New properties for Tyler's covariance matrix estimator

Gordana Draskovic, Frédéric Pascal

HAL Id: hal-01634293
https://hal.science/hal-01634293

Submitted on 29 Feb 2020

New properties for Tyler's covariance matrix estimator

Gordana Drašković
L2S, CentraleSupélec
University Paris-Saclay
3 rue Joliot Curie
91190 Gif-surYvette, France

Frédéric Pascal
L2S, CentraleSupélec
University Paris-Saclay
3 rue Joliot Curie
91190 Gif-surYvette, France

Abstract—In this paper, we deal with covariance matrix estimation in complex elliptically symmetric (CES) distributions. We focus on Tyler's estimator (TyE) and the well-known sample covariance matrix (SCM). TyE is widely used in practice, but its statistical behavior is still poorly understood. On the other hand, under Gaussian assumption, the SCM is Wishart-distributed, but its properties degrade in non-Gaussian environments. The main contribution is the derivation of new properties of TyE under CES framework, in order to approximate its behavior with a simpler one, the Wishart one. Finally, Monte-Carlo simulations support that claims and demonstrate the interest of this result.

Index Terms—Covariance matrix estimation, Complex Elliptically Symmetric distributions, Sample Covariance Matrix, Tyler's estimator, robust estimation

I. INTRODUCTION

Estimation of covariance matrix (CM) of high dimensional distribution is a basic problem in multivariate statistics [1], [2]. It arises in diverse applications such as filtering, whitening, detection, estimation or classification [3]–[7]. In practice, this matrix is unknown and needs to be estimated. Usually, in signal processing applications, one has an N-sample of identically distributed and independent (IID) complex vectors that are used to estimate this matrix. Under the Gaussian assumption, the classical covariance estimator is the well-known sample covariance matrix (SCM) $\widehat{\mathbf{M}}_{SCM}$ that is known to be distributed according to a Wishart distribution for fixed N [1]. Since the SCM is a maximum likelihood estimator (MLE), it exhibits good properties, even for small N. Moreover, it is easy to manipulate thanks to its explicit form. However, it also suffers from major drawbacks. This is case when the distribution of the training data turn to be non-Gaussian, or when the data are corrupted by some outliers. In such cases performances of the SCM may significantly degrade [8].

Robust estimation theory offers an alternative to the SCM thanks to so-called M-estimators, originally introduced by Huber [2] and investigated in the seminal work of Maronna [9]. They have been introduced within the framework of elliptical distributions. Elliptical distributions, originally introduced by Kelker [10], encompass a large number of well-known distributions such as the Gaussian distribution, the multivariate W-distribution or the mulivariate K-distribution. M-estimators of the covariance matrix are however seldom used in the signal processing community. One possible reason for this

is that their statistical properties are not well-known in the signal processing community. Their asymptotical properties have been studied by Tyler in the real case. This has been recently extended to the complex case, which is sometimes more useful for signal processing applications, in [7].

Among these estimators, the Tyler's estimator [11], also referred to as fixed-point (FP) estimator $\widehat{\mathbf{M}}_{FP}$ [3], [12], has been widely used as an alternative to the SCM for radar applications. Asymptotical properties of this estimator in Gaussian context have been analyzed through the bias, the consistency and the second-order moment in [12]. All these results are also valid in the framework of complex elliptically distributed data. Disappointingly, the statistical study of Tyler's estimator is quite problematic, mainly because of its nonexplicit form. With this in mind, we try to take a new look at this estimator. This paper sheds new light on the analysis of its behavior comparing it with the behavior of the SCM. Hence, the main purpose of this work is to provide the asymptotic distribution of the difference between Tyler's estimator in the case of elliptically distributed vectors and the SCM built with their Gaussian kernels, in order to justify the approximation of its behavior by a Wishart distribution.

This paper is organized as follows. Section II introduces some background about CES distributions, the SCM and Tyler's estimator. Then Section III presents the main contribution of the paper with discussions and Section IV shows Monte Carlo simulations validating the theoretical analysis. Finally, Section V concludes this work.

Vectors (resp. matrices) are denoted by bold-faced lower-case letters (resp. uppercase letters). *, ^{T} and ^{H} respectively represent the conjugate, the transpose and the Hermitian operator. N and $\mathbb{C}N$ denote the real and complex normal distributions. $\sim$ means "distributed as", $=_d$ stands for "shares the same distribution as", $\xrightarrow{d}$ denotes convergence in distribution and $\otimes$ denotes the Kronecker product. vec is the operator which transforms a matrix $m \times n$ into a vector of length mn, concatenating its n columns into a single column. Moreover, $\mathbf{I}_m$ is the $m \times m$ identity matrix, $\mathbf{0}$ the matrix of zeros with appropriate dimension and $\mathbf{K}$ is the commutation matrix which transforms $\text{vec}(\mathbf{A})$ into $\text{vec}(\mathbf{A}^T)$.

II. BACKGROUND

A. Complex Elliptically Symmetric distributions

Let be $\mathbf{z}$ an m-dimensional complex circular random vector. The vector has a complex elliptically symmetric (CES) distribution if its probability density function (PDF) can be written as

$$h_{\mathbf{z}}(\mathbf{z}) = |\mathbf{M}|^{-1} g_{\mathbf{z}} \left((\mathbf{z} - \boldsymbol{\mu})^H \mathbf{M}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right) \qquad (1)$$

where $h_{\mathbf{z}} : [0, \infty) \to [0, \infty)$ is any function such that (1) defines a PDF, $\boldsymbol{\mu}$ is the statistical mean and $\mathbf{M}$ is a scatter matrix. The matrix $\mathbf{M}$ reflects the structure of the covariance matrix of $\mathbf{z}$, i.e., the covariance matrix is equal to $\mathbf{M}$ up to a scale factor. This complex elliptically symmetric distribution will be denoted by $\mathbb{C}ES_m(\boldsymbol{\mu}, \mathbf{M}, g_{\mathbf{z}})$.

Stochastic Representation Theorem: A r.v. $\mathbf{z} \sim \mathbb{C}ES_m(\boldsymbol{\mu}, \mathbf{M}, g_{\mathbf{z}})$ with $\text{rank}(\mathbf{M}) = k \leq m$ if and only if it admits the stochastic representation

$$\mathbf{z} =_d \boldsymbol{\mu} + \tau \mathbf{A} \mathbf{u}, \qquad (2)$$

where the non-negative real random variable τ, called the modular variate, is independent of the complex random vector $\mathbf{u} \sim U(\mathbb{C}S^k)$ and $\mathbf{M} = \mathbf{A}\mathbf{A}^H$ is a factorization of $\mathbf{M}$, where $\mathbf{A} \in \mathbb{C}^{m \times k}$ has $\text{rank}(\mathbf{A}) = k$.

In this paper, we will assume that $\boldsymbol{\mu} = \mathbf{0}$. Without loss of generality, the scatter matrix will be taken to be equal to the covariance matrix when the latter exists. If the distribution of the data has a none finite second-order moment, then we will simply consider the scatter matrix estimator.

B. Complex Normal Distribution

The Gaussian distribution is a particular case of CES distributions in which $h_{\mathbf{z}}(z) = e^{-z/2}$. We shall write $\mathbf{z} \sim \mathbb{C}N_m(\mathbf{0}, \mathbf{M})$ to denote this case. In this case, the quadratic form $Q(\mathbf{z}) = \mathbf{z}^H \mathbf{M}^{-1} \mathbf{z}$ has a scaled chi-squared distribution $Q(\mathbf{z}) \sim (1/2)\chi_{2m}^2$, or equivalently $Q(\mathbf{z}) \sim \text{Gam}(m, 1)$ where $\mathbf{z} =_d \tau \mathbf{A} \mathbf{u}$.

C. Sample Covariance Matrix

The complex Wishart distribution $CW(N, \mathbf{M})$ is the distribution of $\sum_{k=1}^{N} \mathbf{x}_k \mathbf{x}_k^H$, when $\mathbf{x}_k$ are complex circular i.i.d. Gaussian vectors with zero mean and covariance matrix $\mathbf{M}$. Let

$$\widehat{\mathbf{M}}_{SCM} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}_k \mathbf{x}_k^H \qquad (3)$$

be the related SCM which will be also referred to as a Wishart matrix. Its asymptotic distribution is given by

$$\sqrt{N}\text{vec}(\widehat{\mathbf{M}}_{SCM} - \mathbf{M}) \xrightarrow{d} \mathbb{C}N(\mathbf{0}, \mathbf{M}^T \otimes \mathbf{M}, (\mathbf{M}^T \otimes \mathbf{M})\mathbf{K}). \quad (4)$$

D. Tyler's estimator

Let $(\mathbf{z}_1, ..., \mathbf{z}_N)$ be an N-sample of m-dimensional complex independent vectors with $\mathbf{z}_k \sim \mathbb{C}ES_m(\mathbf{0}, \mathbf{M}, g_{\mathbf{z}})$. An M-estimator, denoted by $\widehat{\mathbf{M}}$, is defined by the solution of the following M-estimating equation

$$\widehat{\mathbf{M}} = \frac{1}{N} \sum_{k=1}^{N} \varphi(\mathbf{z}_k^H \widehat{\mathbf{M}}^{-1} \mathbf{z}_k) \mathbf{z}_k \mathbf{z}_k^H \qquad (5)$$

where φ is any real-valued weight function on $[0, \infty)$ that does not need to be related to the PDF of the underlying CES distribution. Thus M-estimators constitute a wide class of scatter matrix estimators. Existence and uniqueness of the solution of (5) as well as the convergence of the corresponding recursive algorithm have already been shown, provided the function φ satisfies a set of general assumptions [13]. In particular, the resulting estimators are consistent estimators of the CM (up to a scale factor) [13].

For the particular function $\varphi(s) = m/s$, Tyler's estimator $\widehat{\mathbf{M}}_{FP}$ is the solution (up to a scale factor) of the following equation

$$\widehat{\mathbf{M}}_{FP} = \frac{m}{N} \sum_{k=1}^{N} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \widehat{\mathbf{M}}_{FP}^{-1} \mathbf{z}_k}. \qquad (6)$$

Moreover, for Tyler's estimator $\widehat{\mathbf{M}}_{FP}$ verifying (6), one has (see [12] for more details)

$$\sqrt{N}\text{vec}(\widehat{\mathbf{M}}_{FP} - \mathbf{M}) \xrightarrow{d} \mathbb{C}N(\mathbf{0}, \boldsymbol{\Sigma}_{FP}, \boldsymbol{\Omega}_{FP}) \qquad (7)$$

where $\boldsymbol{\Sigma}_{FP}$ and $\boldsymbol{\Omega}_{FP}$ are defined by

$$\boldsymbol{\Sigma}_{FP} = \frac{m+1}{m} \mathbf{M}^T \otimes \mathbf{M} - \frac{m+1}{m^2} \text{vec}(\mathbf{M})\text{vec}(\mathbf{M})^H,$$
$$\boldsymbol{\Omega}_{FP} = \frac{m+1}{m} (\mathbf{M}^T \otimes \mathbf{M})\mathbf{K} - \frac{m+1}{m^2} \text{vec}(\mathbf{M})\text{vec}(\mathbf{M})^T \qquad (8)$$

This shows that Tyler estimator behaves asymptotically as the SCM, the difference coming from the values of the scale quantities.

III. MAIN CONTRIBUTION

This section is devoted to the main contribution of this paper. We present the main result for the asymptotic distribution of the difference between Tyler's estimator and the SCM. Let $(\mathbf{u}_1, ..., \mathbf{u}_N)$ be an N-sample of m-dimensional complex independent vectors with $\mathbf{u}_n \sim U(\mathbb{C}S^m)$, $\nu_1, ..., \nu_N$ an N-sample of real independent variables with $|\nu_n|^2 \sim (1/2)\chi_{2m}^2$ and $\tau_1, ..., \tau_N$ an N-sample of real independent variables , $n = 1, ..., N$. Let now $\mathbf{x}_n = \nu_n \mathbf{A} \mathbf{u}_n \sim \mathbb{C}N_m(\mathbf{0}, \mathbf{M})$ and $\mathbf{z}_n = \tau_n \mathbf{A} \mathbf{u}_n \sim \mathbb{C}ES_m(\mathbf{0}, \mathbf{M})$ where $\mathbf{M} = \mathbf{A}\mathbf{A}^H$ is a factorization of $\mathbf{M}$. Consider the SCM $\widehat{\mathbf{M}}_{SCM}$ built with $(\mathbf{x}_1, ..., \mathbf{x}_N)$ and the FP estimator $\widehat{\mathbf{M}}_{FP}$ built with $(\mathbf{z}_1, ..., \mathbf{z}_N)$.

Theorem III.1 *The asymptotic distribution of* $\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM}$ *is given by*

$$\sqrt{N}vec(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM}) \xrightarrow{d} \mathbb{C}N(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) \qquad (9)$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are defined by

$$\boldsymbol{\Sigma} = \frac{1}{m}\boldsymbol{M}^T \otimes \boldsymbol{M} + \frac{m-1}{m^2}vec(\boldsymbol{M})vec(\boldsymbol{M})^H,$$
$$\boldsymbol{\Omega} = \frac{1}{m}(\boldsymbol{M}^T \otimes \boldsymbol{M})\boldsymbol{K} + \frac{m-1}{m^2}vec(\boldsymbol{M})vec(\boldsymbol{M})^T. \tag{10}$$

Remark III.1
- *Elements of matrices in* (10) *are smaller than those in* (8) *which reveals that Tyler's estimator asymptotically approaches the SCM much faster than it approaches the real covariance matrix. That justifies the intention of approximating its behavior by the one of the SCM.*
- *It is also crucial to note that the scale factors in* (10) *tend to 0 when the size m increases. Thus, for high dimensional observations, this approximation is even more accurate (since $\boldsymbol{\Sigma} \ll \boldsymbol{\Sigma}_{FP}$ and $\boldsymbol{\Omega} \ll \boldsymbol{\Omega}_{FP}$). This is in agreement with the results obtained in [14] under the assumptions that $N, m \to \infty$ and $N/m \to c > 0$.*

Proof: To prove the statement let us rewrite the right hand side of equation (10) as follows:

$$\sqrt{N}(vec(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM})) =$$
$$\sqrt{N}(vec(\widehat{\mathbf{M}}_{FP} - \mathbf{M} - \widehat{\mathbf{M}}_{SCM} + \mathbf{M}))$$
$$= [1, -1] \begin{bmatrix} \sqrt{N}vec(\widehat{\mathbf{M}}_{FP} - \mathbf{M}) \\ \sqrt{N}vec(\widehat{\mathbf{M}}_{SCM} - \mathbf{M}) \end{bmatrix} \tag{11}$$

Therefore

$$
\begin{aligned}
\boldsymbol{\Sigma}^{(N)} &= NE[vec(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM})vec(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM})^H] \\
&= [1, -1] E\left[\begin{bmatrix} \sqrt{N}vec(\widehat{\mathbf{M}}_{FP} - \mathbf{M}) \\ \sqrt{N}vec(\widehat{\mathbf{M}}_{SCM} - \mathbf{M}) \end{bmatrix} \right. \\
&\quad \times \left. \begin{bmatrix} \sqrt{N}vec(\widehat{\mathbf{M}}_{FP} - \mathbf{M}) \\ \sqrt{N}vec(\widehat{\mathbf{M}}_{SCM} - \mathbf{M}) \end{bmatrix}^H \right] \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
&= \boldsymbol{\Sigma}_1^{(N)} - 2\boldsymbol{\Sigma}_2^{(N)} + \boldsymbol{\Sigma}_3^{(N)} \tag{12}
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{\Sigma}_1^{(N)} &= NE[vec(\widehat{\mathbf{M}}_{FP} - \mathbf{M})vec(\widehat{\mathbf{M}}_{FP} - \mathbf{M})^H] \\
\boldsymbol{\Sigma}_2^{(N)} &= NE[vec(\widehat{\mathbf{M}}_{FP} - \mathbf{M})vec(\widehat{\mathbf{M}}_{SCM} - \mathbf{M})^H] \\
\boldsymbol{\Sigma}_3^{(N)} &= NE[vec(\widehat{\mathbf{M}}_{SCM} - \mathbf{M})vec(\widehat{\mathbf{M}}_{SCM} - \mathbf{M})^H]
\end{aligned}
\tag{13}
$$

where $\boldsymbol{\Sigma}_1^{(N)} \xrightarrow[N\to+\infty]{} \boldsymbol{\Sigma}_1$ given by (8) and $\boldsymbol{\Sigma}_3^{(N)} \xrightarrow[N\to+\infty]{} \boldsymbol{\Sigma}_3$ given by (4). Let us now introduce some notations (see [12] for details)

- $\widehat{\mathbf{M}} = \mathbf{M} + \delta\mathbf{M}$ where $\widehat{\mathbf{M}}$ denote an estimate of $\mathbf{M}$,
- $\boldsymbol{\Delta} = \mathbf{M}^{-1/2}\widehat{\mathbf{M}}\mathbf{M}^{-1/2} - \mathbf{I}$
- $\boldsymbol{\delta} = vec(\boldsymbol{\Delta})$
- TFP (theoretical FP) estimator of $\mathbf{M}$ is equal to

$$\widehat{\mathbf{M}}_{TFP} = \frac{m}{N} \sum_{k=1}^{N} \frac{\mathbf{z}_k\mathbf{z}_k^H}{\mathbf{z}_k^H\mathbf{M}^{-1}\mathbf{z}_k}. \tag{14}$$

In the sequel, these quantities will be indexed according to the studied estimator: SCM, TFP and FP. By using $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$ (see [15]), we have

$$
\begin{aligned}
\boldsymbol{\Sigma}_2^{(N)} &= NE[vec(\mathbf{M}^{1/2}\boldsymbol{\Delta}_{FP}\mathbf{M}^{1/2})vec(\mathbf{M}^{1/2}\boldsymbol{\Delta}_{SCM}\mathbf{M}^{1/2})^H] \\
&= (\mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2})NE[\boldsymbol{\delta}_{FP}\boldsymbol{\delta}_{SCM}^H](\mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2})^H
\end{aligned}
$$

To derive $NE[\boldsymbol{\delta}_{FP}\boldsymbol{\delta}_{SCM}^H]$ we can use the statement that $\sqrt{N}\boldsymbol{\delta}_{FP}$ and $\frac{m+1}{m}\sqrt{N}\boldsymbol{\delta}_{TFP}$ share the same asymptotic distribution for all CES distributions (see [12]). Now, it remains to derive the quantity $NE[\boldsymbol{\delta}_{TFP}\boldsymbol{\delta}_{SCM}^H]$.

Let $\mathbf{v}_k = \mathbf{M}^{-1/2}\mathbf{z}_k$ and $\mathbf{w}_k = \mathbf{M}^{-1/2}\mathbf{x}_k$. Then, for large N, one can write

$$
\begin{aligned}
NE[\boldsymbol{\delta}_{TFP}\boldsymbol{\delta}_{SCM}^H] &= NE\left[vec\left(\frac{m}{N} \sum_{k=1}^{N} \left(\frac{\mathbf{v}_k\mathbf{v}_k^H}{\mathbf{v}_k^H\mathbf{v}_k}\right) - \mathbf{I} \right) \right. \\
&\quad \left. vec\left(\frac{1}{N} \sum_{k=1}^{N} \left(\mathbf{w}_k\mathbf{w}_k^H\right) - \mathbf{I} \right)^H \right] \\
&= \frac{m}{N}E\left[vec\left(\sum_{k=1}^{N} \left(\frac{\mathbf{v}_k\mathbf{v}_k^H}{\mathbf{v}_k^H\mathbf{v}_k}\right) \right) vec\left(\sum_{k=1}^{N} \left(\mathbf{w}_k\mathbf{w}_k^H\right) \right)^H \right] \\
&\quad -Nvec(\mathbf{I})vec(\mathbf{I})^T \\
&= \frac{m}{N}\left(\sum_{k=1}^{N} E\left[vec\left(\frac{\mathbf{v}_k\mathbf{v}_k^H}{\mathbf{v}_k^H\mathbf{v}_k}\right)vec\left(\mathbf{w}_k\mathbf{w}_k^H\right)^H \right] \right. \\
&\quad \left. + \sum_{j\neq i} E\left[vec\left(\frac{\mathbf{v}_i\mathbf{v}_i^H}{\mathbf{v}_i^H\mathbf{v}_i}\right)vec\left(\mathbf{w}_j\mathbf{w}_j^H\right)^H \right] \right) \\
&\quad -Nvec(\mathbf{I})vec(\mathbf{I})^T \\
&= mE\left[vec\left(\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)vec\left(\mathbf{w}\mathbf{w}^H\right)^H \right] \\
&\quad +(N-1)E\left[vec\left(m\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right) \right]E\left[vec\left(\mathbf{w}\mathbf{w}^H\right)^H \right] \\
&\quad -Nvec(\mathbf{I})vec(\mathbf{I})^T \\
&= mE\left[vec\left(\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)vec\left(\mathbf{w}\mathbf{w}^H\right)^H \right] - vec(\mathbf{I})vec(\mathbf{I})^T
\end{aligned}
$$

where $\mathbf{v}_n = \tau_n\mathbf{u}_n \sim \mathbb{C}ES_m(\mathbf{0},\mathbf{I})$ and $\mathbf{w}_n = \nu_n\mathbf{u}_n \sim \mathbb{C}N_m(\mathbf{0},\mathbf{I})$ with τ_n being a r.va. whose PDF is unkown, $|\nu_n|^2 \sim (1/2)\chi_{2m}^2$ and $\mathbf{u}_n \sim U(\mathbb{C}S^m)$. Then, focusing on the following variable:

$$\mathbf{P} = E\left[vec\left(\frac{\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\right)vec\left(\mathbf{w}\mathbf{w}^H\right)^H \right] \tag{15}$$

each element of matrix $\mathbf{P}$ becomes

$$P_{kl} = E\left[u_p u_q^* u_{p'}^* u_{q'} |\nu|^2 \right] \tag{16}$$

with $k = p + m(q-1)$ and $l = p' + m(q'-1)$.

Now, let us define the 4th-order moments of a complex random vector $\mathbf{u}$ by

$$\alpha_{i_1,i_2;j_1,j_2} = E[u_{i_1}u_{i_2}u_{j_1}^*u_{j_1}^*] \tag{17}$$

By the circular symmetry properties all even-order central moments vanish in the cases that the sets (i_1, i_2) and (j_1, j_2) differ. In our case, that means that $P_{kl} = 0$ except for the following indices:

1) $k = p + m(p-1)$
2) $k = p + m(p-1)$, $l = p' + m(p'-1)$ and $p \neq p'$
3) $k = p + m(p'-1)$, $l = p + m(p'-1)$ and $p \neq p'$.
Now, since $\mathbf{u} \sim U(\mathbb{C}S^m)$ one has $|u_i|^2 \sim Beta(1, m-1)$, $E[|u_i|^4] = 2/m(m+1)$ and $E[|u_i|^2|u_j|^2] = 1/m(m+1)$. Also, as $|\nu|^2 \sim (1/2)\chi^2_{2m}$ one obtains the following results:
1) $P_{p+m(p-1),p+m(p-1)} = 2/(m+1)$;
2) $P_{p+m(p-1),p'+m(p'-1)} = 1/(m+1)$;
3) $P_{p+m(q-1),p+m(q-1)} = 1/(m+1)$;
and thus

$$NE[\boldsymbol{\delta}_{TFP}\boldsymbol{\delta}_{SCM}^H] \xrightarrow[N\to+\infty]{} \mathbf{B} \qquad (18)$$

where

$$\mathbf{B} = \frac{m}{m+1}\left(\mathbf{I} - \frac{1}{m}\text{vec}(\mathbf{I})\text{vec}(\mathbf{I})^T\right) \qquad (19)$$

Therefore

$$NE[\boldsymbol{\delta}_{FP}\boldsymbol{\delta}_{SCM}^H] \xrightarrow[N\to+\infty]{} \frac{m+1}{m}\mathbf{B} \qquad (20)$$

Since the covariance matrix is Hermitian and using the property $(\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}) = (\mathbf{AB} \otimes \mathbf{CD})$ (see [15]), one has $\boldsymbol{\Sigma}_2^{(N)} \xrightarrow[N\to+\infty]{} \boldsymbol{\Sigma}_2$ where

$$
\begin{aligned}
\boldsymbol{\Sigma}_2 &= (\mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2})(\mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2})^H \\
&- \frac{1}{m}(\mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2})\text{vec}(\mathbf{I})\text{vec}(\mathbf{I})^T(\mathbf{M}^{T/2} \otimes \mathbf{M}^{1/2})^H \\
&= (\mathbf{M}^T \otimes \mathbf{M}) - \frac{1}{m}\text{vec}(\mathbf{M})\text{vec}(\mathbf{M})^H \qquad (21)
\end{aligned}
$$

Finally, we obtain the expression of $\boldsymbol{\Sigma}$

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_1 - 2\boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_3 \\
&= (\frac{m+1}{m} - 2 + 1)(\mathbf{M}^T \otimes \mathbf{M}) \\
&+ (-\frac{m+1}{m^2} + \frac{2}{m})\text{vec}(\mathbf{M})\text{vec}(\mathbf{M})^H \\
&= \frac{1}{m}(\mathbf{M}^T \otimes \mathbf{M}) + \frac{m-1}{m^2}\text{vec}(\mathbf{M})\text{vec}(\mathbf{M})^H \quad (22)
\end{aligned}
$$

The asymptotic pseudo-covariance matrix $\boldsymbol{\Omega}$ is defined as

$$\boldsymbol{\Omega} = NE[\text{vec}(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM})\text{vec}(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM})^T] \quad (23)$$

As derived in [7], $\boldsymbol{\Omega} = \boldsymbol{\Sigma}\mathbf{K}$ and $\text{vec}(\mathbf{M})^H\mathbf{K} = \text{vec}(\mathbf{M})^T$, which leads to the result of theorem III.1 and concludes the proof. ∎

IV. SIMULATIONS

In this section some simulation results are presented. The results are obtained for complex Gaussian-distributed[1] zero-mean data with covariance matrix $\mathbf{M}$ whose elements are defined by $M_{i,j} = \rho^{|i-j|}$, $i = 1,...,m$. Operator $\overline{\mathbf{A}}$ is defined as the empirical mean of the quantities $\mathbf{A}(i)$ obtained from I Monte Carlo runs. For each iteration i, we generate a new set of N secondary data $\mathbf{x}_1,...,\mathbf{x}_N$ to compute $\mathbf{A}(i)$.

The correlation coefficient ρ is first set to 0. In other words, the true covariance matrix is equal to identity. It should be

---

[1]Note that a change in the underlying distribution will not affect the results since Tyler's estimator is distribution-free over the class of CES.
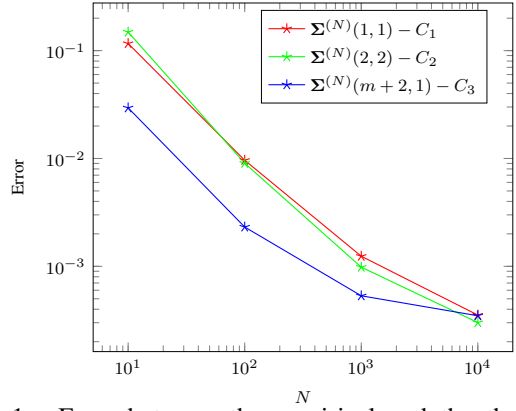

Figure 1: Error between the empirical and the theoretical results for three different elements

noticed that in that case the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ contain three different sorts of nonzero elements: $C_1 = \frac{1}{m} + \frac{m-1}{m^2}$, $C_2 = \frac{1}{m}$ ans $C_3 = \frac{m-1}{m^2}$.

Figure 1 presents the difference between an element of the covariance matrix of $\sqrt{N}\text{vec}(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{M}}_{SCM})$ (this matrix will be denoted as $\boldsymbol{\Sigma}^{(N)}$) and its theoretical asymptotic value. The results for the first diagonal, the second diagonal and a nondiagonal element are plotted. The horizontal scale presents the number of samples N used to estimate the covariance matrix. The dimension of the data is set to 5. One can see that the error tends to zero as the number of samples increases confirming our previous findings (Theorem III.1).

Figure 2 depicts the values of three different element versus the dimension of the data. The number of samples is set to 1000. The red dashed line presents the empirical results, while the green smooth one draws the theoretical ones. First, one can see that the empirical results coincide with the theoretical ones if N is not very large. Second, one can notice that the values of elements decrease as the dimension m increases (tend to 0 when $m \to \infty$) which justifies the statement that the behavior of Tyler's estimator gets closer to the one of the SCM as the dimension of vectors increases and reinforces the usefulness of the result in the large dimension case.

From (10) one can conclude that the asymptotic pseudo-covariance matrix $\boldsymbol{\Omega}$ contains the same elements as the covariance matrix $\boldsymbol{\Sigma}$ but at different positions, due to the commutation matrix $\mathbf{K}$.

Let us now focus on a more general case when the data are correlated with the correlation factor equal to 0.5. Figure 3 illustrates the first diagonal element of $\boldsymbol{\Sigma}^{(N)}$ and its theoretical asymptotic value versus the dimension m. One can note that the results are a little bit degraded but still good, knowing that the number of samples N is 1000. It worthwhile noting that with a bigger number of samples N the empirical results would be more accurate.

V. CONCLUSION

In this paper we have analyzed the asymptotic distribution of Tyler's estimator when the observed data have a CES distribu-
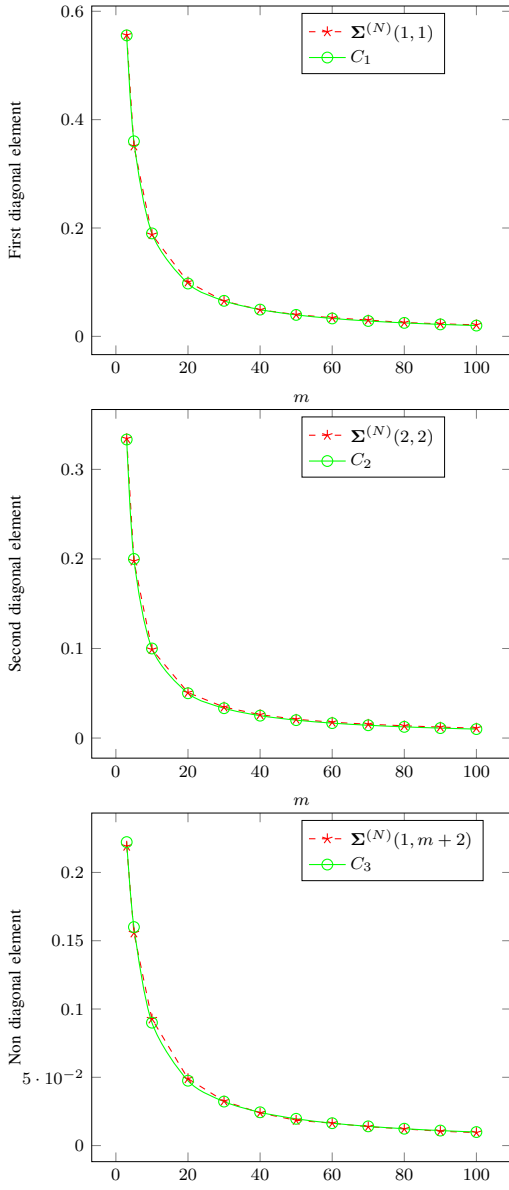
Figure 2: Three different elements of $\mathbf{\Sigma}^{(N)}$ compared with their theoretical asymptotic values

Figure 3: First diagonal element of $\mathbf{\Sigma}^{(N)}$ compared with its theoretical asymptotic value from equation (10) for $\rho = 0.5$

REFERENCES

[1] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*, ser. New York, NY. USA:Springer-Verlag, 1999.

[2] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, January 1964.

[3] F. Pascal, Y. Chitour, J.-P. 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, January 2008.

[4] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4097–4107, 2011.

[5] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 29–38, 2012.

[6] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized gaussian distributions," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5960–5971, December 2013.

[7] 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, July 2013.

[8] M. Mahot, P. Forster, J.-P. Ovarlez, and F. Pascal, "Robustness analysis of covariance matrix estimates," *European Signal Processing Conference (EUSIPCO), Aalborg, Denmark*, August 2010.

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

[10] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 32, no. 4, pp. 419–430, December 1970.

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

[12] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Performance analysis of covariance matrix estimates in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2206–2217, June 2008.

[13] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, November 2012.

[14] R. Couillet, F. Pascal, and J. W. Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples," *Journal of Multivariate Analysis*, vol. 139, pp. 56–78, July 2015.

[15] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," nov 2012, version 20121115. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274

tion. We have demonstrated that it behaves as the SCM in the Gaussian context, even if the samples have a CES distribution. The upshot of this is the possibility of using Tyler's estimator in practice and approximating its properties by the ones of Wishart distributions. Since the Wishart distribution is well-known, this allows us to better understand the behavior of Tyler's estimator in the context of CES-distributed data. For instance, in signal detection problems, this result allows us to use Tyler's estimator and to theoretically adjust parameters (e.g., the detection threshold) 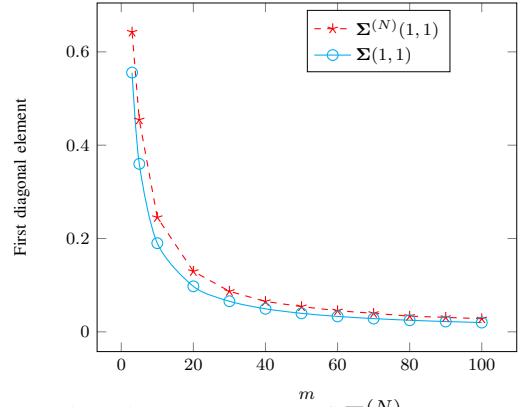thanks to the properties of the SCM.