EFFICIENT ESTIMATION OF SCATTER MATRIX WITH CONVEX STRUCTURE UNDER T-DISTRIBUTION

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

* SONDRA, CentraleSupélec, 91192 Gif-sur-Yvette, France
† Université Paris-Nanterre/LEME, 92410 Ville d'Avray, France
‡ SATIE, ENS Paris-Saclay, CNRS, 94230 Cachan, France

ABSTRACT

This paper addresses structured covariance matrix estimation under t-distribution. Covariance matrices frequently reveal a particular structure due to the considered application and taking into account this structure usually improves estimation accuracy. In the framework of robust estimation, the t-distribution is particularly suited to describe heavy-tailed observation. In this context, we propose an efficient estimation procedure for covariance matrices with convex structure under t-distribution. Numerical examples for Hermitian Toeplitz structure corroborate the theoretical analysis.

***Index Terms*—** Robust estimation, structured covariance matrix, convex structure, t-distribution, M-estimators.

1. INTRODUCTION

In adaptive signal processing, the estimation of the Covariance Matrix (CM) is a step at the center of most of the existing algorithms [1]. In addition to the Hermitian and positive properties, CM generally exhibits a specific structure due to the considered application. For example, using uniform linear arrays, CM reveals the Toeplitz structure [2]. For certain statistical models such as MIMO communications, CM can be expressed as a Kronecker product of two smaller dimension matrices, which could be themselves structured [3] (e.g., the reader is referred to [4] for further examples of structured CM). Taking into account this structure, leads to a better estimation accuracy, since the degree of freedom in the estimation problem decreases. This problem has been widely explored in the Gaussian framework [5]. However, the Gaussian case is not suited for heavy tailed observations. Conversely, the class of Complex Elliptically Symmetric [6, 7] (CES) includes most of usual non Gaussian distributions [8–11]. Notably, it includes the t-distribution, which gives a convenient extension of the normal distribution that can accurately model spiky radar clutter measurements [8, 10]. The extra-parameter of this distribution, called degree of freedom, provides a more flexible modeling with moderate increase in computational complexity [12]. In this context, the main contribution of this paper is to propose a consistent and efficient estimator for CM with convex structures under t-distributed data.

This paper is organized as follows. In section 2, we relate our contribution to prior work. In section 3, a brief review on t-distribution and the Fisher Information Matrix (FIM) is presented. Section 4 focuses on the proposed estimator. The performance analysis is treated in Section 5. Section 6 gives a particular application considering a Hermitian Toeplitz structure with simulation results.

In the following, the notation $\stackrel{d}{=}$ indicates "has the same distribution as". Convergence in distribution and in probability are, respectively, denoted by $\xrightarrow{d}$ and $\xrightarrow{\mathcal{P}}$. For a matrix $\mathbf{A}$, $|\mathbf{A}|$ and $\mathrm{Tr}\,(\mathbf{A})$ denote the determinant and the trace of $\mathbf{A}$. $\mathbf{A}^T$ (respectively $\mathbf{A}^H$) stands for the transpose (respectively conjugate transpose) matrix. The vec-operator $\mathrm{vec}(\mathbf{A})$ stacks all columns of $\mathbf{A}$ into a vector. The operator $\otimes$ refers to Kronecker matrix product and finally, the subscript "e" refers to the true value.

2. RELATION TO PRIOR WORK

In recent works, unstructured CM estimator under t-distribution with unknown degree of freedom has been studied in [13] by combining Maximum Likelihood (ML) and Method of Moments (MoM) approaches. However, the convergence of this latter iterative procedure is not guaranteed. Furthermore, robust estimation of CM with convex structures from normalized observations has been inspired by the unstructured distribution-free scatter matrix estimator proposed by Tyler [14]. Specifically, in [15], a COnvexly ConstrAined (COCA) CM estimator is proposed, leaning on the General MoM for the Tyler's estimator subject to convex constraints. This estimator is consistent but suffers from heavy computational cost. In [16,17], estimators have been proposed minimizing Tyler's cost function under structure constraints with iterative Majorization-Minimization algorithms. In [18], an efficient estimator for convex structured scatter matrix and normalized data is derived based on the COvariance Matching Estimation Technique (COMET) approach [5]. To the best of our knowledge, no estimator has been carried out for convex structured CM in a t-distribution context. In this paper, we fill this lack.

3. BACKGROUND AND PROBLEM SETUP

3.1. Background on the complex t-distribution

A m-dimensional zero mean random vector (r.v.), $\mathbf{y}_n \in \mathbb{C}^m$ follows a complex t-distribution with d degrees of freedom, denoted by $\mathbf{y}_n \sim \mathbb{C}t_{m,d}\,(\mathbf{0}, \mathbf{R})$, if it has the following probability density function (pdf) [7]:

$$p(\mathbf{y}_n; \mathbf{R}, d) = \frac{\Gamma(d+m)}{\pi^m d^m \Gamma(d)} |\mathbf{R}|^{-1} g\left(\mathbf{y}_n^H \mathbf{R}^{-1} \mathbf{y}_n\right) \qquad (1)$$

where $\mathbf{R}$ is the scatter matrix, d is a positive real number and the function $g(\cdot)$, called the density generator function, is given

Thanks to the Direction Générale de l'Armement (D.G.A) for its financial participation to this work. This work is also partially funded by the ANR ASTRID referenced ANR-17-ASTR-0015.

by $g(s) = (1+s/d)^{-(d+m)}$. Furthermore, the covariance matrix of the observations is related to the scatter matrix by $\text{Cov}(\mathbf{y}_n) = \mathbb{E}\left[\mathbf{y}_n\mathbf{y}_n^H\right] = \frac{d}{d-1}\mathbf{R}$, $d \neq 1$. This distribution has heavier tail than the Gaussian distribution. For example, the case $d = 0.5$ corresponds to the complex Cauchy distribution and the limit case $d \to \infty$ coincides with the Gaussian distribution. It has finite 2nd-order moment for $d > 1$. The 2nd-order modular variate, Q_n defined as $Q_n \overset{d}{=} \mathbf{y}_n^H\mathbf{R}^{-1}\mathbf{y}_n$ is a non-negative random variable, whose pdf is expressed by

$$p(q;\mathbf{R},d) = \frac{\Gamma(d+m)}{d^m\Gamma(d)\Gamma(m)}q^{m-1}g(q) \tag{2}$$

The t-distribution belongs to the subclass of Compound-Gaussian distributions [7, 19]. Indeed, a t-distributed r.v. with d degrees of freedom, $\mathbf{y}$ has the following stochastic representation

$$\mathbf{y} \overset{d}{=} \sqrt{\frac{2d}{x}}\mathbf{n}, \text{ with } x \sim \chi^2_{2d} \text{ and } \mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0},\mathbf{R}) \tag{3}$$

where χ^2_ν denotes the central chi-squared distribution with ν degrees of freedom.

3.2. Fisher Information Matrix

The Fisher Information Matrix (FIM) is a useful tool to study the ultimate performance of unbiased estimators. Specifically, the Cramér-Rao bound (CRB), which is the inverse of the FIM, is a lower bound on the mean squarre error. In the Gaussian framework, the FIM can be easily derived using the so-called Slepian-Bang formula [20, 21]. In [22], an extension of this formula has been derived for the CES distribution. Considering the particular case of the complex t-distribution with zero mean and a scatter matrix $\mathbf{R}$ parameterized by $\boldsymbol{\mu}$, the (k,ℓ) element of the FIM for a single vector of observation is given by [22]:

$$[\mathbf{F}]_{(k,\ell)} = \frac{(d+m)\text{Tr}\left(\mathbf{R}^{-1}\tilde{\mathbf{R}}_k\mathbf{R}^{-1}\tilde{\mathbf{R}}_\ell\right) - \text{Tr}\left(\mathbf{R}^{-1}\tilde{\mathbf{R}}_k\right)\text{Tr}\left(\mathbf{R}^{-1}\tilde{\mathbf{R}}_\ell\right)}{d+m+1} \tag{4}$$

in which $\tilde{\mathbf{R}}_k = \frac{\partial\mathbf{R}}{\partial\mu_k}$. As noticed previously, for $d \to \infty$, we retrieve the Slepian-Bang's formula in the Gaussian case.

3.3. Problem setup

Let us consider N i.i.d. zero mean t-distributed observations, $\mathbf{y}_n \sim \mathbb{C}t_{m,d}(\mathbf{0},\mathbf{R}_\text{e})$, $n = 1,\ldots,N$. We assume that the scatter matrix belongs to a convex subset $\mathscr{S}$ of Hermitian positive-definite matrices, and that there exists a one-to-one differentiable mapping $\boldsymbol{\mu} \mapsto \mathbf{R}(\boldsymbol{\mu})$ from $\mathbb{R}^P$ to $\mathscr{S}$. The vector $\boldsymbol{\mu}$ is the unknown parameter of interest with exact value $\boldsymbol{\mu}_\text{e}$, and $\mathbf{R}_\text{e} = \mathbf{R}(\boldsymbol{\mu}_\text{e})$ corresponds to the exact scatter matrix.

The negative log-likelihood function is given, up to an additive constant, by

$$\mathcal{L}(\mathbf{y}_1,..,\mathbf{y}_N;\boldsymbol{\mu}) = \frac{(d+m)}{N}\sum_{n=1}^N \log\left(1+\frac{\mathbf{y}_n^H\mathbf{R}(\boldsymbol{\mu})^{-1}\mathbf{y}_n}{d}\right) + \log|\mathbf{R}(\boldsymbol{\mu})| \tag{5}$$

The above function is non-convex w.r.t $\mathbf{R}$, its minimization w.r.t. $\boldsymbol{\mu}$ is therefore a difficult and time consuming problem. To overcome this issue, we propose in the next section a new estimation method that gives unique, consistent and asymptotically efficient estimates. Furthermore, for linear structures, we obtain closed form expressions of the estimates.

4. PROPOSED ALGORITHM

In this section, we propose a two step estimation procedure of $\boldsymbol{\mu}$. The first step consists in computing the unstructured ML-estimator of $\mathbf{R}$. The estimation of $\boldsymbol{\mu}$ is then obtained by solving a weighted least squares problem, derived from the so-called EXIP (EXtended Invariance Principle) approach [23]. For notational convenience, we omit the dependence on N for the estimators based on N observations when there is no ambiguity.

4.1. Step 1: unstructured ML-estimation of R

Let N i.i.d. observations, $\mathbf{y}_n \sim \mathbb{C}t_{m,d}(\mathbf{0},\mathbf{R})$ with $N > m$. The unstructured ML-estimator for the scatter matrix, denoted by $\widehat{\mathbf{R}}$, minimizing the negative log-likelihood (5) is the solution of the following fixed point equation:

$$\widehat{\mathbf{R}} = \frac{d+m}{N}\sum_{n=1}^N \frac{\mathbf{y}_n\mathbf{y}_n^H}{d+\mathbf{y}_n^H\widehat{\mathbf{R}}^{-1}\mathbf{y}_n} \triangleq \mathcal{H}_N(\widehat{\mathbf{R}}) \tag{6}$$

The reader can refer to [6, 24] for the study of existence and uniqueness related to (6), for which the iterative algorithm $\mathbf{R}_{k+1} = \mathcal{H}_N(\mathbf{R}_k)$ converges to $\widehat{\mathbf{R}}$ for any initialization point [7]. Moreover, the consistency and the asymptotic Gaussianity of this estimator have been proved in [7, Section VI]. Specifically, one has the following properties by [25]

$$\sqrt{N}\text{vec}\left(\widehat{\mathbf{R}}-\mathbf{R}\right) \xrightarrow{d} \mathbb{C}\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma},\boldsymbol{\Omega}) \tag{7}$$

with

$$\begin{cases} \boldsymbol{\Sigma} = \sigma_1\left(\mathbf{R}^T\otimes\mathbf{R}\right) + \sigma_2\text{vec}(\mathbf{R})\,\text{vec}(\mathbf{R})^H \\ \boldsymbol{\Omega} = \sigma_1\left(\mathbf{R}^T\otimes\mathbf{R}\right)\mathbf{K} + \sigma_2\text{vec}(\mathbf{R})\,\text{vec}(\mathbf{R})^T \end{cases}$$

where $\sigma_1 = \frac{d+m+1}{d+m}$ and $\sigma_2 = \frac{d+m+1}{d(d+m)}$ [13].

4.2. Step 2: Estimation of $\boldsymbol{\mu}$

For the second step, the estimator on $\boldsymbol{\mu}$ is obtained from the FIM and the unstructured ML-estimator for $\mathbf{R}$. It is determined by minimizing the following function, designed from the EXIP method [23]:

$$\begin{aligned} \widehat{\boldsymbol{\mu}} &= \arg\min_{\boldsymbol{\mu}}\ (\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}(\boldsymbol{\mu}))^T\,\mathbf{F}\,(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}(\boldsymbol{\mu})) \\ &= \arg\min_{\boldsymbol{\mu}}\ \sum_{k,\ell}(\widehat{\gamma}_k-\gamma_k(\boldsymbol{\mu}))\,[\mathbf{F}]_{(k,\ell)}\,(\widehat{\gamma}_\ell-\gamma_\ell(\boldsymbol{\mu})) \end{aligned}$$

where $\widehat{\gamma}_k$ denotes the k-th component of $\widehat{\boldsymbol{\gamma}} = \mathbf{J}\,\text{vec}\left(\widehat{\mathbf{R}}\right) = \mathbf{J}\,\widehat{\mathbf{r}}$, $\boldsymbol{\gamma}(\boldsymbol{\mu}) = \mathbf{J}\,\text{vec}(\mathbf{R}(\boldsymbol{\mu})) = \mathbf{J}\,\mathbf{r}(\boldsymbol{\mu})$ with $\mathbf{J}$ the matrix transforming the complex-valued vector $\mathbf{r}$ into a real-valued vector $\boldsymbol{\gamma}$ by the Hermitian property. The matrix $\mathbf{J}$ is invertible since the mapping between $\mathbf{r}$ and $\boldsymbol{\gamma}$ is one-to-one. In practice, the FIM is unknown, then we use its estimate, denoted by $\widehat{\mathbf{F}}$ and obtained from (4) by plugging an estimate of $\mathbf{R}$. Therefore, we obtain, by noting $\mathbf{R}_k = \frac{\partial\mathbf{R}}{\partial\gamma_k}$,

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}}\ (d+m)\text{Tr}\left(\widehat{\mathbf{R}}^{-1}\mathbf{Q}\widehat{\mathbf{R}}^{-1}\mathbf{Q}\right) - \text{Tr}\left(\widehat{\mathbf{R}}^{-1}\mathbf{Q}\right)\text{Tr}\left(\widehat{\mathbf{R}}^{-1}\mathbf{Q}\right)$$

with $\mathbf{Q} = \sum_k (\widehat{\gamma}_k - \gamma_k(\boldsymbol{\mu}))\,\mathbf{R}_k$. Furthermore, we remark that

$$\mathbf{Q} = \mathbf{R}_{\mathrm{e}} + \sum_k (\widehat{\gamma}_k - \gamma_{\mathrm{e}_k})\,\mathbf{R}_k - \left(\mathbf{R}_{\mathrm{e}} + \sum_k (\gamma_k(\boldsymbol{\mu}) - \gamma_{\mathrm{e}_k})\,\mathbf{R}_k\right)$$

In the above expression, we notice two first order Taylor expansions of $\widehat{\mathbf{R}}$, respectively $\mathbf{R}(\boldsymbol{\mu})$, around $\mathbf{R}_{\mathrm{e}}$ justified by the consistency of $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\mu}}$, whose the latter is provided by the EXIP method [23]. We finally achieve the following criterion

$$\begin{aligned}\widehat{\boldsymbol{\mu}} &= \arg\min_{\boldsymbol{\mu}} \mathcal{J}(\boldsymbol{\mu}) \quad \text{with} \\ \mathcal{J}(\boldsymbol{\mu}) &= (d+m)\mathrm{Tr}\left(\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\mu})\right)\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\mu})\right)\right) \\ &\quad - \left[\mathrm{Tr}\left(\widehat{\mathbf{R}}^{-1}\left(\widehat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\mu})\right)\right)\right]^2 \end{aligned} \tag{8}$$

Using the following relations $\mathrm{Tr}\left(\mathbf{A}^H\mathbf{B}\right) = \mathrm{vec}\,(\mathbf{A})^H\,\mathrm{vec}\,(\mathbf{B})$ and $\mathrm{vec}\,(\mathbf{AXB}) = \left(\mathbf{B}^T \otimes \mathbf{A}\right)\mathrm{vec}\,(\mathbf{X})$, we rewrite (8) as

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}}\ (\widehat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\mu}))^H\,\widehat{\mathbf{Y}}\,(\widehat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\mu})) = \left\|\widehat{\mathbf{Y}}^{1/2}\,(\widehat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\mu}))\right\|^2 \tag{9}$$

with $\widehat{\mathbf{Y}} = (d+m)\widehat{\mathbf{W}}^{-1} - \mathrm{vec}\left(\widehat{\mathbf{R}}^{-1}\right)\mathrm{vec}\left(\widehat{\mathbf{R}}^{-1}\right)^H$ and $\widehat{\mathbf{W}} = \widehat{\mathbf{R}}^T \otimes \widehat{\mathbf{R}}$. Given $\widehat{\mathbf{R}}$, the function $\mathcal{J}(\boldsymbol{\mu})$ is convex w.r.t $\mathbf{R}(\boldsymbol{\mu})$. Therefore, for $\mathbf{R} \in \mathscr{S}$ convex set, the minimization of (9) w.r.t. $\mathbf{R}(\boldsymbol{\mu})$ is a convex problem that admits a unique solution. In the following, we address the study of consistency and efficiency of $\widehat{\boldsymbol{\mu}}$ given as minimizer of (9).

5. ASYMPTOTIC PERFORMANCES

This section provides a statistical analysis of the proposed estimator $\widehat{\boldsymbol{\mu}}$, which is the unique solution minimizing the criterion (8) w.r.t $\boldsymbol{\mu} = (\mu_1, \dots, \mu_P)^T \in \mathbb{R}^P$ as already mentioned.

Theorem 1. The estimator $\widehat{\boldsymbol{\mu}}$, given by (8), is a consistent estimator of $\boldsymbol{\mu}_{\mathrm{e}}$. Likewise, $\mathbf{R}(\widehat{\boldsymbol{\mu}})$ is a consistent estimator of $\mathbf{R}(\boldsymbol{\mu}_{\mathrm{e}})$.

Proof. Using the consistency of $\widehat{\mathbf{R}}$ [7] and for large N, we obtain $\widehat{\mathbf{Y}} \xrightarrow{\mathcal{P}} \mathbf{Y}_{\mathrm{e}}$, $\widehat{\mathbf{r}} \xrightarrow{\mathcal{P}} \mathbf{r}_{\mathrm{e}}$. Consequently, (9) becomes

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}}\ \left\|\mathbf{Y}_{\mathrm{e}}^{1/2}\,(\mathbf{r}_{\mathrm{e}} - \mathbf{r}(\boldsymbol{\mu}))\right\|^2. \tag{10}$$

Since $\mathbf{Y}_{\mathrm{e}}^{1/2}$ is a full-rank matrix, the unique solution of the above problem satisfies $\mathbf{r}_{\mathrm{e}} = \mathbf{r}(\widehat{\boldsymbol{\mu}})$. Hence, under the assumption of a one-to-one mapping, the only solution is $\boldsymbol{\mu}_{\mathrm{e}}$, which establishes the consistency of $\widehat{\boldsymbol{\mu}}$. Finally, the continuous mapping implies $\mathbf{R}(\widehat{\boldsymbol{\mu}}) \xrightarrow{\mathcal{P}} \mathbf{R}(\boldsymbol{\mu}_{\mathrm{e}})$ ∎

Theorem 2. Let $\widehat{\boldsymbol{\mu}}_N$ be the estimator of $\boldsymbol{\mu}_{\mathrm{e}}$ defined by (8) and based on N i.i.d. observations, $\mathbf{y}_n \sim \mathbb{C}t_{m,d}\,(\mathbf{0}, \mathbf{R}(\boldsymbol{\mu}_{\mathrm{e}}))$. $\widehat{\boldsymbol{\mu}}_N$ is asymptotically unbiased, efficient and Gaussian distributed. Specifically,

$$\sqrt{N}\,(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}}) \xrightarrow{d} \mathcal{N}\,(\mathbf{0}, \mathbf{CRB}) \tag{11}$$

with $\mathbf{CRB} = \dfrac{d+m+1}{N}\left(\dfrac{\partial \mathbf{r}}{\partial \boldsymbol{\mu}}^H \mathbf{Y}_{\mathrm{e}} \dfrac{\partial \mathbf{r}}{\partial \boldsymbol{\mu}}\right)^{-1}$, in which we note $\mathbf{Y}_{\mathrm{e}} = (d+m)\mathbf{W}_{\mathrm{e}}^{-1} - \mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}^{-1}\right)\mathrm{vec}\left(\mathbf{R}_{\mathrm{e}}^{-1}\right)^H$ and $\dfrac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$ refers to the Jacobian matrix of $\mathbf{r}(\boldsymbol{\mu})$.

Proof. The estimate $\widehat{\boldsymbol{\mu}}_N$ is given by minimizing the function $\mathcal{J}(\boldsymbol{\mu})$. The consistency of $\widehat{\boldsymbol{\mu}}_N$ allows us to write the following Taylor expansion around $\boldsymbol{\mu}_{\mathrm{e}}$:

$$0 = \left.\frac{\partial \mathcal{J}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\widehat{\boldsymbol{\mu}}_N} = \left.\frac{\partial \mathcal{J}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}} + \left(\left.\frac{\partial^2 \mathcal{J}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}\partial \boldsymbol{\mu}^T}\right|_{\boldsymbol{\mu}=\boldsymbol{\xi}_N}\right)(\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}})$$

with $\boldsymbol{\xi}_N$ such as $|\boldsymbol{\xi}_N - \boldsymbol{\mu}_{\mathrm{e}}| \leq |\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}}|$, leading to

$$\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}} = -\left(\left.\frac{\partial^2 \mathcal{J}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}\partial \boldsymbol{\mu}^T}\right|_{\boldsymbol{\mu}=\boldsymbol{\xi}_N}\right)^{-1}\left.\frac{\partial \mathcal{J}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}} \quad \text{s.t. invertibility}$$

Therefore $\lim_{N\to\infty} \boldsymbol{\xi}_N = \boldsymbol{\mu}_{\mathrm{e}}$ by consistency of $\widehat{\boldsymbol{\mu}}_N$. According to [5, 26], we obtain for large N

$$\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}} \overset{d}{=} -\mathbf{H}(\boldsymbol{\mu}_{\mathrm{e}})^{-1}\mathbf{g}_N(\boldsymbol{\mu}_{\mathrm{e}})$$

with $\mathbf{g}_N(\boldsymbol{\mu}) = \dfrac{\partial \mathcal{J}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$ and $\mathbf{H}(\boldsymbol{\mu}) = \lim_{N\to\infty}\left(\dfrac{\partial^2 \mathcal{J}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}\partial \boldsymbol{\mu}^T}\right)$ invertible in the neighborhood of $\boldsymbol{\mu}_{\mathrm{e}}$. After some calculus, we obtain from (9)

$$\begin{aligned}\mathbf{g}_N(\boldsymbol{\mu}) &= -2\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^H \widehat{\mathbf{Y}}\,(\widehat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\mu})) \\ \mathbf{H}(\boldsymbol{\mu}_{\mathrm{e}}) &= 2\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}}^H \mathbf{Y}_{\mathrm{e}} \left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}}\end{aligned}$$

By consistency of $\widehat{\mathbf{Y}}$, we obtain for large N

$$\mathbf{g}_N(\boldsymbol{\mu}_{\mathrm{e}}) \simeq -2\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}}^H \mathbf{Y}_{\mathrm{e}}\,(\widehat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\mu}_{\mathrm{e}}))$$

Then,

$$\mathrm{Cov}\,(\widehat{\boldsymbol{\mu}}_N) \simeq \mathbf{H}(\boldsymbol{\mu}_{\mathrm{e}})^{-1}\mathbb{E}\left[\mathbf{g}_N(\boldsymbol{\mu}_{\mathrm{e}})\mathbf{g}_N(\boldsymbol{\mu}_{\mathrm{e}})^H\right]\mathbf{H}(\boldsymbol{\mu}_{\mathrm{e}})^{-1}$$

Using the asymptotic distribution of $\widehat{\mathbf{r}}$ given by (7), we finally obtain, in asymptotic regime,

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}}\right] &\simeq -\mathbb{E}\left[\mathbf{H}(\boldsymbol{\mu}_{\mathrm{e}})^{-1}\mathbf{g}_N(\boldsymbol{\mu}_{\mathrm{e}})\right] \\ &\simeq 2\mathbf{H}(\boldsymbol{\mu}_{\mathrm{e}})^{-1}\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}}^H \mathbf{Y}_{\mathrm{e}}\mathbb{E}\left[\widehat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\mu}_{\mathrm{e}})\right]\end{aligned}$$

$$\mathbb{E}\left[\mathbf{g}_N(\boldsymbol{\mu}_{\mathrm{e}})\mathbf{g}_N(\boldsymbol{\mu}_{\mathrm{e}})^H\right] \simeq \frac{4(d+m+1)}{N}\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}}^H \mathbf{Y}_{\mathrm{e}}\left.\frac{\partial \mathbf{r}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{e}}}$$

Hence, concerning the bias and the covariance on $\widehat{\boldsymbol{\mu}}_N$, we obtain asymptotically $\mathbb{E}\left[\widehat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_{\mathrm{e}}\right] \underset{N\to+\infty}{\longrightarrow} 0$ and

$$\mathrm{Cov}\,(\widehat{\boldsymbol{\mu}}_N)^{-1} = \frac{N}{d+m+1}\frac{\partial \mathbf{r}}{\partial \boldsymbol{\mu}}^H \mathbf{Y}_{\mathrm{e}}\frac{\partial \mathbf{r}}{\partial \boldsymbol{\mu}} = \mathbf{CRB}(\boldsymbol{\mu})^{-1}$$

where the CRB was obtained from the equation (4). It follows from the Delta method [27, Chapter 3] generalized for complex-valued estimators connected by a $\mathbb{C}$-differentiable function [28, 29] that $\sqrt{N}\mathrm{vec}\,(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\mathrm{e}}) \xrightarrow{d} \mathcal{N}\,(\mathbf{0}, \mathbf{CRB})$ ∎

6. NUMERICAL RESULTS

In this section, we illustrate the results of the previous statistical analysis for an Hermitian Toeplitz scatter matrix, which has indeed a convex structure. A natural parameterization for the Toeplitz structure is to take the real and imaginary parts of the first row in the matrix as interest parameter. For $m = 4$, the Toeplitz scatter matrix is generated from the Vandermonde matrix, $\mathbf{A}$ with $[\mathbf{A}]_{k,\ell} = e^{j2\pi(k-1)f_\ell}$, $f_\ell > 0$ and the positive diagonal matrix, $\mathbf{D}$, by $\mathbf{R} = \mathbf{ADA}^H$ and a trace equal to m. We generate 5000 sets of N independent m-dimensional t-distributed samples, $\mathbf{y}_n \sim \mathbb{C}t_{m,d}(\mathbf{0}, \mathbf{R}_e)$, $n = 1, \dots, N$ with $d = 5$ degrees of freedom, using the stochastic representation (3).

We compare the performance of the proposed algorithm to the state-of-the-art and the CRB. Furthermore, we display the performance of the proposed estimation scheme by replacing the first step by the joint-algorithm proposed in [13] (to deal with the possibility of unknown parameter d). Our algorithm is compared to RCOMET from [18] and COCA from [15]. Both methods are based on the Tyler's scatter estimator [14] using normalized observations $\mathbf{z}_n = \mathbf{y}_n/\|\mathbf{y}_n\|$. To the best of our knowledge, there is no other algorithm specifically derived for t-distributed observations dealing with structured scatter matrix. Finally, we compare to the intuitive estimate $\boldsymbol{\mu}$ obtained by averaging the real and imaginary parts of diagonals of the unstructured ML estimator (projection onto the Toeplitz set). The algorithms proposed in [16, 17] are not adapted for the considered case.

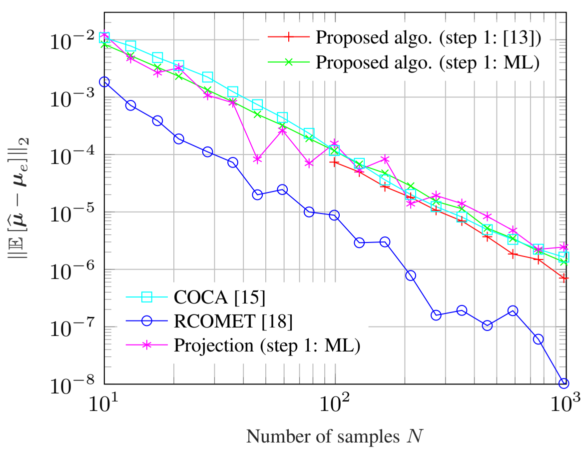

Fig. 1. Bias simulation

Fig. 1 presents the Euclidean norm of the estimated bias for $\widehat{\boldsymbol{\mu}}$ based on 5000 runs for each N. As shown previously, our algorithm with the unstructured ML estimator is asymptotically unbiased as well as the other algorithms. Performance of the proposed estimation scheme with the joint-algorithm as first step are not displayed for small N, since the joint-algorithm does not converge for part of the 5000 runs.

The asymptotic efficiency of our estimator is checked on Fig. 2: it reaches the CRB as N increases. RCOMET, and COCA do not reach this bound since they do not take into account the underlying distribution of the data. Despite the absence of convergence proof for the joint-algorithm, we remark that optimal asymptotic performances for $\boldsymbol{\mu}$ may be approached with unknown d.

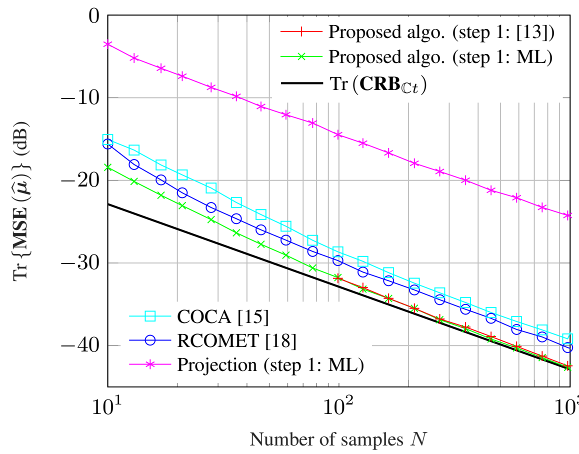

Fig. 2. Efficiency simulation

N	RCOMET [18]	Projection (step1: ML)	Proposed algo.(step1: ML)	Proposed algo.(step1: [13])	COCA [15]
100	0.012 s	0.017 s	0.018 s	0.46 s	2.18 s
500	0.048 s	0.086 s	0.086 s	1.94 s	15.1 s
1000	0.090 s	0.17 s	0.17 s	3.63 s	50.4 s

Table 1. Average calculation time

The Table 1 recaps the averaging calculation time of the different algorithms. As already mentioned, the COCA estimator suffers from heavy computational cost, since the number of constraints grows linearly in N. The estimation scheme with the joint-algorithm is slower than the one using the exact ML-estimator, which makes sense since the degree of freedom of the t-distribution is also estimated.

7. CONCLUSION

In this paper, we addressed structure covariance estimation for convex structures. We proposed a consistent, asymptotically unbiased and efficient estimator for t-distribution. Numerical simulations validate the theoretical analysis and show a performance gain compared to other algorithms, which normalize first the observations.

8. 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] M. Haardt, M. Pesavento, F. Röemer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," in *Array and Statistical Signal Processing*, ser. Academic Press Library in Signal Processing. Elsevier, Jan. 2014, vol. 3, ch. 15, pp. 651–717.

[3] 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.

[4] T. A. Barton and D. R. Fuhrmann, “Covariance structures for multidimensional data,” *Multidimensional Systems and Signal Processing*, vol. 4, no. 2, pp. 111–123, 1993.

[5] B. Ottersten, P. Stoica, and R. Roy, “Covariance matching estimation techniques for array signal processing applications,” *ELSEVIER Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

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

[7] 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.

[8] K. J. Sangston, F. Gini, and M. S. Greco, “Adaptive detection of radar targets in compound-Gaussian clutter,” in *Proc. of IEEE Radar Conference*, May, 2015, pp. 587–592.

[9] S. Watts, “Radar detection prediction in sea clutter using the compound K-distribution model,” *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 132, no. 7, pp. 613–620, Dec. 1985.

[10] K. D. Ward, C. J. Baker, and S. Watts, “Maritime surveillance radar. part 1: Radar scattering from the ocean surface,” *IEE Proceedings F (Communications, Radar and Signal Processing)*, vol. 137, no. 2, pp. 51–62, Apr. 1990.

[11] S. Kotz and S. Nadarajah, *Multivariate T-Distributions and Their Applications*. Cambridge University Press, 2004.

[12] 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.

[13] 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.

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

[15] 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.

[16] Y. Sun, P. Babu, and D. P. Palomar, “Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr., 2015, pp. 5693–5697.

[17] 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.

[18] B. Mériaux, C. Ren, M. N. El Korso, A. Breloy, and P. Forster, “Robust-COMET for covariance estimation in convex structures: algorithm and statistical properties,” in *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2017.

[19] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, and P. Larzabal, “Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[20] D. Slepian, “Estimation of signal parameters in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 3, no. 3, pp. 68–89, Mar. 1954.

[21] W. J. Bangs, “Array processing with generalized beamformers,” Ph.D. dissertation, Yale University, 1971.

[22] 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.

[23] 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.

[24] J. T. Kent and D. E. Tyler, “Redescending M-estimates of multivariate location and scatter,” *The Annals of Statistics*, vol. 19, no. 4, pp. 2102–2119, Dec. 1991.

[25] M. Mahot, F. Pascal, P. Forster, and J. P. Ovarlez, “Asymptotic properties of robust complex covariance matrix estimates,” *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3348–3356, Jul. 2013.

[26] L. Ljung, *System Identification : Theory for the User*, 2nd ed. Prentice Hall PTR, 1999.

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

[28] 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.

[29] 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.