A Toeplitz-Tyler Estimation of the Model Order in Large Dimensional Regime
Eugénie Terreaux, Jean-Philippe Ovarlez, Frédéric Pascal

To cite this version:
Eugénie Terreaux, Jean-Philippe Ovarlez, Frédéric Pascal. A Toeplitz-Tyler Estimation of the Model Order in Large Dimensional Regime. ICASSP 2018 - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr 2018, Calgary, Canada. pp.4489-4493, 10.1109/ICASSP.2018.8461915. hal-02124624

HAL Id: hal-02124624
https://hal-centralesupelec.archives-ouvertes.fr/hal-02124624
Submitted on 17 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
This paper presents a new algorithm to estimate the number of sources embedded in a correlated Complex Elliptically Distributed (CES) noise in the context of large dimensional regime. The proposed method is a two-steps ones: first the data covariance matrix is estimated with a robust and consistent estimator exploiting the Toeplitz structure assumption of the true scatter matrix. Then, after whitening the signal thanks to the first estimator, the distribution of its Tyler estimator eigenvalues is studied, as in [1]. This allows to derive a threshold, estimated thanks to the Marchenko-Pastur law, to separate the eigenvalues corresponding to the noise and those corresponding to the sources. The number of sources can therefore be deduced. The proposed method is compared to classical ones as the Akaike Information Criterion (AIC) or other algorithms recently developed.

1. INTRODUCTION

The problem of estimating the number of sources, or equivalently the model order, is a recurrent and challenging issue in a lot of signal processing fields (see e.g., [2], [3] or [4]). To that end, statistical methods often rely on the estimation of the covariance matrix eigenvalues. This is the case for instance for the MUSIC method [5], the Akaike Information Criterion (AIC) [6] or others methods of model order selection [7]. Thereby, with such methods, the better the covariance matrix estimation is, the better the number of sources estimation is. In this context of number of sources estimation, this article introduces new results when the observations turn to be non-Gaussian and of large dimension.

Large dimensional regime corresponds to the following regime: the number of received signals \( N \) and the dimension \( m \) tend to infinity at the same rate (see e.g., [8]). For white or whitened noise, several methods, based on the Random Matrix Theory (RMT) have been proposed to extract information of interest from the received signals. One can cite for instance the number of embedded sources estimation [9], the problem of radar detection [10], signal subspace estimation with an adapted MUSIC detection algorithm [11]. However, when the additive noise is correlated, some RMT methods require the estimation of a specific threshold which has no explicit expression and can be very difficult to obtain [7, 12] while the others assume that the covariance matrix is known and use it to whiten the signal.

Concerning non-Gaussian signals, robust methods have been recently developed based on Complex Elliptically Symmetric (CES) distribution modelling. CES distributions belong to a wide-class of distributions including the Gaussian distribution, the K-distribution and others, as detailed in [13]. Its flexibility enables to model a large panel of random signals and justifies its wide use in signal processing applications. To deal with non-Gaussian data, robust methods adapted to large dimensional observations [14] have been developed in the case of white (or perfectly whitened) noise.

To fill this gap, [15] proposed a method that relaxes this strong assumption thanks to a Toeplitz-based correlation model. Moreover, this methodology is based on Maronna’s assumptions [16], that is to say, for well-behaved weighting functions \( u(\cdot) \) appearing in the definition of the covariance \( M \)-estimators. For model order selection, the threshold proposed in [15] depends on the selected function \( u(\cdot) \) and remains quite difficult to derive. This is the case for instance for the Huber [17] \( M \)-estimators.

This article expands former results of [15] to the case of Tyler [18] \( M \)-estimators for correlated CES noise. Note that Tyler \( M \)-estimators do not respect Maronna’s conditions but provide more flexibility since it is a distribution-free estimator [19]. Moreover, recent works in [1] proved that the eigenvalues distribution of the Tyler \( M \)-estimator follows, in the white CES case, the Marchenko-Pastur distribution [20]. The main contribution of this article is first to propose a consistent estimator of the covariance matrix for a correlated CES noise. Then, it aims at providing an estimation of the number of sources when the signal is constituted of a correlated CES...
This estimation process is shown to be easier than those proposed in [15]. The proposed method splits into two steps as in [15]: the first one is the whitening of the signal with a good estimated covariance matrix. This estimated whitening covariance matrix is obtained thanks to a Tyler estimator and a Toeplitz rectification. The second step consists in estimating the eigenvalues of the Tyler covariance matrix of the whitened signal and then, it exploits the support of the Marchenko-Pastur law to find the number of sources in the signal. These steps are described in the two first parts of this article. The third part shows a comparison in terms of performance between the proposed method, the method proposed in [15] and AIC through some simulations.

Notations: Matrices are in bold and capital, vectors in bold. Let \( \mathbf{X} \) be a square matrix of size \( s \times s \), \( (\lambda)_i(\mathbf{X}) \), \( i \in [1, ..., s] \), are the eigenvalues of \( \mathbf{X} \). \( \operatorname{Tr}(\mathbf{X}) \) is the trace of the matrix \( \mathbf{X} \). \( \| \mathbf{X} \| \) stands for the spectral norm. Let \( \mathbf{A} \) be a matrix, \( \mathbf{A}^T \) is the transpose of \( \mathbf{A} \) and \( \mathbf{A}^H \) the Hermitian transpose of \( \mathbf{A} \). \( \mathbf{I}_n \) is the \( n \times n \) identity matrix. For any \( m \)-vector \( \mathbf{x} \), \( \mathcal{L} : \mathbf{x} \mapsto \mathcal{L}(\mathbf{x}) \) is the \( m \times m \) matrix defined as the Toepplitz operator: \( ([\mathcal{L}(\mathbf{x})]_{i,j})_{i,j \leq j} = x_{i-j} \) and \( ([\mathcal{L}(\mathbf{x})]_{i,j})_{i,j > j} = x_{i-j}^* \). For any matrix \( \mathbf{A} \) of size \( m \times m \), \( \mathcal{T}(\mathbf{A}) \) (the scatter matrix \( \mathcal{L}(\mathbf{a}) \) where \( \mathbf{a} \) is a vector for which each component \( a_i, 0 < i < m - 1 \) contains the sum of the \( i \)-th diagonal of \( \mathbf{A} \) divided by \( m \). For \( x \in \mathbb{R} \), \( \delta_0 \) is the Dirac measure at \( x \). \( \text{supp}(\mathbf{a}) \) is the support of \( \mathbf{a} \). The notation \( \overset{\text{a.s.}}{\rightarrow} \) means "tends to almost surely". The notation \( \text{dist} \) stands for the distance associated to the \( L_1 \) norm.

1.1. Model and Assumptions

The \( N \) observations \( \mathbf{Y} = \{\mathbf{y}_0, ..., \mathbf{y}_{N-1}\} \) are made of a correlated CES noise added to \( p \) mixed sources. The equation (1) sets down the model:

\[
\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j} \mathbf{m}_j + \sqrt{\tau_i} \mathbf{C}^{1/2} \mathbf{x}_i, \quad i \in [0, N-1],
\]

where the \( \{\tau_i\}_{i \in [0,N-1]} \) are positive random variables and where \( \mathbf{X} = [\mathbf{x}_0, ..., \mathbf{x}_{N-1}] \) is the matrix of independant multivariate \( m \)-vectors uniformly distributed on the sphere of dimension \( m \). Moreover, the \( m \times p \) matrix \( \mathbf{M} \) with elements \( M_{i,j} = (\mathbf{M})_{i,j} = (\mathbf{m}_j)_i \) is referred to as the mixing matrix and contains the \( p \) vectors of the sources. Each element \( s_{i,j} \) of the \( p \times N \) matrix \( \mathbf{S} \) corresponds to the amplitude variation of each source in the received vector.

The matrix \( \mathbf{C} \) is assumed to be a nonnegative definite Toeplitz matrix:

\[
\mathbf{C} = \begin{pmatrix}
    c_0 & c_1 & \ldots & c_{m-1} \\
    c_1 & c_0 & \ldots & c_{m-2} \\
    \vdots & \vdots & \ddots & \vdots \\
    c_{m-1} & c_{m-2} & \ldots & c_0
\end{pmatrix}.
\]

The signal can thus be written as:

\[
\mathbf{Y} = \mathbf{M} \mathbf{S} + \mathbf{C}^{1/2} \mathbf{X} \mathbf{T}^{1/2},
\]

where \( \mathbf{T} \) is the \( N \times N \)-diagonal matrix containing the \( \{\tau_i\}_{i \in [0,N-1]} \). In the sequel, we will consider the following assumptions: first, \( N \to \infty, m \to \infty \) and \( c_N = \frac{m}{N} \to c > 0 \).

Then \( \mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\tau_i} \) satisfies \( \tau \mu_N(d\tau) \to 1 \) almost surely and \( \frac{1}{N} \sum_{i=1}^{N} \delta_{\tau_i} \) converges almost surely toward the true measure \( \nu \). We will also assume that \( \max \text{dist}(\lambda_i(\mathbf{C}), \text{supp}(\nu)) \to 0 \) and that \( \{c_k\}_{k \in [0,m-1]} \) are absolutely summable coefficients, such that \( c_0 \neq 0 \).

The problem considered here is the estimation of \( p \), the number of sources. Next section presents the main contribution of this article.

2. RESULTS

2.1. Signal Whitening

The first step of the proposed algorithm is to whiten the signal with a consistent estimation of \( \mathbf{C} \), the scatter matrix of the CES noise. The proposed estimator is a Tyler \( M \)-estimator of the scatter matrix \( \hat{\mathbf{C}} \) enforced to be Toeplitz-structured with the operator \( \mathcal{T}(\hat{\mathbf{C}}) \). The Tyler \( M \)-estimator \( \hat{\mathbf{C}} \) is defined as the unique solution if it exists of:

\[
\Sigma = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{y}_i \mathbf{y}_i^H}{\mathbf{y}_i^H \Sigma^{-1} \mathbf{y}_i}.
\]

The consistency of these estimator is proven thanks to [1], [15] and the following theorem:

**Theorem 1** (A consistent estimation of \( \mathbf{C} \)). Under the prior assumptions, we have the following convergence:

\[
\left\| \mathcal{T} (\hat{\mathbf{C}}) - \mathbf{C} \right\| \overset{\text{a.s.}}{\rightarrow} 0.
\]

**Proof.** Let \( \hat{\mathbf{S}} \) be the Sample Covariance Matrix (SCM) that is \( \hat{\mathbf{S}} = \frac{1}{N} \mathbf{X} \mathbf{X}^H \). Then:

\[
\left\| \mathcal{T} (\hat{\mathbf{C}}) - \mathbf{C} \right\| \leq \left\| \mathcal{T}(\hat{\mathbf{C}}) - \mathcal{T}(\hat{\mathbf{S}}) \right\| + \left\| \mathcal{T}(\hat{\mathbf{S}}) - \mathbf{C} \right\|.
\]

Moreover, it is proven in [15] that \( \left\| \mathcal{T}(\hat{\mathbf{S}}) - \mathbf{C} \right\| \overset{\text{a.s.}}{\rightarrow} 0 \). Since [1] shows that \( \hat{\mathbf{C}} - \mathbf{S} \overset{\text{a.s.}}{\rightarrow} 0 \) and as \( \mathcal{T} \) is a linear operator, it appears that \( \left\| \mathcal{T} (\hat{\mathbf{C}}) - \mathcal{T}(\hat{\mathbf{S}}) \right\| \) and that \( \left\| \mathcal{T} (\hat{\mathbf{C}}) - \mathbf{C} \right\| \overset{\text{a.s.}}{\rightarrow} 0 \). \( \square \)

The consistency of this estimator is shown on Figure 1 where the \( \{\tau_i\}_{i \in [1,N]} \) are inverse gamma distributed. The
rate of the convergence is relatively slow but it is possible to enhance it with banding or tapering (see for example [21]).

\[ \hat{W} = \hat{C}^{-1/2} \hat{W} C^{1/2} \hat{C}^{-1/2} \]  

which leads to:

\[ \hat{W} = \hat{C}^{-1/2} \hat{W} C^{1/2} \hat{C}^{-1/2} \]  

The equation (8) can be rewritten as:

\[ \|\hat{W} - \hat{S}_w\| \leq \|\hat{W} - W\| + \|\hat{S}_w - W\| . \]  

(11)

Moreover, it is proven in [1] that \( \|\hat{W} - \hat{S}_w\| \xrightarrow{a.s.} 0 \). Concerning the first term, it can be rewritten as:

\[ \|\hat{W} - \hat{W}\| \leq \|\hat{C}^{-1/2} \hat{W} C^{1/2} \hat{C}^{-1/2} - \hat{W} C^{1/2} \hat{C}^{-1/2}\| + \|\hat{W} C^{1/2} \hat{C}^{-1/2} - \hat{W}\| . \]  

(12)

After left and right factorization, one obtains:

\[ \|\hat{W} - \hat{W}\| \leq \|\hat{C}^{-1/2} \hat{W} C^{1/2} - I_n\| \|\hat{W}\| \left( \|C^{1/2} \hat{C}^{-1/2}\| + 1 \right) . \]

As \( \|C\| \) has a bounded support, \( \|\hat{W}\| \) has a bounded support too thanks to Theorem 1. Moreover, this theorem proves the consistency as \( \|C - \hat{C}\| \xrightarrow{a.s.} 0 \) and this ensures the proof.

Fig. 1. Consistency of the proposed estimator of \( C, N \in [5, 10000], \) 

c = 0.7, \{\tau\} \in [1, N] following an inverse gamma distribution

Let us note \( \hat{C} \) the estimator \( \mathcal{T}(\hat{C}) \). The signal can be whitened as following:

\[ \hat{Y}_w = \hat{C}^{-1/2} MS + \hat{C}^{-1/2} C^{1/2} XT^{1/2} , \]

with \( \hat{Y}_w = [\hat{y}_{w1}, \ldots, \hat{y}_{wN}] \).

2.2. Estimation

The signal being whitened, it is now possible to apply a Tyler M-estimator in order to threshold its eigenvalues. Let \( \hat{W} \) denote a Tyler estimation of the “white” scatter matrix, that is, \( \hat{W} \) is the unique solution if it exists of:

\[ \Sigma_w = m \sum_{i=1}^{N} \frac{\hat{y}_{wi} \hat{y}_{wi}^H}{\hat{y}_{wi}^H \hat{y}_{wi}} . \]

Let us define \( \hat{S}_w = \frac{1}{N} XX^H \). In order to threshold its eigenvalues and estimate the number of sources, Theorem 2 enables to choose the threshold as the right edge of the Marchenko-Pastur law support [20]:

Theorem 2 (Convergence of \( \hat{W} \)). Under same assumptions as previously, 

\[ \|\hat{W} - \hat{S}_w\| \xrightarrow{a.s.} 0 . \]  

(8)

Proof. It follows the one of Theorem 2 in [15]. Let \( \hat{W} \) denote the Tyler M-estimator of \( Y_w = C^{-1/2} MS + X T^{1/2} \) with \( Y_w = [y_{w1}, \ldots, y_{wN}] \), i.e., \( \hat{W} = m \sum_{i=1}^{N} \frac{y_{wi} y_{wi}^H}{y_{wi}^H y_{wi}} . \)

Since \( \hat{W} = m \sum_{i=1}^{N} \frac{\hat{y}_{wi} \hat{y}_{wi}^H}{\hat{y}_{wi}^H \hat{y}_{wi}} \), if we replace \( \hat{y}_{wi} \) by \( \hat{C}^{-1/2} C^{1/2} y_{wi} \), \( 0 \leq i \leq N \), one obtains:

\[ C^{-1/2} \hat{C}^{1/2} \hat{W} \hat{C}^{1/2} C^{-1/2} = m \sum_{i=1}^{N} \frac{y_{wi} y_{wi}^H}{y_{wi}^H \left( C^{-1/2} \hat{C}^{1/2} \hat{W} C^{1/2} C^{-1/2} \right)^{-1} y_{wi}} . \]

As detailed in [15], the equation (8) can be rewritten as:

\[ \|\hat{W} - \hat{S}_w\| \leq \|\hat{W} - W\| + \|\hat{S}_w - W\| . \]  

(11)

Moreover, it is proven in [1] that \( \|\hat{W} - \hat{S}_w\| \xrightarrow{a.s.} 0 \). Concerning the first term, it can be rewritten as:

\[ \|\hat{W} - \hat{W}\| \leq \|\hat{C}^{-1/2} \hat{W} C^{1/2} \hat{C}^{-1/2} - \hat{W} C^{1/2} \hat{C}^{-1/2}\| + \|\hat{W} C^{1/2} \hat{C}^{-1/2} - \hat{W}\| . \]  

(12)

After left and right factorization, one obtains:

\[ \|\hat{W} - \hat{W}\| \leq \|\hat{C}^{-1/2} \hat{W} C^{1/2} - I_n\| \|\hat{W}\| \left( \|C^{1/2} \hat{C}^{-1/2}\| + 1 \right) . \]

As \( \|C\| \) has a bounded support, \( \|\hat{W}\| \) has a bounded support too thanks to Theorem 1. Moreover, this theorem proves the consistency as \( \|C - \hat{C}\| \xrightarrow{a.s.} 0 \) and this ensures the proof.

2.3. Simulations

In order to enlighten these theoretical results, Figure 2 shows that the eigenvalues distribution of the estimated scatter matrix \( \hat{W} \) follows the Marchenko-Pastur distribution. On this figure, the \( \{\tau\} \in [1, N] \) are inverse gamma distributed. The 4 sources with SNR=10 dB present in the signal can be detected since 4 eigenvalues are found upon the Marchenko-Pastur distribution support, even if the SNR is low.

Then, Figure 3 shows the eigenvalues distribution with the same parameters as in Figure 2 but for a Fixed Point estimation of \( C \) noted FP, directly defined as the unique solution of \( \Sigma = \sum_{i=1}^{N} u \left( y_{i} H \Sigma^{-1} y_{i} \right) y_{i} y_{i}^H \) build with the function \( u(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{t + x} \), \( t = 0.1 \) and compared to the Marchenko-Pastur law. As expected, this distribution does not follow the Marchenko-Pastur law and the chosen threshold is no more relevant. To find a correct threshold for this estimator, it is necessary to whiten the signal first and then calculate a more complicated threshold, without explicit expression, as detailed in [15].
Eigenvalues and the Marchenko-Pastur distribution

Fig. 2. The Marchenko-Pastur distribution and the eigenvalues distribution of $\tilde{W}$, 4 sources, SNR=10 dB, $N = 2000$, $m = 900$, $c = 0.7$, $\{\tau \}_{i \in [1, N]}$ are inverse gamma distributed.

Detection of the number $\hat{p}$ of sources

Fig. 3. 4 sources, $N = 2000$, $m = 900$, $\{\tau \}_{i \in [1, N]}$ are inverse-gamma distributed.

Detection of the number $\hat{p}$ of sources

Fig. 4. 4 sources, $N = 2000$, $m = 900$, $\{\tau \}_{i \in [1, N]}$ are inverse-gamma distributed.

The number $\hat{p}$ of sources detected in the signal is plotted on Figure 4, for different SNR. Our method, denoted TylTER, is compared to the method of [15] noted FPTER, and to the AIC method. The true number of sources is also plotted on the Figure. The scale is in logarithm and the SNR comes from 0 to 30 dB. The AIC method overestimates the number of sources as our method finds the true number of sources from a SNR = 20dB. The method FPTER begins to find sources for higher SNR than TylTER.

On Figure 5 are drawn the same results but for the $\{\tau \}_{i \in [1, N]}$ Student-t distributed. As expected, the proposed method still detects the sources for correct SNR. The results are almost the same except that the SNR required for detecting sources is higher for the methods TylTER and FPTER. Moreover, the FPTER method detects sources for a higher SNR than for the distribution of the $\tau_i$ in Figure 4. Indeed, in Figure 4, the chosen function $u$ for FPTER is the optimal one for the $\tau_i$. As the $\tau_i$ are slightly different but the function $u$ is the same, the method FPTER is still robust but need higher SNR than TylTER to detect all the sources. These results are very promising and encourage the application of the proposed method to a large panel of signals.

3. CONCLUSION

A new robust algorithm has been proposed in order to estimate the number of sources in a signal. The algorithm consists first in a whitening the signals and then in estimating the scatter matrix eigenvalues distribution. Simulations show the efficiency of the method, even if the noise is different from what expected. Two major results of this paper are first that this algorithm is derived for correlated CES noise in large dimensional regime and secondly that it proves it is no longer necessary to determine a complicated threshold as in [15]: the right edge of the Marchenko-Pastur law support can simply be used. Thereby, independently of the $\tau_i$ distribution, the proposed threshold does not require any a priori on the noise.
4. REFERENCES


