{-1}) \text{vec}(\mathbf{R}_e^{-1})^H \right] \mathbf{P} \boldsymbol{\Gamma}_e \end{cases}$$

The proof is concluded by the Delta-method derived for complex-valued parameters [20], which leads to

$$\sqrt{K}\text{vec}(\widehat{\mathbf{R}}_{\text{KPr}} - \mathbf{R}_e) \xrightarrow{\mathcal{L}} \mathcal{GCN}(\mathbf{0}, \mathbf{\Xi}, \mathbf{\Xi}\mathbf{K}_{np,np})$$

where $\mathbf{\Xi} = \mathbf{P}\mathbf{\Gamma}_e\mathbf{H}_\infty^\dagger\mathbf{R}_\infty\mathbf{H}_\infty^\dagger\mathbf{\Gamma}_e^H\mathbf{P}^H$, can be simplified as requested, by using Lemma 1 in Appendix with $\alpha_1 = 1$ and $\alpha_2 = 0$. ■

Remark 6. The exhibited multivariate non-circular complex Gaussian distribution is degenerate w.r.t the Lebesgue measure, since the matrix $\mathbf{\Xi}$ is not full rank.

Theorem 3. The estimator $\widehat{\mathbf{R}}_{\text{KPr}}$ is asymptotically efficient.

Proof. From (2), the FIM w.r.t. $\boldsymbol{\theta}$ is given by

$$\mathbf{F}(\boldsymbol{\theta}_e) = \mathbf{\Gamma}_e^H\mathbf{P}^H \left(\sigma_1^{-1}\mathbf{W}_{\mathbf{R}_e}^{-1} + (\sigma_1^{-1} - 1)\text{vec}(\mathbf{R}_e^{-1})\text{vec}(\mathbf{R}_e^{-1})^H \right) \mathbf{P}\mathbf{\Gamma}_e$$

The latter 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 $\text{vec}(\mathbf{R})$, we use the results of [21], which yields

$$\text{CRB}_r = \mathbf{P}\mathbf{\Gamma}_e\mathbf{F}^\dagger(\boldsymbol{\theta}_e)\mathbf{\Gamma}_e^H\mathbf{P}^H = \mathbf{\Xi} \quad (10)$$

where the last equality is obtained by using Lemma 1 in Appendix with $\alpha_1 = \sigma_1^{-1}$ and $\alpha_2 = \sigma_1^{-1} - 1$. The parameter vector $\text{vec}(\mathbf{R})$ being complex-valued, the intended result is easily achievable by working on the vector concatenating the real and imaginary parts of $\text{vec}(\mathbf{R})$ and noting that $\mathbf{\Xi}^T = \mathbf{K}_{np,np}\mathbf{\Xi}\mathbf{K}_{np,np}$. ■

Remark 7. In order to obtain the CRB for estimators of $\text{vec}(\mathbf{R})$, we could have used the framework of constrained CRB [22], yielding the same result of (10).

Corollary 1. The estimator $\widehat{\mathbf{R}}_{\text{KPr}}$ obtained after $N_{\text{it}} < \infty$ iterations of step 3, is consistent, asymptotically efficient and Gaussian distributed.

Proof. Theorems 1, 2 and 3 are valid at each iteration, which concludes the proof. ■

V. NUMERICAL RESULTS

In this section, we illustrate the results of the previous theoretical analysis for a KP structured scatter matrix, whose Kronecker factors are Hermitian Toeplitz, whose first rows are

$$\begin{cases} [\mathbf{A}_e]_{1,n} = [0.6324, -0.3369 - 0.2994i, -0.152 + 0.535i] \\ [\mathbf{B}_e]_{1,p} = [0.6478, -0.3152 - 0.1261i, -0.2961 + 0.0034i] \end{cases}$$

The minimal parameterization for the Toeplitz structure consists in stacking the real and imaginary parts of the first row of the matrix. We generate 5000 sets of K i.i.d. t -distributed samples $\mathbf{y}_k \sim \mathcal{C}t_{m,d}(\mathbf{0}, \mathbf{R}_e)$, $k = 1, \dots, K$ with $d = 3$ degrees of freedom and $n = p = 3$. We compare the performance of the proposed method with the estimator

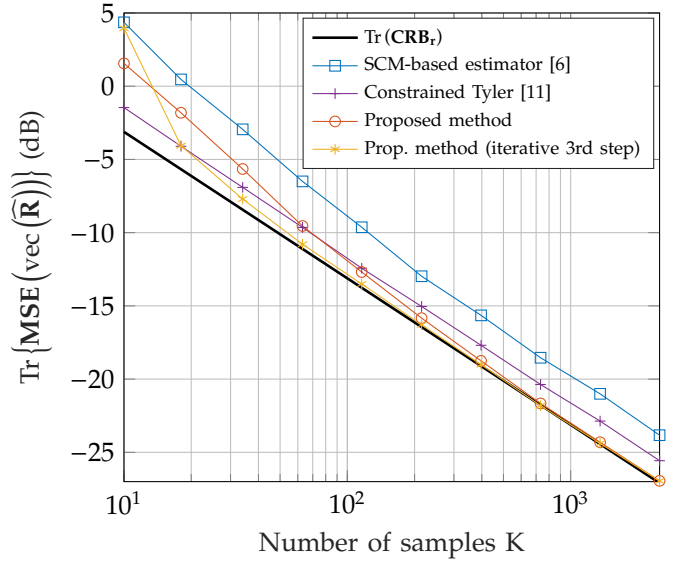


Figure 1. Efficiency simulation

of [6] based on the Sample Covariance Matrix (SCM) and the estimator of [11] obtained by minimizing the Tyler's cost function under structure constraints. Being also based on the SCM, the estimator derived in [5], not drawn for comparison, would have the same asymptotic performance as the one of [6].

From Fig. 1, we verify that the MSE of the proposed method asymptotically reaches the CRB, unlike the others estimators, which are not optimal for the t -distribution. We can also note that the proposed estimate, where step 3 has been iterated, converge faster than without iterations at low sample support. In addition, the asymptotic unbiasedness of the proposed method as well as this of the other algorithms can be indirectly noticed on the Fig. 1.

VI. CONCLUSION

In this paper, we addressed robust estimation of KP structured scatter matrices, with potential linear structured Kronecker factors. We proposed a consistent, asymptotically efficient and Gaussian distributed estimator for the t -distribution. Numerical results illustrate the theoretical analysis.

VII. APPENDIX

This appendix introduces a lemma, which is required for the proof of theorem 2 and 3.

Lemma 1. Let α_1 and α_2 be two scalars such that $\alpha_1 > 0$ and $\alpha_2 \neq -\alpha_1/(np)$. For any admissible $\boldsymbol{\theta}_1$ such

that $\mathbf{R}_1 = \mathbf{R}(\boldsymbol{\theta}_1) = \mathbf{A}_1 \otimes \mathbf{B}_1$, let us introduce $\boldsymbol{\Gamma}_1 = \begin{bmatrix} \boldsymbol{\theta}_{B_1} \otimes \mathbf{I}_{n_A} & \mathbf{I}_{n_B} \otimes \boldsymbol{\theta}_{A_1} \end{bmatrix}$. We have the following equality

$$\begin{aligned} & \mathbf{P}\boldsymbol{\Gamma}_1 \left(\boldsymbol{\Gamma}_1^H \mathbf{P}^H \left[\alpha_1 \mathbf{W}_{\mathbf{R}_1}^{-1} + \alpha_2 \text{vec}(\mathbf{R}_1^{-1}) \text{vec}(\mathbf{R}_1^{-1})^H \right] \mathbf{P}\boldsymbol{\Gamma}_1 \right)^\dagger \boldsymbol{\Gamma}_1^H \mathbf{P}^H = \\ & \mathcal{P}_{\mathbf{R}} \left(\frac{1}{p\alpha_1} (\mathbf{b}_1 \mathbf{b}_1^H) \otimes \mathbf{P}_A \mathbf{F}_A^{-1} \mathbf{P}_A^H + \frac{1}{n\alpha_1} \mathbf{P}_B \mathbf{F}_B^{-1} \mathbf{P}_B^H \otimes (\mathbf{a}_1 \mathbf{a}_1^H) \right. \\ & \quad \left. - \left(\frac{\alpha_1 + 2np\alpha_2}{np\alpha_1(\alpha_1 + np\alpha_2)} \right) (\mathbf{b}_1 \mathbf{b}_1^H) \otimes (\mathbf{a}_1 \mathbf{a}_1^H) \right) \mathcal{P}_{\mathbf{R}}^H \end{aligned}$$

Proof. From $\mathcal{P}_{\mathbf{R}}^H \mathbf{W}_{\mathbf{R}_1}^{-1} \mathcal{P}_{\mathbf{R}} = \mathbf{W}_{\mathbf{B}_1}^{-1} \otimes \mathbf{W}_{\mathbf{A}_1}^{-1}$, we can obtain

$$\boldsymbol{\Gamma}_1^H \mathbf{P}^H \left[\alpha_1 \mathbf{W}_{\mathbf{R}_1}^{-1} + \alpha_2 \text{vec}(\mathbf{R}_1^{-1}) \text{vec}(\mathbf{R}_1^{-1})^H \right] \mathbf{P}\boldsymbol{\Gamma}_1 = \mathbf{G} + \boldsymbol{\gamma}_B \boldsymbol{\gamma}_A^H$$

where

$$\mathbf{G} = \begin{bmatrix} p\alpha_1 \mathbf{F}_A + p^2 \alpha_2 \mathbf{F}_A \boldsymbol{\theta}_{A_1} \boldsymbol{\theta}_{A_1}^H \mathbf{F}_A^H & \lambda \mathbf{F}_A \boldsymbol{\theta}_{A_1} \boldsymbol{\theta}_{B_1}^H \mathbf{F}_B^H \\ \mathbf{0} & n\alpha_1 \mathbf{F}_B + n^2 \alpha_2 \mathbf{F}_B \boldsymbol{\theta}_{B_1} \boldsymbol{\theta}_{B_1}^H \end{bmatrix},$$

$$\boldsymbol{\gamma}_B = \begin{pmatrix} \mathbf{0} \\ \sqrt{\lambda} \mathbf{F}_B \boldsymbol{\theta}_{B_1} \end{pmatrix}, \boldsymbol{\gamma}_A = \begin{pmatrix} \sqrt{\lambda} \mathbf{F}_A \boldsymbol{\theta}_{A_1} \\ \mathbf{0} \end{pmatrix} \text{ and } \lambda = (\alpha_1 + np\alpha_2).$$

The matrix \mathbf{G} is invertible, such that

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

in which \mathbf{G}_{11}^{-1} and \mathbf{G}_{22}^{-1} are given by the Sherman-Morrison formula

$$\mathbf{G}_{11}^{-1} = \frac{1}{p} \left(\alpha_1 \mathbf{F}_A + p\alpha_2 \mathbf{F}_A \boldsymbol{\theta}_{A_1} \boldsymbol{\theta}_{A_1}^H \mathbf{F}_A^H \right)^{-1} = \frac{1}{p\alpha_1} \mathbf{F}_A^{-1} - \frac{\alpha_2}{\lambda \alpha_1} \boldsymbol{\theta}_{A_1} \boldsymbol{\theta}_{A_1}^H$$

$$\mathbf{G}_{22}^{-1} = \frac{1}{n} \left(\alpha_1 \mathbf{F}_B + n\alpha_2 \mathbf{F}_B \boldsymbol{\theta}_{B_1} \boldsymbol{\theta}_{B_1}^H \right)^{-1} = \frac{1}{n\alpha_1} \mathbf{F}_B^{-1} - \frac{\alpha_2}{\lambda \alpha_1} \boldsymbol{\theta}_{B_1} \boldsymbol{\theta}_{B_1}^H$$

Indeed, since $\alpha_2 \neq -\alpha_1/(np)$, the matrix

$\mathbf{G}_{11} = p\mathbf{P}_A^H \mathbf{W}_{A_1}^{-1/2} (\alpha_1 \mathbf{I}_{n^2} + p\alpha_2 \text{vec}(\mathbf{I}_n) \text{vec}(\mathbf{I}_n)^T) \mathbf{W}_{A_1}^{-1/2} \mathbf{P}_A$ is full-rank as well as \mathbf{G}_{22} .

Furthermore, we verify that $1 + \boldsymbol{\gamma}_A^H \mathbf{G}^{-1} \boldsymbol{\gamma}_B = 0$, we can then apply [23, Theorem 6], which yields

$$(\mathbf{G} + \boldsymbol{\gamma}_B \boldsymbol{\gamma}_A^H)^\dagger = \mathbf{G}^{-1} - \mathbf{k} \mathbf{k}^\dagger \mathbf{G}^{-1} - \mathbf{G}^{-1} \mathbf{h} \mathbf{h}^\dagger + \kappa \mathbf{h} \mathbf{h}^\dagger \quad (11)$$

with $\mathbf{k} = \frac{1}{n\sqrt{\lambda}} \begin{bmatrix} -\boldsymbol{\theta}_{A_1} \\ \boldsymbol{\theta}_{B_1} \end{bmatrix}$, $\mathbf{h} = \frac{1}{p\sqrt{\lambda}} \begin{bmatrix} \boldsymbol{\theta}_{A_1} \\ -\boldsymbol{\theta}_{B_1} \end{bmatrix}^T$, $\kappa = \mathbf{k}^\dagger \mathbf{G}^{-1} \mathbf{h}^\dagger$ and where the pseudo-inverse of a vector is given by $\mathbf{x}^\dagger = \frac{\mathbf{x}^H}{\|\mathbf{x}\|^2}$ [23]. The intended result is obtained after simplification of $\mathbf{P}\boldsymbol{\Gamma}_1 (\mathbf{G} + \boldsymbol{\gamma}_B \boldsymbol{\gamma}_A^H)^\dagger \boldsymbol{\Gamma}_1^H \mathbf{P}^H$ with (11). ■

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] 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.
- [6] 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.
- [7] I. Solovychik and D. Trushin, "Gaussian and robust Kronecker product covariance estimation: Existence and uniqueness," *Journal of Multivariate Analysis*, vol. 149, pp. 92–113, 2016.
- [8] 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.
- [9] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the *t*-distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6182–6189, Dec. 2012.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] A. W. Van der Vaart, *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, Jun. 2000, vol. 3.
- [20] 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.
- [21] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 87–90, 2001.
- [22] 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.
- [23] C. D. Meyer, Jr, "Generalized inversion of modified matrices," *SIAM Journal on Applied Mathematics*, vol. 24, no. 3, pp. 315–323, 1973.