Robust Model Order Selection in Large Dimensional Elliptically Symmetric Noise

Eugénie Terreaux*, Jean-Philippe Ovarlez*†, and Frédéric Pascal‡

*CentraleSupélec-SONDRA, 3 rue Joliot-Curie, 91190 Gif-sur-Yvette, France

(e-mail: eugenie.terreaux@centralesupelec.fr)

†ONERA, DEMR/TSI, Chemin de la Hunière, 91120 Palaiseau, France

*L2S/CentraleSupélec-Université Paris-Sud, 3 rue Joliot-Curie, 91190 Gif-sur-Yvette, France

arXiv:1710.06735v1 [stat.ME] 18 Oct 2017

Abstract—This paper deals with model order selection in context of correlated noise. More precisely, one considers sources embedded in an additive Complex Elliptically Symmetric (CES) noise, with unknown parameters. The main difficulty for estimating the model order lies into the noise correlation, namely the scatter matrix of the corresponding CES distribution. In this work, to tackle that problem, one adopts a two-step approach: first, we develop two different methods based on a Toeplitz-structured model for estimating this unknown scatter matrix and for whitening the correlated noise. Then, we apply Maronna's M-estimators on the whitened signal to estimate the covariance matrix of the "decorrelated" signal in order to estimate the model order. The proposed methodology is based both on robust estimation theory as well as large Random Matrix Theory, and original results are derived, proving the efficiency of this methodology. Indeed, the main theoretical contribution is to derive consistent robust estimators for the covariance matrix of the signal-plus-correlated noise in a large dimensional regime and to propose efficient methodology to estimate the rank of signal subspace. Finally, as shown in the analysis, these results show a great improvement compared to the state-of-the-art, on both simulated and real hyperspectral images.

Index Terms—Model order selection, RMT, correlated noise, CES distribution, robust estimation.

I. INTRODUCTION

MODEL order selection is a challenging issue in signal processing for example in wireless communication [1], array processing [2], or other related problems [3], [4]. Classically, for a white noise, statistical methods such as the one based on the application of the information theoretic criteria for model order selection, allow to estimate the model order thanks to eigenvalues and eigenvectors of the covariance matrix of the signal. This is the case of the Akaike Information Criterion (AIC) [5] or the Minimum Description Length (MDL) [6], [7]. Other examples are the problem of source localization [8], where the estimation of the signal subspace is done by the estimation of the eigenvalues of the covariance matrix, channel identification [9], waveform estimation [10] and many other parametric estimation problems. Though, all these methods are no more relevant for large dimensional and correlated data. Even if particular cases have been studied for correlated signals as in [11] or [12], these methods can not be generalized for all kind of signals and a whitening step, when possible, can not be systematically set up [13]. Moreover, the commonly used statistical model for this problem has not the same matrix properties when the data are large and when they

are not: the covariance matrix is not correctly apprehended and the methods fail to estimate the model order, for example in [14], in [15] or in [16]. In the field of model order selection for large dimensional regime, that is when the number of snapshots N and the dimension of the signal m tend to infinity with a constant positive ratio, and for white or whitened noise, the Random Matrix Theory (RMT) proposes methods to estimate the model order selection relying on the study of the largest eigenvalues distribution of the covariance matrix [17]. The RMT introduces new methodologies which correctly handle the statistical properties of large matrices thanks to a statistical and probability approach: see [18] for a review of this theory, [19] for a general detection algorithm, [20] for an adapted MUSIC detection algorithm, [21] for applications to radar detection and [22] for an application on hyperspectral imaging. When the noise is spatially correlated, it is still possible to estimate the model order for example by evaluating the distance between the eigenvalues of the covariance matrix [23]. Nevertheless, these methods require a threshold that has no explicit expression and can be fastidious to obtain [24]. In addition to the problem of the large dimension and the correlation, another recurrent problem in signal processing is the non-Gaussianity of the noise. To be less dependent of the noise statistic, that is for the model order selection not to be degraded with a noise more or less sightly different than targeted, robust methods for model order selection have been developed [25] in hyperspectral imaging [26]. Nevertheless, these methods depend on unknown parameters [1] or are not adapted for large data. Recent results in RMT enable to correctly estimate the covariance matrix for textured signals [27]. But the correlation matrix is assume to be known and the signal is whitened before processed.

In this works, one considers a Complex Elliptically Symmetric (CES) noise. The CES distributions modelling is often exploited in signal processing, because of its flexibility, that is the ability to model a large panel of random signals. The signal can be split in two parts: a texture and a speckle. They are rather often used in various fields, as in [28] for hyperspectral imaging, or [29] for radar clutter echoes modelling. This article deals with large dimansional non-Gaussian data, and proposes a robust method to estimate the model order. The robustness of our method comes from the robust estimation of the covariance matrix, with a Maronna M-estimator [30] which assigns different weights according to the Mahalanobis

distance between the signals received by the different sensors. It is a generalization of [27] and [31] to the case of left hand side correlation (with an unknown covariance matrix). Moreover, this article proposes a new algorithm to estimate the model order.

In a first part, an estimator for the correlation matrix is presented: the toeplitzified Sample Covariance Matrix (SCM), that is, the SCM enforced to be of Toeplitz form [32]. Indeed, as the covariance matrix is supposed to be Toeplitz, the SCM is toeplitzified as in [33] to enhance the estimation. The data are then whitened with this Toeplitz matrix and a robust Maronna M-estimator of the covariance matrix is then used after the data whitening. This robust estimation is studied and a threshold on its eigenvalues can be derived to select the model order. A second part presents the same procedure for the toeplitzified Fixed-Point (FP) estimator [34] and [35]. The third part presents some simulations on both simulated and real hyperspectral images. Proofs of the main results are postponed in the appendices.

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 $\mathbf{X}$. $Tr(\mathbf{X})$ 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 Toeplitz operator : $([\mathcal{L}(\mathbf{x})]_{i,j})_{i \leq j} = x_{i-j}$ and $([\mathcal{L}(\mathbf{x})]_{i,j})_{i > j} = x_{i-j}^*$. For any matrix $\mathbf{A}$ of size $m \times m$, $\mathcal{T}(\mathbf{A})$ represents the matrix $\mathcal{L}(\check{\mathbf{a}})$ where $\check{\mathbf{a}}$ is a vector for which each component $\check{\mathbf{a}}_{i, 0 < i < m-1}$ contains the sum of the $i-$th diagonal of $\mathbf{A}$ divided by m. For $x \in \mathbb{R}$, δ_x is the Dirac measure at x. For any complex z, $z^\star$ is the conjugate of z. The notation *dist* stands for the distance associated to the L_1 norm. supp is the support of a set. Eventually, $\mathcal{R}e$ and $\mathcal{I}m$ stand respectively for the real and the imaginary part for a complex number. The notation $\xrightarrow{a.s.}$ means "tends to almost surely".

II. MODEL AND ASSUMPTIONS

This section introduces the model as well as the general assumptions needed to derive the results. Let us consider the following general sources-plus-noise model. Let $\mathbf{Y} = [\mathbf{y}_0, \ldots, \mathbf{y}_{N-1}]$ be a matrix of size $m \times N$, containing N observations $\{\mathbf{y}_i\}_{i \in [\![0, N-1]\!]}$ of size m, constituted of p mixed sources corrupted with an additive noise:

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

which can be rewritten as

$$\mathbf{Y} = \mathbf{MS} + \mathbf{C}^{1/2} \, \mathbf{X} \, \mathbf{T}^{1/2}, \quad (2)$$

where the $\{\tau_i\}_{i \in [\![0, N-1]\!]}$ are positive random variables, and $\mathbf{T}$ is the $N \times N$-diagonal matrix containing the $\{\tau_i\}_{i \in [\![0, N-1]\!]}$. 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. In this work, the additive noise is modelled thanks to the general family of Complex Elliptically Symmetric (CES) distributions [36], [37]

(see also [38] for more details on CES as well as their use in signal processing). Thus, each component of the noise is characterized by a random vector $\mathbf{x}_i$ uniformly distributed on a sphere times an independent positive random scalar τ_i with unspecified probability distribution function. The left hand side spectral correlation is handled by the scatter matrix $\mathbf{C}$.

Each element $s_{i,j}$ of the $p \times N$ matrix $\mathbf{S}$ corresponds to the power variation of each source in the received vector. This matrix can be written $\mathbf{S} = \boldsymbol{\delta}^H \, \boldsymbol{\Gamma}^{1/2}$ where $\boldsymbol{\delta}$ is a $N \times p$ random matrix, independent of $\mathbf{X}$, whose elements are normally distributed with zero-mean and unit variance. $\boldsymbol{\Gamma}$ is a $N \times N$ Hermitian covariance matrix. Eventually, $\mathbf{C} = \mathcal{L}\left([c_0, \ldots, c_{m-1}]^T\right)$ is a $m \times m$ Hermitian nonnegative definite Toeplitz matrix:

$$\mathbf{C} = \begin{pmatrix} c_0 & c_1 & ... c_{m-1} \\ c_1^\star & c_0 & ... c_{m-2} \\ ... & & \\ c_{m-1}^\star & c_{m-2}^\star ... & c_0 \end{pmatrix}.$$

In the sequel, we will consider the following assumptions:
Assumption 1: One assumes the usual random matrix regime, *i.e.*: $N \to \infty$, $m \to \infty$ and $c_N = \dfrac{m}{N} \to c > 0$,
Assumption 2: for matrices $\mathbf{C}$, $\mathbf{T}$ and $\mathbf{X}$ of equation (2), one has:

- $\text{dist}(\lambda_i(\mathbf{C}), \text{supp}(\nu)) \to 0$ with ν the limit of $\frac{1}{N} \sum \delta_{\lambda_i(\mathbf{C})}$ when $N \to \infty$,
- $\{c_k\}_{k \in [\![0, m-1]\!]}$ are absolutely summable coefficients, such that $c_0 \neq 0$,
- The random measure $\mu_N = \dfrac{1}{N} \sum_{i=1}^N \delta_{\tau_i}$ satisfies $\displaystyle\int \tau \, \mu_N(d\tau) \xrightarrow{a.s.} 1$,
- $\mathbf{X}$ is a white noise, with independent and identically distributed entries with zero-mean and with unit variance,
- $\|\boldsymbol{\Gamma}\| < \infty$ and $\|\mathbf{M}\| < \infty$.

Assumption 3:
- In each column of $\mathbf{M}$, the coefficients are absolutely summable that is, for all fixed j, $\sum_{i=1}^m |M_{i,j}| < \infty$. This is a common assumption in several applications and especially in hyperspectral imaging.
- $\boldsymbol{\Gamma}$ has coefficients absolutely summable.

Assumption 4: Let $[\mathbf{Y}]_{i,j} = [\mathbf{y}_i]_j$, then the coefficients $[\mathbf{y}_i]_j$ are absolutely summable, that is, for a fixed i, $\sum_j |[\mathbf{y}_i]_j|$ exists.

III. MODEL ORDER SELECTION: A GAUSSIAN APPROACH

In this section, the consistency of the SCM is used to whiten the signal and to estimate the model order thanks to a Maronna M-estimator. The step which consists in directly evaluating the model order with a Maronna M-estimator has been already studied in [27] for the special case of *spiked* model with CES white noise. In this work, one considers the more challenging problem of correlated CES noise.

A. Whitening Step

The noise being correlated, one proposes here to whiten it using the Toeplitz structure of the noise covariance matrix. As a reminder, the model under consideration is the following:

$$\mathbf{Y} = \mathbf{M}\,\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}^{1/2} + \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}. \tag{3}$$

Let $\mathbf{Y}$ be written as $\mathbf{Y} = [\mathbf{y}_0, ..., \mathbf{y}_{N-1}]$ where $\mathbf{y}_j = (y_{0,j}, y_{1,j}, \ldots, y_{m-1,j})^T$, $j \in [\![0, N-1]\!]$ and let $\widetilde{\mathbf{C}}_{SCM}$ be a biased Toeplitz estimation of the covariance matrix $\mathbf{C}$ such that:

$$\left[\widetilde{\mathbf{C}}_{SCM}\right]_{i,j} = [\mathcal{L}(\tilde{\mathbf{c}}_{SCM})]_{i,j} \tag{4}$$

with $\tilde{c}_{SCM,k} = \frac{1}{mN}\sum_{i=0}^{m-1}\sum_{j=0}^{N-1} y_{i,j}\, y_{i+k,j}^\star\, \mathbb{1}_{0\le i+k<m}$

It can be equivalently stated as $\widehat{\mathbf{C}}_{SCM} = \frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H$ and $\widetilde{\mathbf{C}}_{SCM} = \mathcal{T}(\widehat{\mathbf{C}}_{SCM})$. The following theorem establishes the consistency of $\widetilde{\mathbf{C}}_{SCM}$.

Theorem 1 (Consistency of $\widetilde{\mathbf{C}}_{SCM}$). *Under above assumptions, one has the following result:*

$$\left\|\widetilde{\mathbf{C}}_{SCM} - \mathbb{E}[\tau]\,\mathbf{C}\right\| \xrightarrow{a.s.} 0. \tag{5}$$

The covariance matrix defined by $\check{\mathbf{C}}_{SCM} = \frac{1}{\mathbb{E}(\tau)}\widetilde{\mathbf{C}}_{SCM}$ characterizes the biased Toeplitz estimation of $\mathbf{C}$.

Proof: The complete proof is in Appendix A. ∎

This estimator is then used to whiten the samples:

$$\check{\mathbf{Y}}_{wSCM} = \check{\mathbf{C}}_{SCM}^{-1/2}\mathbf{M}\,\boldsymbol{\delta}^H\boldsymbol{\Gamma}^{1/2} + \check{\mathbf{C}}_{SCM}^{-1/2}\mathbf{C}^{1/2}\mathbf{X}\,\mathbf{T}^{1/2}. \tag{6}$$

In practice, $\mathbb{E}(\tau)$ can be empirically estimated or is supposed to be equal to 1.

B. Estimation of the covariance matrix

Once the signal $\mathbf{Y}$ has been whitened, a robust estimation of the (unobservable) covariance matrix $\mathbb{E}\left[\mathbf{X}\mathbf{X}^H\right]$ can be performed through the samples $\check{\mathbf{Y}}_{wSCM}$. This estimation is said to be robust in the sense that it can annihilate the high values of the texture τ, which can alter the structure quality of the estimated covariance matrix. The chosen estimator is a Maronna's M-estimator [30], which gives good performances for CES signals. This robust estimation of the scatter matrix is therefore a fixed-point estimator noted $\check{\boldsymbol{\Sigma}}_{SCM}$ and defined through $\check{\mathbf{Y}}_{wSCM} = [\check{\mathbf{y}}_{wS0}, ..., \check{\mathbf{y}}_{wS\,N-1}]$ as the unique solution of the following equation:

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\check{\mathbf{y}}_{wSi}^H\boldsymbol{\Sigma}^{-1}\check{\mathbf{y}}_{wSi}\right)\check{\mathbf{y}}_{wSi}\check{\mathbf{y}}_{wSi}^H, \tag{7}$$

under
(i) u: $[0, +\infty) \mapsto (0, +\infty)$ nonnegative, continuous et non-increasing function derived thanks to the probability distribution function of the CES (for the complete calculus, see [39]),
(ii) $\phi : x \mapsto x\,u(x)$ increasing and bounded, with $\lim_{x\to\infty}\phi(x) = \Phi_\infty > 1$,

(iii) $\lim_{N\to\infty} c_N < \Phi_\infty^{-1}$.

Next step consists in evaluating the rank of the signal subspace from this matrix.

C. Model order selection

The mean idea is to study the eigenvalues distribution of this Maronna M-estimator to find the model order or the number of sources. Indeed, in a non-RMT regime, that is if Assumption 1 is not satisfied, and in the case of a white Gaussian noise, it is possible to set a threshold such that no eigenvalues of the noise can be found upon. If eigenvalues are found beyond this threshold, they are due to sources. Here, under Assumption 1 and thanks to [11] in the case of a white Gaussian noise plus an additive signal, no eigenvalues outside the support of the Marchenko-Pastur law can belong to the noise. However, due to the presence of the texture matrix $\mathbf{T}$, some eigenvalues could exist upon the right edge of the Marchenko-Pastur distribution support. A more precise threshold can then be derived to ensure that no eigenvalue found upon are due to the noise. However, it does not ensure that all the sources eigenvalues will be located beyond this threshold. Indeed, this depends of the sources Signal to Noise Ratio (SNR).

The proposed estimator $\check{\boldsymbol{\Sigma}}_{SCM}$ has so to be analysed for CES distribution. However, some characteristics such as its eigenvalues distribution can not be easily and theoretically studied when both m and $N \to \infty$ as the term $u\left(\frac{1}{m}\check{\mathbf{y}}_{wSi}^H\check{\boldsymbol{\Sigma}}_{SCM}^{-1}\check{\mathbf{y}}_{wSi}\right)$ is not independent on $\check{\mathbf{y}}_{wSi}$. To fill this gap, the following white model [27] is considered:

$$\mathbf{Y}_w = [\mathbf{y}_{w0}, \ldots, \mathbf{y}_{w\,N-1}] = \mathbf{C}^{-1/2}\mathbf{M}\,\boldsymbol{\delta}^H\boldsymbol{\Gamma}^{1/2} + \mathbf{X}\,\mathbf{T}^{1/2}. \tag{8}$$

Notice that the difference between models (8) and (6) lies in the empirical whitening. Then,

$$\hat{\mathbf{S}} \triangleq \frac{1}{N}\sum_{i=0}^{N-1} v\left(\tau_i\,\gamma\right)\mathbf{y}_{wi}\mathbf{y}_{wi}^H, \tag{9}$$

which can be rewritten as $\hat{\mathbf{S}} = \mathbf{Y}_w\mathbf{D}_\nu\mathbf{Y}_w^H$ where $\mathbf{D}_\nu$ a diagonal matrix containing the $\{v(\tau_i\,\gamma)\}_i$, where:
(i) $g: x \mapsto \dfrac{x}{1 - c\,\phi(x)}$,
(ii) $v: x \mapsto u \circ g^{-1}(x)$, $\psi: x \mapsto x\,v(x)$, with $\lim_{x\to\infty}\psi(x) = \dfrac{\Phi_\infty}{1 - c\,\Phi_\infty}$,
(iii) γ is the unique solution, if defined, of the equation in γ:
$$1 = \frac{1}{N}\sum_{i=1}^{N}\frac{\psi\left(\tau_i\,\gamma\right)}{1 + c\,\psi(\tau_i\,\gamma)}.$$
Moreover, it is proved in [27] that:
$$\left\|\hat{\boldsymbol{\Sigma}} - \hat{\mathbf{S}}\right\| \xrightarrow{a.s.} 0. \tag{10}$$
where $\hat{\boldsymbol{\Sigma}}$ is the unique solution (if it exists) of:
$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\mathbf{y}_{wi}^H\boldsymbol{\Sigma}^{-1}\mathbf{y}_{wi}\right)\mathbf{y}_{wi}\mathbf{y}_{wi}^H.$$

The distribution of the eigenvalues of $\hat{\mathbf{S}}$ can hence be more efficiently studied, the terms $[v(\tau_i\,\gamma)]_{i\in[\![0,\,N-1]\!]}$ being independent of the $\{\mathbf{x}_i\}_i$. The goal being to study $\check{\mathbf{\Sigma}}_{SCM}$ which is the unique solution of (7), the following theorem enables to establish the relationship between $\check{\mathbf{\Sigma}}_{SCM}$ and $\hat{\mathbf{S}}$ thanks to (10).

Theorem 2 (Convergence of $\check{\mathbf{\Sigma}}_{SCM}$).

With previous definitions, one has the following convergence:

$$\left\|\check{\mathbf{\Sigma}}_{SCM}-\hat{\mathbf{S}}\right\| \xrightarrow{a.s.} 0\,. \tag{11}$$

Proof: The proof is provided in Appendix B. ∎

As the eigenvalues distribution of $\hat{\mathbf{S}}$ can be theoretically analysed when $N,\,m\to\infty$, it can characterize also those of $\check{\mathbf{\Sigma}}_{SCM}$ thanks to (11). Under the hypothesis that there is no source present in the signal, it is possible to set a threshold similarly to [27]. Indeed, in this case:

$$\left\|\hat{\mathbf{S}}\right\| = \left\|\frac{1}{N}\sum_{i=0}^{N-1}\tau_i\,v(\tau_i\,\gamma)\,\mathbf{x}_i\,\mathbf{x}_i^H\right\| = \left\|\frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{\gamma}\,\psi(\tau_i\,\gamma)\,\mathbf{x}_i\,\mathbf{x}_i^H\right\|\,,$$

$$\leq \frac{\Phi_\infty}{\gamma\,(1-c\,\Phi_\infty)}\left\|\frac{1}{N}\sum_{i=0}^{N-1}\mathbf{x}_i\,\mathbf{x}_i^H\right\|\,.$$

Thanks to [11], and the bounds of the Marchenko Pastur distribution support, this inequality becomes

$$\left\|\hat{\mathbf{S}}\right\| \leq t\,, \tag{12}$$

where the threshold t is defined for the covariance matrix $\check{\mathbf{\Sigma}}_{SCM}$ by:

$$t = \frac{\Phi_\infty\,(1+\sqrt{c})^2}{\gamma\,(1-c\,\Phi_\infty)}\,. \tag{13}$$

Then, if the signal contains sources of sufficiently high SNR, eigenvalues might be found upon this threshold t and all these eigenvalues correspond to sources. Let $\left\{\lambda_i(\check{\mathbf{\Sigma}}_{SCM})\right\}_{i\in[\![1,N]\!]}$ be the sorted eigenvalues of $\check{\mathbf{\Sigma}}_{SCM}$ when sources are present in the samples. As all sources are assumed to be independent, the estimated number of sources $\hat{p}$ which corresponds to the rank of the signal subspace is then given by $\hat{p} = \min_{k}(\lambda_k > t)$, if $p << \min(N,m)$.

D. Results

This section is devoted to the presentation of some simulations relative to the estimation of the covariance matrix. Samples are considered here sources-free. The parameters are set to $c=0.45$, $m=900$ and $N=2000$. Thus, $\mathbf{Y} = \mathbf{C}^{1/2}\mathbf{X}\mathbf{T}^{1/2}$ with $\mathbf{C} = \mathcal{L}\left((\rho^0,\rho^1,\dots,\rho^{m-1})^T\right)$ where $\rho=0.7$ and $\mathbf{X}$ is a zero-mean complex Gaussian noise with identity covariance matrix. The texture matrix $\mathbf{T}$ is a diagonal $N\times N$-matrix containing the $\{\tau_i\}_{i\in[\![0,N-1]\!]}$ on its diagonal where $\{\tau_i\}_i$ are i.i.d. inverse gamma distributed with mean equal to 1 and with shape parameter equal to 10. The function u is here defined as $u:x\mapsto \dfrac{1+\alpha}{x+\alpha}$ where α is a fixed parameter equal to 0.1.

Figure 1 shows the eigenvalues of the estimated covariance matrix $\check{\mathbf{\Sigma}}_{SCM}$ when samples $\mathbf{Y}$ have been whitened by $\check{\mathbf{C}}_{SCM}^{-1/2}$. On the figure 2, the signal $\mathbf{Y}$ has not been
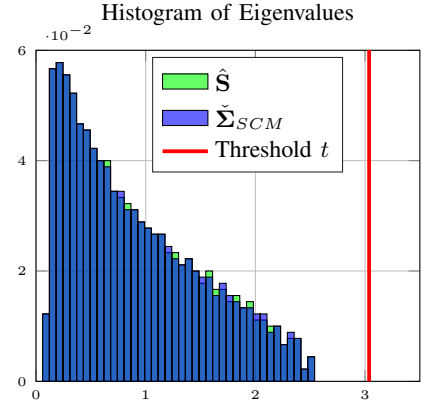

Fig. 1. Eigenvalues of the covariance matrices $\check{\mathbf{\Sigma}}_{SCM}$ and $\hat{\mathbf{S}}$ when the signal $\mathbf{Y}$ has been whitened by $\check{\mathbf{C}}_{SCM}$ and the corresponding threshold t ($\rho=0.7$, $m=900$, $N=2000$, $\tau=$ inverse gamma, $\alpha=0.1$).

Fig. 2. Eigenvalues of the covariance matrices $\check{\mathbf{\Sigma}}_{SCM}$ and $\hat{\mathbf{S}}$ when the signal $\mathbf{Y}$ has not been whitened by $\check{\mathbf{C}}_{SCM}$ and the corresponding threshold t ($\rho=0.7$, $m=900$, $N=2000$, $\tau=$ inverse gamma, $\alpha=0.1$).
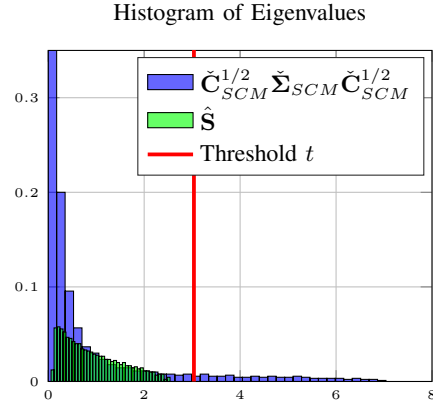
whitened. The green histogram corresponds to the eigenvalues distribution of $\hat{\mathbf{S}}$ whose histogram is expected to coincide with the distribution of the eigenvalues of $\check{\mathbf{\Sigma}}_{SCM}$ as the equation (11) indicates. Moreover, the threshold $t = \dfrac{(1+\alpha)(1+\sqrt{c})^2}{\gamma\,(1-c\,(1+\alpha))}$ given by (13) has been estimated and drawn in red, in order to confirm that the eigenvalues are all smallest than the threshold.

As the eigenvalues distribution of $\check{\mathbf{\Sigma}}_{SCM}$ are closed to those of $\hat{\mathbf{S}}$, the fixed-point estimator correctly annihilates the influence of the textures τ_i's and the whitening balances the matrix of correlation. On Figure 1, we can observe that the eigenvalues do not exceed the upper bound t. When the signal has not been whitened, this threshold t does not theoretically correspond. Indeed, in Figure 2, the threshold is found to be smaller than the largest eigenvalues of the estimated covariance matrix. These figures illustrate first the results of Theorem 2 and show the importance of the whitening process.

Figure 3 presents the eigenvalues distributions of $\hat{\mathbf{S}}$ and $\check{\mathbf{C}}_{SCM}$ for samples distributed according to a different CES distribution. Here, the texture $\mathbf{T}$ is a diagonal matrix containing the $\{\tau_i\}_{i\in[\![0,,N-1]\!]}$ on its diagonal where each τ_i is independent and identically distributed and follows a distribution

equal to t^2 where t is a Student-t distributed random variable with parameter 100 and $\alpha = 0.1$. The eigenvalues are not so close than the eigenvalues of $\hat{\mathbf{S}}$ and are found to get closer to the threshold t. If the distribution of τ is getting away to the one for which the function u has been calculated, the method seems so to be less reliable. To fill this gap, we propose to enhance the proposed SCM-based method for the whitening through robust M-estimators-based method.
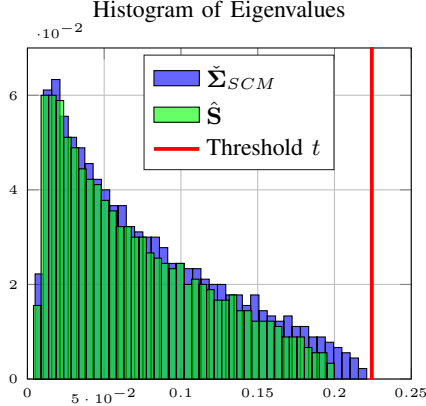

Fig. 3. Eigenvalues of the covariance matrices $\check{\mathbf{\Sigma}}_{SCM}$ and $\hat{\mathbf{S}}$ when the signal is whitened by $\check{\mathbf{C}}_{SCM}$ and the calculated threshold ($\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2$, $\alpha = 0.1$).

IV. MODEL ORDER SELECTION: A ROBUST METHOD APPROACH

This section aims at developing a robust estimator based technique to whiten the signal instead of the previous SCM-based one. This section follows the same steps than in the previous section but by using a M-estimator in the whitening process.

A. Whitening Step

Let $\widetilde{\mathbf{C}}_{FP}$ be an biased estimator of the covariance matrix $\mathbf{C}$ such that $\widetilde{\mathbf{C}}_{FP} = \mathcal{T}(\widehat{\mathbf{C}}_{FP})$ where $\widehat{\mathbf{C}}_{FP}$ is the unique solution to the Maronna's M-estimator [30]:

$$\mathbf{Z} = \frac{1}{N} \sum_{i=0}^{N-1} u\left(\frac{1}{m} \mathbf{y}_i^H \mathbf{Z}^{-1} \mathbf{y}_i\right) \mathbf{y}_i \mathbf{y}_i^H .$$

As in the previous section, $u(.)$ is a function derived thanks to the probability distribution function of the CES noise: u: $[0, +\infty) \mapsto (0, +\infty)$ nonnegative, continuous and non-increasing. The following theorem stands for $\widetilde{\mathbf{C}}_{FP}$:

Theorem 3 (Consistency of $\widetilde{\mathbf{C}}_{FP}$). *Let $\widetilde{\mathbf{C}}_{FP}$ be a fixed-point estimator of the covariance matrix $\mathbf{C}$ as defined above, the following result holds:*

$$\left\|\widetilde{\mathbf{C}}_{FP} - \mathbb{E}\left[v(\tau\,\gamma)\,\tau\right] \mathbf{C}\right\| \xrightarrow{a.s.} 0 , \tag{14}$$

where:

i) $\phi : x \mapsto x\,u(x)$ *increasing and bounded, with* $\lim_{x\to\infty} \phi(x) = \Phi_\infty > 1$,

ii) $\lim_{N\to\infty} c_N < \Phi_\infty^{-1}$,

iii) $g : x \mapsto \dfrac{x}{1 - c\,\phi(x)}$,

iv) $v : x \mapsto u \circ g^{-1}(x)$, $\psi : x \mapsto x\,v(x)$,

v) γ *is the unique solution, if defined, of:* $1 = \dfrac{1}{N} \sum_{i=1}^{N} \dfrac{\psi\,(\tau_i\,\gamma)}{1 + c\,\psi(\tau_i\,\gamma)}$.

The covariance matrix $\check{\mathbf{C}}_{FP} = \dfrac{\widetilde{\mathbf{C}}_{FP}}{\mathbb{E}\left[v(\tau\,\gamma)\,\tau\right]}$ characterizes the estimator of the true covariance matrix $\mathbf{C}$.

Proof: The proof, inspired by [31] and [40], is provided in Appendix B. ∎

Remark: [17] proves that $\tilde{\mathbf{C}}_{SCM} = \phi^{-1}(1)\,\tilde{\mathbf{C}}_{FP}$. When the function u is well chosen, it is possible to have $\phi^{-1}(1) = 1$ and $\tilde{\mathbf{C}}_{SCM} = \tilde{\mathbf{C}}_{FP}$, as it will be the case for the u chosen in the following sections. But even in this case, $\check{\mathbf{C}}_{SCM}$ and $\check{\mathbf{C}}_{FP}$ differ up to a scale factor as $\check{\mathbf{C}}_{SCM} = \dfrac{\mathbb{E}\left[v(\tau\gamma)\,\tau\right]}{\mathbb{E}(\tau)}\,\check{\mathbf{C}}_{FP}$.

As in the previous section, the samples $\mathbf{Y}$ can then be whitened thanks to $\check{\mathbf{C}}_{FP}^{-1/2}$. Let $\check{\mathbf{Y}}_{wFP} = [\check{\mathbf{y}}_{wF0}, ...\check{\mathbf{y}}_{wF\,N-1}]$ be the whitened samples:

$$\check{\mathbf{Y}}_{wFP} = \check{\mathbf{C}}_{FP}^{-1/2}\,\mathbf{M}\,\boldsymbol{\delta}^H\,\mathbf{\Gamma}^{1/2} + \check{\mathbf{C}}_{FP}^{-1/2}\,\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2} .$$

The parameter $\mathbb{E}[\tau]$ can be in practice evaluated with the empirical estimator of the mean, or, as in the previous section, be considered as equal to one. The quantity $\mathbb{E}[v(\tau\,\gamma)\,\tau]$ can be also evaluated through an estimate $\hat{\gamma}$ of γ as explained in the Results section.

B. Robust estimation of the covariance matrix and model order selection

The robust estimation of the covariance matrix and the model order selection are done as previously. The robust estimator of the scatter matrix of the whitened signal $\check{\mathbf{Y}}_{wFP}$ is a fixed-point estimator denoted by $\check{\mathbf{\Sigma}}_{FP}$ and defined as the unique solution of the equation:

$$\mathbf{\Sigma} = \frac{1}{N} \sum_{i=0}^{N-1} u\left(\frac{1}{m} \check{\mathbf{y}}_{wFi}^H \mathbf{\Sigma}^{-1} \check{\mathbf{y}}_{wFi}\right) \check{\mathbf{y}}_{wFi} \check{\mathbf{y}}_{wFi}^H . \tag{15}$$

Thus, $\check{\mathbf{\Sigma}}_{FP}$ is a robust estimator of the covariance matrix of the whitened signal.

The equation (11) is still effective when replacing $\check{\mathbf{\Sigma}}_{SCM}$ by $\check{\mathbf{\Sigma}}_{FP}$. Indeed, Theorem 2 can be adapted as follows:

Theorem 4. *The following convergence holds:*

$$\left\|\check{\mathbf{\Sigma}}_{FP} - \hat{\mathbf{S}}\right\| \xrightarrow{a.s.} 0 . \tag{16}$$

Proof: The proof is the same as in Theorem 2 and is provided in Appendix B. ∎

The same threshold t given by equation (13) can be used on the eigenvalues of $\check{\mathbf{\Sigma}}_{FP}$ to estimate p. The final corresponding algorithm is presented below.
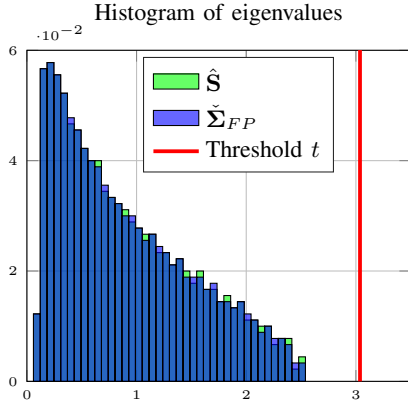
Fig. 4. Eigenvalues of the covariance matrices $\hat{\mathbf{S}}$ and $\check{\boldsymbol{\Sigma}}_{FP}$ when the signal is whitened through $\check{\mathbf{C}}_{FP}$ and the corresponding threshold t ($\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma).
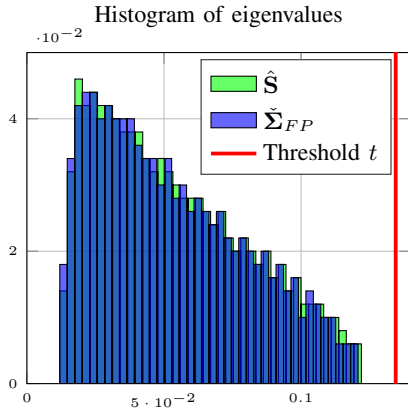

Fig. 5. Eigenvalues of the covariance matrices $\hat{\mathbf{S}}$ and $\check{\boldsymbol{\Sigma}}_{FP}$ when the signal is whitened through $\check{\mathbf{C}}_{FP}$ and the corresponding threshold ($\rho = 0.7$, $m = 900$, $N = 2000$, $\tau = t^2$, $t \sim$ student)

C. Results

As in the previous section, it seems interesting to analyse the eigenvalues distributions of $\hat{\mathbf{S}}$ and $\check{\boldsymbol{\Sigma}}_{FP}$. For the next simulations, source-free samples are considered and the parameters are set to $c = 0.45$, $m = 900$ and $N = 2000$. The function u chosen for the FP and Maronna M-estimators is the same function as before with $\alpha = 0.1$.

Figure 4 presents the eigenvalues distribution of the covariance matrices $\hat{\mathbf{S}}$ and $\check{\boldsymbol{\Sigma}}_{FP}$ when the signal has been whitened by $\check{\mathbf{C}}_{FP}$. One can notice that the results are the same as Figure 1: for N large, the distribution of eigenvalues is almost the same as those of $\hat{\mathbf{S}}$. However, as the rate of convergence of (14) is faster than in (5), it is more interesting to consider the robust method. Moreover, if a robust estimator is not used after the whitening process, the eigenvalues distribution will not follow those of $\hat{\mathbf{S}}$ and will exceed the threshold t.

For robustness analysis (not the same texture distribution for the u function and the observed samples), Figure 5 shows quite good results when $\mathbf{T}$ is a diagonal matrix containing the $\{\tau_i\}_{i \in [\![0, N-1]\!]}$ on its diagonal where $\{\tau_i\}_{i \in [\![0, N-1]\!]}$ are i.i.d. and follow a distribution equal to t^2 with t a Student-t random variable with parameter 100 and $\alpha = 0.1$.

Figure 6 presents the same histograms as in Figure 4 for a single source of SNR equal to 10 dB present in the samples. One can observe that only single eigenvalue exceeds the threshold and that the noise eigenvalues distribution of $\check{\boldsymbol{\Sigma}}_{FP}$ fits well those of $\hat{\mathbf{S}}$.
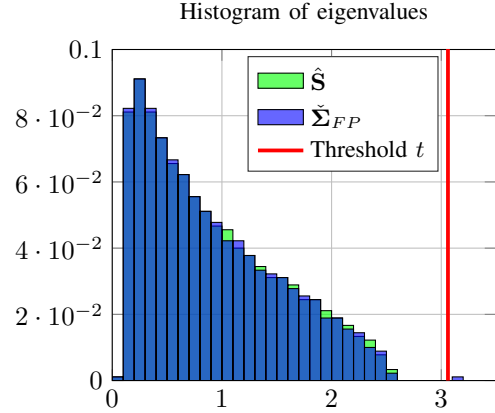

Fig. 6. Eigenvalues of the covariance matrices $\hat{\mathbf{S}}$ and $\check{\boldsymbol{\Sigma}}_{FP}$ for a single source with $SNR = 10dB$ present in the samples and the calculated threshold $\rho = 0.7$, $m = 900$, $N = 2000$, $\tau \sim$ inverse gamma

The results are better for this robust method than in the previous section (e.g. figure 3). Indeed, the robust method provides robustness with respect to the distribution of τ: if the distribution of the texture differs to those for which the function u has been computed, the method is still reliable, this can be explained by the robustness of the covariance matrix estimation. As for the non-whitening case, the eigenvalues get over the threshold and no conclusion or model order can be deducted. These results have so extended the paper of [31] to the left hand side correlated noise case. The L2-norm of the estimated covariance matrix compared to the SCM tends to zero when N and m tends to infinity with a constant ratio c. As a lot of estimation methods for the rank of the signal subspace are based on the estimation of the eigenvalues of the covariance matrix, this new estimator improves the consistency for resolution of this problem.

V. RESULTS AND COMPARISONS

In this section some results of order selection are presented, on both simulated and real hyperspectral images. The simulations are based on $\check{\boldsymbol{\Sigma}}_{SCM}$ and $\check{\boldsymbol{\Sigma}}_{FP}$.

A. Estimation of the model order

In order to test the proposed method, we simulate hyperspectral images, before dealing with real images. As a reminder, we first whiten the received signal thanks to a Toeplitz matrix coming from the SCM or a Fixed-Point estimator. Thus, a M-estimator is used to estimate the scatter matrix of the whitened signal. The distribution of its eigenvalues is then studied: a threshold is applied to count how many eigenvalues are higher than this threshold, providing the estimated model order $\hat{p}$.
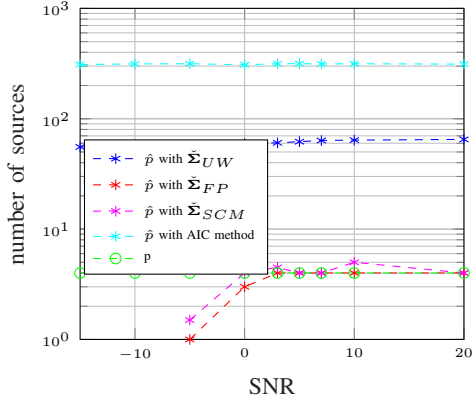
Fig. 7. Estimation of the number $\hat{p}$ of sources (4 trials) embedded in CES correlated noise ($m = 400$, $c = 0.2$, $p = 4$ source, $\rho = 0.7$) versus SNR.
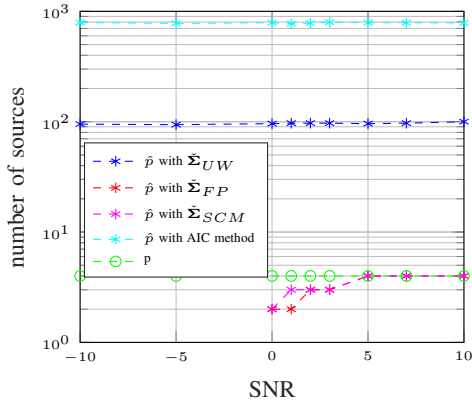

Fig. 8. Estimation of the number p of sources (4 trials) embedded in CES correlated noise ($m = 400$, $c = 0.2$, $p = 4$ sources, $\rho = 0.7$) versus SNR.

For simulated and correlated ($\rho = 0.7$) CES noise, the $\{\tau_i\}_{i \in [\![0, N-1]\!]}$ are inverse gamma distributed with parameter $\nu = 0.1$. On Figure 7 ($m = 400$ and $N = 2000$), $p = 4$ sources are added in the observations with a SNR varying from -15 to 20dB. For this figure, the number of sources $\hat{p}$ (average on 4 trials) is estimated through three methods: AIC, the non-whitened signal and the two proposed methods: when the signal is whitened with the Toeplitz version of the SCM and the one of the FP. The proposed method starts to find sources from a SNR equal to -5dB. The FP method seems to better evaluate the number of sources. For a greater SNR, whereas it systematically gives the correct number of sources, the other methods overestimate it. On Figure 8 the same simulation is done for $p = 4$ but with the $\{\tau_i\}_i$ following a distribution equal to t^2 where t is a Student-t random variable, as before. On Figure 8, one notice that the proposed estimators still present better performance than the others, and allow to find sources with SNR greater than 0 dB.

Now, we compare the results obtained with three different methods on several real hyperspectral images found in public access: *Indian Pines*, *SalinasA* from AVIRIS database and *PaviaU* from ROSIS database [41]. Let $M1$ be the proposed method with a whitening made with the SCM estimator, $M2$ be the proposed method with a whitening made with a Fixed-Point estimator, $M3$ be the method consisting in thresholding the eigenvalues of the Fixed-Point estimator without the whitening step, and the usual AIC method. For the function $u(.)$ corresponding to the Student-t distribution, we choose $\nu = 0.1$ for the whitening process if it is done by a fixed-point estimator, and zero for the estimation process. As we do not have any access to the true distribution of the noise, an empirical estimator of γ is used, $\hat{\gamma} = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} \dfrac{1}{m} \mathbf{y}_i^H \, \check{\boldsymbol{\Sigma}}_{(i)}^{-1} \, \mathbf{y}_i$,

where $\check{\boldsymbol{\Sigma}}_{(i)} = \check{\boldsymbol{\Sigma}} - \dfrac{1}{N} u\left(\dfrac{1}{m} \mathbf{y}_i^H \, \check{\boldsymbol{\Sigma}}^{-1} \, \mathbf{y}_i\right) \mathbf{y}_i \, \mathbf{y}_i^H$. Then [27] shows that $\gamma - \hat{\gamma} \xrightarrow{a.s.} 0$. Moreover, as the distribution of τ is unknown, we choose to consider that $\mathbb{E}\left[\tau\right]$ and $\mathbb{E}\left[v(\tau \gamma)\tau\right]$ are equal to 1. Further works can be carried out to estimate correctly these unknown quantities. However, we can reasonably assume than $\mathbb{E}\left[v(\tau \gamma)\tau\right]$ and $\mathbb{E}\left[\tau\right]$ are not to large and that the estimation error will not impact the results a lot. The results are summarized in table I. On each image, the result tends to be better than those of classical methods.

TABLE I
ESTIMATED p FOR DIFFERENT HYPERSPECTRAL IMAGES.

Images	Indian Pines	SalinasA	PaviaU	Cars
p	16	9	9	6
$\hat{p}$ M1	11	9	1	3
$\hat{p}$ M2	12	9	1	3
$\hat{p}$ M3	220	204	103	1
$\hat{p}$ AIC	219	203	102	143

VI. CONCLUSION

The model order selection for large dimensional data and for sources embedded in correlated CES noise is tackled in this article. Two Toeplitz-based covariance matrix estimators are first introduced, and their consistency has been proved. As for the CES texture, it is handled with any M-estimator, which can then be used to estimate the correct structure of the scatter matrix built on whitened observations. The Random Matrix Theory provides tools to correctly estimate the model order. Results obtained on real and simulated hyperspectral images are promising. Moreover, the proposed method can be applied on a lot of other kind of model order selection problems such as radar clutter rank estimation, sources localization or any hyperspectral problems such as anomaly detection or linear or non-linear unmixing techniques.

APPENDIX A
PROOFS OF THEOREM 1 AND THEOREM 3

The proofs of Theorem 1 and Theorem 3 are inspired by [31]. For these theorems, we will use the lemma 4.1 in [42], that is, for $\mathbf{T} = \mathcal{L}\left((t_0, \ldots, t_{m-1})^T\right)$ a Toeplitz Hermitian $m \times m$-matrix with $\{t_k\}_{k \in [\![0, m-1]\!]}$ absolutely summable ($t_{-k} = t_k^*$), we can define the function $f(.)$ such that for any $\lambda \in [0, 2\pi]$, $f(\lambda) = \displaystyle\sum_{k=1-m}^{m-1} t_k \, e^{i\lambda k}$ and M_f characterizes its essential supremum:

$$\|\mathbf{T}\| \leq M_f = \sup_{\lambda \in [0, 2\pi)} \left| \sum_{k=1-m}^{m-1} t_k \, e^{i\lambda k} \right|. \qquad (17)$$

A. Proof of Theorem 1

As in the main body of this article, let $\mathbf{Y} = \mathbf{M}\,\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}^{1/2} + \mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$ and let $\mathcal{T}$ be the Toeplitz operator as defined in the introduction and for any m-vector $\mathbf{x}$, $\left([\mathcal{L}(\mathbf{x})]_{i,j}\right)_{i \leq j} = x_{i-j}$ and $\left([\mathcal{L}(\mathbf{x})]_{i,j}\right)_{i > j} = x_{i-j}^*$, of size $m \times m$. Under Assumption 1, Assumption 2, Assumption 3 and as $\mathcal{T}\left(\frac{1}{N}\,\mathbf{Y}\,\mathbf{Y}^H\right)$ and $\mathbb{E}[\tau]\,\mathbf{C}$ are Toepltiz matrices, one can write, thanks to (17):

$$\left\| \mathcal{T}\left(\frac{1}{N}\,\mathbf{Y}\,\mathbf{Y}^H\right) - \mathbb{E}[\tau]\,\mathbf{C}\right\| \leq \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}_m(\lambda) - \mathbb{E}[\tau]\,\gamma_m(\lambda)\right|, \tag{18}$$

where $\gamma_m(\lambda) = \displaystyle\sum_{k=1-m}^{m-1} c_{k,m}\,e^{i\,k\,\lambda}$ with $c_{-k} = c_k^\star$ and $\hat{\gamma}_m(\lambda) = \displaystyle\sum_{k=1-m}^{m-1} \check{c}_{k,m}\,e^{i\,k\,\lambda}$ with $\check{c}_{-k} = \check{c}_k^\star$.

The following lemma is essential for the development of the proof:

Lemma 5. *The quantity $\hat{\gamma}_m(\lambda)$ can be rewritten as:*

$$\hat{\gamma}_m(\lambda) = \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda), \tag{19}$$

with $\mathbf{d}_m(\lambda) = \dfrac{1}{\sqrt{m}}\left(1, e^{-i\,\lambda}, \ldots, e^{-i\,(m-1)\,\lambda}\right)^T$.

Proof: The proof draws his inspiration from the one of Appendix A1 in [31]. Equation (19) can be rewritten as:

$$\mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda) = \frac{1}{m\,N}\sum_{l,l'=0}^{m-1} e^{-i\,(l'-l)\,\lambda}\,\left[\mathbf{Y}\,\mathbf{Y}^H\right]_{l,l'},$$

$$= \sum_{k=1-m}^{m-1} e^{-i\,k\,\lambda}\,\frac{1}{m\,N}\sum_{i=0}^{m-1}\sum_{j=0}^{N-1} y_{i,j}\,y_{i+k,j}^\star \mathbb{1}_{0 \leq i+k \leq m},$$

$$= \sum_{k=1-m}^{m-1} \check{c}_k\,e^{-i\,k\,\lambda}.$$

Thereby, we have:

$$\hat{\gamma}_m(\lambda) = \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{Y}\,\mathbf{Y}^H}{N}\,\mathbf{d}_m(\lambda),$$

$$= \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{M}\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}\,\boldsymbol{\delta}\,\mathbf{M}^H}{N}\,\mathbf{d}_m(\lambda)$$

$$+ \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,\boldsymbol{\Gamma}^{1/2}\,\boldsymbol{\delta}\,\mathbf{M}^H}{N}\,\mathbf{d}_m(\lambda)$$

$$+ \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{M}\,\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}^{1/2}\,\mathbf{T}^{1/2}\,\mathbf{X}^H\,\mathbf{C}^{1/2}}{N}\,\mathbf{d}_m(\lambda)$$

$$+ \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}\,\mathbf{X}^H\,\mathbf{C}^{1/2}}{N}\,\mathbf{d}_m(\lambda). \tag{20}$$

And we note : $\hat{\gamma}_m^{sign}(\lambda) = \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{M}\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}\,\boldsymbol{\delta}\,\mathbf{M}^H}{N}\,\mathbf{d}_m(\lambda)$,

$$\hat{\gamma}_m^{cross}(\lambda) = \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}\,\boldsymbol{\Gamma}^{1/2}\,\boldsymbol{\delta}\,\mathbf{M}^H + \mathbf{M}\,\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}^{1/2}\,\mathbf{T}^{1/2}\,\mathbf{X}^H\,\mathbf{C}^{1/2}}{N}\,\mathbf{d}_m(\lambda)$$

$$\hat{\gamma}_m^{noise}(\lambda) = \mathbf{d}_m^H(\lambda)\,\frac{\mathbf{C}^{1/2}\,\mathbf{X}\,\mathbf{T}\,\mathbf{X}^H\,\mathbf{C}^{1/2}}{N}\,\mathbf{d}_m(\lambda).$$

And the equation (18) becomes:

$$\left\| \mathcal{T}\left(\frac{1}{N}\,\mathbf{Y}\,\mathbf{Y}^H\right) - \mathbb{E}[\tau]\,\mathbf{C}\right\| \leq \sup_{\lambda \in [0,2\pi)} |\hat{\gamma}_m^{noise}(\lambda)$$

$$+ \hat{\gamma}_m^{sign}(\lambda) + \hat{\gamma}_m^{cross}(\lambda) - \mathbb{E}[\tau]\,\gamma_m(\lambda)|. \tag{21}$$

This leads to:

$$\left\| \mathcal{T}\left(\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right) - \mathbb{E}[\tau]\,\mathbf{C}\right\|$$

$$\leq \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]\right|$$

$$+ \sup_{\lambda \in [0,2\pi)} \left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] - \mathbb{E}(\tau)\,\gamma_m(\lambda)\right|$$

$$+ \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}_m^{sign}(\lambda)\right| + \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}_m^{cross}(\lambda)\right|. \tag{22}$$

We will now analyse each term of (22).

1) Analysis of $\displaystyle\sup_{\lambda \in [0,2\pi)} \left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] - \mathbb{E}(\tau)\,\gamma_m(\lambda)\right|$: We first need the following lemma:

Lemma 6.

$$\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \mathbb{E}[\tau]\,\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda) = \mathbb{E}[\tau]\,\gamma_m(\lambda). \tag{23}$$

Proof: The equation (20) gives $\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\dfrac{\mathbf{C}^{1/2}\mathbf{X}\,\mathbf{T}\,\mathbf{X}^H\mathbf{C}^{1/2}}{N}\right]\,\mathbf{d}_m(\lambda)$. Let $\mathbf{V} = \mathbf{C}^{1/2}\mathbf{X}$ and $(\mathbf{V})_{i,j} = v_{i,j}$. We obtain $\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\dfrac{\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H}{N}\right]\,\mathbf{d}_m(\lambda)$. As $\left(\mathbb{E}\left[\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H\right]\right)_{ij} = \displaystyle\sum_{k=1}^{N}\mathbb{E}[\tau]\,\mathbb{E}\left[v_{j,k}^*\,v_{ik}\right]$ and $c_{k'} = \mathbb{E}\left[v_{p,n}\,v_{p+k',n}^*\right]$, we have $\left(\mathbb{E}\left[\mathbf{V}\,\mathbf{T}\,\mathbf{V}^H\right]\right)_{ij} = \displaystyle\sum_{k=1}^{N}\mathbb{E}[\tau]\,c_{j-i} = N\,\mathbb{E}[\tau]\,c_{j-i}$. This leads to

$$\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] = \frac{N\,\mathbb{E}(\tau)}{N}\,\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda) = \mathbb{E}[\tau]\,\gamma_m(\lambda).$$

∎

Thereby, the second term of (22) leads to:

$$\sup_{\lambda \in [0,2\pi)} \left|\mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] - \mathbb{E}(\tau)\,\gamma_m(\lambda)\right|$$

$$= \sup_{\lambda \in [0,2\pi)} |\mathbb{E}(\tau)\,\gamma_m(\lambda) - \mathbb{E}(\tau)\,\gamma_m(\lambda)| = 0.$$

The second term is equal to zero.

2) Analysis of $\displaystyle\sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right]\right|$: As in [31], the method consists in proving for a $\lambda_i \in [0,2\pi)$ and a real $x > 0$ that $\mathbb{P}\left[\left|\hat{\gamma}_m^{noise}(\lambda_i) - \mathbb{E}\left[\gamma_m^{noise}(\lambda_i)\right]\right| > x\right] \to 0$. After that, it remains to prove that $\mathbb{P}\left[\displaystyle\sup_{\lambda \in [\lambda_i,\lambda_{i+1})}\left|\gamma_m^{noise}(\lambda) - \gamma_m^{noise}(\lambda_i)\right| > x\right] \to 0$ and that $\mathbb{P}\left[\displaystyle\sup_{\lambda \in [\lambda_i,\lambda_{i+1})}\left|\mathbb{E}\gamma_m^{noise}(\lambda_i) - \mathbb{E}\gamma_m^{noise}(\lambda)\right| > x\right] \to 0$.

Let $\lfloor \cdot \rfloor$ be the floor function, choosing a $\beta > 2$, $\mathcal{I} = [0, \dots, \lfloor N^\beta \rfloor - 1]$, $\lambda_i = 2\pi \frac{i}{\lfloor N^\beta \rfloor}$, $i \in \mathcal{I}$:

$$\sup_{\lambda \in [0,2\pi)} \left| \hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] \right|$$
$$\leqslant \max_{i \in \mathcal{I}} \sup_{\lambda \in [\lambda_i \lambda_{i+1}]} \left| \hat{\gamma}_m^{noise}(\lambda) - \hat{\gamma}_m^{noise}(\lambda_i) \right|$$
$$+ \max_{i \in \mathcal{I}} \left| \hat{\gamma}_m^{noise}(\lambda_i) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda_i)\right] \right|$$
$$+ \max_{i \in \mathcal{I}} \sup_{\lambda \in [\lambda_i \lambda_{i+1}]} \left| \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda_i)\right] - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] \right|,$$
$$\overset{\triangle}{=} \chi_1 + \chi_2 + \chi_3. \tag{24}$$

The idea of the proof in [31] is then to provide concentration inequalities for the term χ_1 and χ_2 (random terms) and a bound on χ_3. The only difference with [31] is the presence of the matrix $\mathbf{T}$ in $\hat{\gamma}_m^{noise}(\lambda)$ and the left side correlation of the noise. Let note $\|\boldsymbol{\gamma}\|_\infty$ the sup norm of the function $\boldsymbol{\gamma}$: $\lambda \longrightarrow \sum_{k=-\infty}^{\infty} c_k\, e^{-ik\lambda}$ for $\lambda \in [0\, 2\pi)$. The convergence of the first term χ_1 is proposed in the following lemma.

Lemma 7. *A constant $A > 0$ can be found*

such that, for any $x > 0$ and N large enough,

$$\mathbb{P}[\chi_1 > x] \leq \exp\left(-\frac{cN^2}{\|\mathbf{T}\|_\infty}\left(\frac{x N^{\beta-2}}{A\|\boldsymbol{\gamma}\|_\infty} - \log\left(\frac{x N^{\beta-2}}{A\|\boldsymbol{\gamma}\|_\infty}\right) - 1\right)\right).$$

Proof: As already mentioned, the proof is the same as in [31] except for two points: the presence of the matrix $\mathbf{T}$ and the left side correlation of the noise instead of right side in [31]. The inequality: $\|\mathbf{V}_N \mathbf{T} \mathbf{V}_N^H\| \leqslant \|\mathbf{T}\|_\infty \|\mathbf{V}_N \mathbf{V}_N^H\| \leq \|\mathbf{T}\|_\infty \|\mathbf{C}\| \|\mathbf{X}\mathbf{X}^H\|$ enables to write:

$$\left| \hat{\gamma}_m^{noise}(\lambda) - \hat{\gamma}_m^{noise}(\lambda_i) \right|$$
$$= \left| \mathbf{d}_m^H(\lambda) \frac{\mathbf{V}\mathbf{T}\mathbf{V}^H}{N} \mathbf{d}_m(\lambda) - \mathbf{d}_m^H(\lambda_i) \frac{\mathbf{V}\mathbf{T}\mathbf{V}^H}{N} \mathbf{d}_m(\lambda_i) \right|,$$
$$\leq \frac{2}{N} |\mathbf{d}_m(\lambda) - \mathbf{d}_m(\lambda_i)| \|\mathbf{C}\| \|\mathbf{T}\|_\infty \|\mathbf{X}\mathbf{X}^H\|.$$

And then the end of the proof is exactly the same as those of the Lemma 4 in [31] replacing c by $\frac{c}{\|\mathbf{T}\|_\infty}$ in the exponential. The left correlation is without consequences on the proof. ∎

The convergence of the second term χ_2 is proposed in the following lemma.

Lemma 8.

$$\mathbb{P}[\chi_2 > x] \leq 2N^\beta \exp\left(-\frac{cN}{\|\mathbf{T}\|_\infty}\left(\frac{x}{\|\boldsymbol{\gamma}\|_\infty} - \log\left(\frac{x}{\|\boldsymbol{\gamma}\|_\infty} + 1\right)\right)\right).$$

Proof: The proof is the same as those of the Lemma 5 in [31], with the $\frac{c}{\|\mathbf{T}\|_\infty}$ on the denominator. ∎

The convergence of the third term χ_3 is proposed in the following lemma.

Lemma 9.

$$\chi_3 \leq A\|\boldsymbol{\gamma}\|_\infty N^{-\beta+1}.$$

Proof: The proof is the same as those of the lemma 6 of [31], still with the $\frac{c}{\|\mathbf{T}\|_\infty}$ on the denominator. ∎

These inequalities proves that
$$\mathbb{P}\left[\sup_{\lambda \in [0,2\pi)} \left| \hat{\gamma}_m^{noise}(\lambda) - \mathbb{E}\left[\hat{\gamma}_m^{noise}(\lambda)\right] \right| > x\right] \xrightarrow{a.s.} 0 \text{ for}$$
any x positive real and with a e^{-N^2} rate of decrease.

3) Analysis of $\sup_{\lambda \in [0,2\pi)} |\hat{\gamma}_m^{cross}(\lambda)|$: To prove the convergence of the last term of (22), let us recall that

$$\hat{\gamma}_m^{cross}(\lambda) = \mathbf{d}_m^H(\lambda) \frac{\mathbf{C}^{1/2}\mathbf{X}\mathbf{T}^{1/2}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\delta}\mathbf{M}^H}{N} \mathbf{d}_m(\lambda)$$
$$+ \mathbf{d}_m^H(\lambda) \frac{\mathbf{M}\boldsymbol{\delta}^H\boldsymbol{\Gamma}^{1/2}\mathbf{T}^{1/2}\mathbf{X}^H\mathbf{C}^{1/2}}{N} \mathbf{d}_m(\lambda).$$

Let $\mathbb{I}_m$ be a $m \times m$ matrix containing 1 everywhere and $\mathbf{D}_m(\lambda)$ be the matrix containing the elements of $\mathbf{d}_m(\lambda)$ on its diagonal. It can be easily verified that, for any matrix $\mathbf{A}$, $\mathbf{d}_m^H(\lambda)\mathbf{A}\mathbf{d}_m(\lambda) = \text{Tr}\left(\mathbf{D}_m^H(\lambda)\mathbf{A}\mathbf{D}_m(\lambda)\mathbb{I}_m\right)$. We obtain:

$$\hat{\gamma}_m^{cross}(\lambda) =$$
$$2\,\mathcal{R}e\left[\frac{1}{N}\text{Tr}\left(\mathbf{X}\mathbf{T}^{1/2}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\delta}\mathbf{M}^H\mathbf{D}_m(\lambda)\mathbb{I}_m\mathbf{D}_m^H(\lambda)\mathbf{C}^{1/2}\right)\right].$$

For readability, let $\mathbf{E}(\lambda) = \mathbf{D}_m(\lambda)\mathbb{I}_m\mathbf{D}_m^H(\lambda)$ defined as:

$$\mathbf{E}(\lambda) = \begin{pmatrix} 1 & e^{i\lambda} \dots e^{i(m-1)\lambda} \\ e^{-i\lambda} & 1 \dots e^{i(m-2)\lambda} \\ e^{-i(m-1)\lambda} & \dots\dots 1 \end{pmatrix},$$

let $\mathbf{G}(\lambda) = \mathbf{M}^H\mathbf{E}(\lambda)\mathbf{C}^{1/2}\mathbf{X}$ and $\mathbf{J} = \mathbf{T}^{1/2}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\delta}$, two matrices respectively of size $p \times N$ and $N \times p$. Moreover, let $\mathbf{g}(\lambda) = [g_1(\lambda), \dots, g_{Np}(\lambda)]^T = \text{vec}(\mathbf{G}(\lambda))$ and $\mathbf{j} = [j_1, \dots, j_{Np}]^T = \text{vec}(\mathbf{J})$. We obtain:

$$\hat{\gamma}_m^{cross}(\lambda) = \frac{2}{N}\,\mathcal{R}e\left(\text{vec}^T(\mathbf{G}(\lambda))\text{vec}(\mathbf{J})\right),$$
$$= \frac{2}{N}\sum_{k=1}^{Np}\mathcal{R}e(g_k(\lambda))\mathcal{R}e(j_k) - \mathcal{I}m(g_k(\lambda))\mathcal{I}m(j_k).$$

This expression can be transformed by introducing $\mathbf{A} = \mathbf{M}^H\mathbf{E}\mathbf{C}^{1/2}\otimes\mathbf{I}_N$, $\mathbf{B} = \mathbf{T}^{1/2}\boldsymbol{\Gamma}^{1/2}\otimes\mathbf{I}_p$, $\tilde{\mathbf{g}}(\lambda) = \mathbf{A}^{-1}\mathbf{g}(\lambda)$ $\tilde{\mathbf{j}} = \mathbf{B}^{-1}\mathbf{j}$, $a_k = \sum_{l=1}^{Np}(\mathbf{A}^T)_{l,k}$ and $b_k = \sum_{s=1}^{Np}(\mathbf{B})_{s,k}$:

$$\hat{\gamma}_m^{cross}(\lambda)$$
$$= \frac{2}{N}\,\mathcal{R}e\left(\left(\mathbf{A}^{-1}\mathbf{g}(\lambda)\right)^T\mathbf{A}^T\mathbf{B}\left(\mathbf{B}^{-1}\mathbf{j}\right)\right)$$
$$= \frac{2}{N}\sum_{k=1}^{Np}\mathcal{R}e\left(\sum_{l=1}^{Np}(\mathbf{A}^T)_{l,k}\tilde{\mathbf{g}}_k(\lambda)\right)\mathcal{R}e\left(\sum_{s=1}^{Np}(\mathbf{B})_{s,k}\tilde{\mathbf{j}}_k\right)$$
$$- \mathcal{I}m\left(\sum_{l=1}^{Np}(\mathbf{A}^T)_{l,k}\tilde{\mathbf{g}}_k(\lambda)\right)\mathcal{I}m\left(\sum_{s=1}^{Np}(\mathbf{B})_{s,k}\tilde{\mathbf{j}}_k\right),$$
$$= \frac{2}{N}\sum_{k=1}^{Np}\mathcal{R}e\left(a_k\,\tilde{g}_k(\lambda)\right)\mathcal{R}e\left(b_k\,\tilde{j}_k\right)$$
$$- \mathcal{I}m\left(a_k\,\tilde{g}_k(\lambda)\right)\mathcal{I}m\left(b_k\,\tilde{j}_k\right).$$

The variables $a_k\,\tilde{g}_k(\lambda)$ and $b_k\,\tilde{j}_k$ are two independent complex Gaussian variables with variances respectively equal to $|\tilde{a}_k(\lambda)|^2$ and $|b_k|^2$. We can apply the following lemma:

Lemma 10. *Let x and y be two independent Gaussian $\mathcal{N}(0,1)$ scalar random variables, then for any $\tau \in (-1\ 1)$, then $\mathbb{E}\left[\exp\left(\tau\,x\,y\right)\right] = (1-\tau^2)^{-1/2}$.*

Proof: The proof is derived in [31] through lemma 13. ∎

Let $\nu > 0$ a real such that : $\nu^{-1} > \sup_{k\in[\![1,Np]\!],\lambda\in[0,2\pi)} \left(|\tilde{a}_k(\lambda)|^2\,|b_k|^2\right)$. Then, for a fixed $\lambda \in [0\ 2\pi)$, from Lemma 10 and from the Markov Inequality:

$$\mathbb{P}\left[\hat{\gamma}_m^{cross}(\lambda) > x \mid \mathbf{T}\right]$$
$$= \mathbb{P}\left[\exp\left(N\,\nu\,\hat{\gamma}_m^{cross}(\lambda)\right) > \exp\left(N\,\nu\,x\right) \mid \mathbf{T}\right]$$
$$\leq \exp\left(-N\,\nu\,x\right)\mathbb{E}\left[\exp\left(2\,\nu\sum_{k=1}^{Np}\left[\mathcal{R}e\left(a_k\tilde{g}_k(\lambda)\right)\mathcal{R}e\left(b_k\tilde{j}_k\right)\right.\right.\right.$$
$$\left.\left.\left.-\mathcal{I}m\left(a_k\,\tilde{g}_k(\lambda)\right)\mathcal{I}m\left(b_k\,\tilde{j}_k\right)\right]\right)\right]$$
$$\leq \exp\left(-N\,\nu\,x\right)\prod_{k=1}^{Np}\left(1-4\,\nu^2\frac{|\tilde{a}_k(\lambda)|^2}{2}\frac{|b_k|^2}{2}\right)^{-1/2}$$
$$\left(1-4\,\nu^2\frac{|\tilde{a}_k(\lambda)|^2}{2}\frac{|b_k|^2}{2}\right)^{-1/2}$$
$$\leq \exp\left(-N\nu x-\sum_{k=1}^{Np}\log\left(1-\nu^2\,|\tilde{a}_k(\lambda)|^2\,|b_k|^2\right)^{-1}\right)$$

Moreover, since the $\Gamma_{i,j}$ are absolutely summable (Assumption 3), it exists a constant K such that:

$$|b_k|^2 = \left|\sum_{l=1}^{Np}\sqrt{\tau_l}\,\Gamma_{l,k}^{1/2}\right|^2 \leq K\sum_{l=1}^{Np}\tau_l\,.$$

Furthermore, since $\frac{1}{N}\sum_{l=1}^{Np}\tau_l \xrightarrow[N\to\infty]{} \mathbb{E}(\tau_i) = 1$, we obtain $|b_k|^2 \leq N\,K$. To deal with $|\tilde{a}_k(\lambda)|$, let K_1 and K_2 be some constants and remind that, for a fixed j, the $\{c_{i,j}\}_i$ and the $\{M_{i,j}\}_i$ are absolutely summable:

$$|a_k(\lambda)| = \frac{1}{m}\left|\sum_{s=1}^{p}\sum_{l,j=1}^{m}c_{l,k}\,M_{j,s}^{\star}\,e^{i(l-j)\,\lambda}\right|,$$
$$\leq \frac{1}{m}\sum_{l=1}^{m}|c_{l,k}|\,m\sum_{s,j=1}^{p,m}|M_{j,s}^{\star}|,$$
$$\leq p\,K_2\max_s\left(\sum_{j=1}^{m}|M_{j,s}^{\star}|\right) = p\,K_1\,.$$

We obtain $\nu^2\,|\tilde{a}_k(\lambda)|^2\,|b_k|^2 \leq \nu^2\,N^2\,p^2\,K\,K_1^2$ with $p \ll N$. Let q and ϵ be two positive reals small enough and such that:

$$\nu^2 = \left(\frac{q}{N^{1/2+\epsilon}}\right)^2 < \frac{K\,K_1^2}{N}\,.$$

Then $\lim_{N\to\infty}\nu^2\,|a_k(\lambda)|^2\,|b_k|^2 = 0$ and $\log\left(1-\nu^2\,|\tilde{a}_k(\lambda)|^2\,|b_k|^2\right)^{-1} \sim \nu^2\,|\tilde{a}_k(\lambda)|^2\,|b_k|^2$. Thereby, with A defining a constant, it can be obtained:

$$\mathbb{P}\left[\hat{\gamma}_m^{cross}(\lambda) > x \mid \mathbf{T}\right] \leq \exp\left(-N^{1/2-\epsilon}\,q\,x - A\right).$$

Then, integrating with respect to any density $p_{\mathbf{T}}(.)$ of $\mathbf{T}$ leads to:

$$\mathbb{P}\left[\hat{\gamma}_m^{cross}(\lambda) > x\right] = \int \mathbb{P}\left[\hat{\gamma}_m^{cross}(\lambda) > x \mid \mathbf{T}\right]p_{\mathbf{T}}(\mathbf{T})\,d\mathbf{T}$$
$$\leq \exp\left(-N^{1/2-\epsilon}\,q\,x - A\right).$$

This proves that, for any λ_i, $\mathbb{P}\left[\hat{\gamma}_m^{cross}(\lambda_i) > x\right] \xrightarrow[N\to\infty]{} 0$. It remains now to prove that $\max_{i\in\mathcal{I}}\sup_{\lambda\in[\lambda_i\ \lambda_{i+1}]}|\hat{\gamma}_m^{cross}(\lambda) - \hat{\gamma}_m^{cross}(\lambda_i)| \xrightarrow{a.s.} 0$. This will be left to the reader as it follows the same proof as for χ_1 of (24). We have so $\mathbb{P}\left[\sup_{\lambda\in[0\ 2\pi)}\hat{\gamma}_m^{cross}(\lambda) > x\right] \xrightarrow[N\to\infty]{} 0$.

4) Analysis of $\sup_{\lambda\in[0,2\pi)}|\hat{\gamma}_m^{sign}(\lambda)|$: The proof of convergence of this quantity follows the same principles. We have:

$$\hat{\gamma}_m^{sign}(\lambda) = \mathbf{d}_m^H(\lambda)\frac{\mathbf{M}\,\boldsymbol{\delta}^H\,\boldsymbol{\Gamma}\,\boldsymbol{\delta}\,\mathbf{M}^H}{N}\mathbf{d}_m(\lambda)\,.$$

As previously, let $\mathbb{I}_m$ be a $m \times m$ matrix containing 1 everywhere and let $\mathbf{E}(\lambda) = \mathbf{D}_m(\lambda)\,\mathbb{I}_m\,\mathbf{D}_m^H(\lambda)$. Then:

$$\hat{\gamma}_m^{sign}(\lambda) = 2\,\mathcal{R}e\left[\frac{1}{N}\,\mathrm{Tr}\left(\mathbf{M}\,\boldsymbol{\delta}\,\boldsymbol{\Gamma}\,\boldsymbol{\delta}^H\,\mathbf{M}^H\mathbf{E}\right)\right].$$

Let $\mathbf{A}(\lambda) = \mathbf{M}^H\,\mathbf{E}\,\mathbf{M}\,\boldsymbol{\delta}$ and $\mathbf{B} = \boldsymbol{\Gamma}\,\boldsymbol{\delta}^H$ be two matrix respectively of size $p \times N$ and $N \times p$. Defining $\mathbf{a}(\lambda) = \mathrm{vec}(\mathbf{A}(\lambda))$ and $\mathbf{b} = \mathrm{vec}(\mathbf{B})$, we have:

$$\hat{\gamma}_m^{sign}(\lambda) = \frac{2}{N}\mathcal{R}e\left(\mathrm{vec}^T(\mathbf{A}(\lambda))\,\mathrm{vec}(\mathbf{B})\right),$$
$$= \frac{2}{N}\mathcal{R}e\left(\mathbf{a}^T(\lambda)\left(\mathbf{M}^H\,\mathbf{E}\,\mathbf{M}\otimes\mathbf{I}_N\right)^{-T}\right.$$
$$\left.\left(\mathbf{M}^H\,\mathbf{E}\,\mathbf{M}\otimes\mathbf{I}_N\right)^T\left(\boldsymbol{\Gamma}\otimes\mathbf{I}_p\right)\left(\boldsymbol{\Gamma}\otimes\mathbf{I}_p\right)^{-1}\mathbf{j}\right),$$
$$= \frac{2}{N}\sum_{k=1}^{Np}\mathcal{R}e(a_k(\lambda))\,\mathcal{R}e(b_k) - \mathcal{I}m(a_k(\lambda))\,\mathcal{I}m(b_k)\,.$$

Let us define $\mathbf{C}(\lambda) = \mathbf{M}^H\,\mathbf{E}\,\mathbf{M}\otimes\mathbf{I}_N$, $\mathbf{D} = \boldsymbol{\Gamma}\otimes\mathbf{I}_p$, $\tilde{\mathbf{a}}(\lambda) = \mathbf{C}^{-1}(\lambda)\,\mathbf{a}(\lambda)$, $\tilde{\mathbf{b}} = \mathbf{D}^{-1}\mathbf{b}$, $c_k = \sum_{l=1}^{Np}\left(\mathbf{C}^T(\lambda)\right)_{l,k}$ and $d_k = \sum_{s=1}^{Np}(\mathbf{D})_{s,k}$. Using Lemma 10 and the Markov inequality, it can be shown that, for any fixed $\lambda \in [0\ 2\pi)$ and a constant μ such that $0 < \mu < \left(\sup_{\lambda\in[0,2\pi)}\|\mathbf{C}(\lambda)\|\sup\|\mathbf{D}\|\right)^{-1}$:

$$\mathbb{P}\left[\hat{\gamma}_m^{sign}(\lambda) > x\right]$$
$$\leq \exp\left(-N\,\nu\,x - \sum_{k=1}^{Np}\log\left(1-\mu^2\,|c_k(\lambda)|^2\,|d_k|^2\right)^{-1}\right).$$

As the matrix $\boldsymbol{\Gamma}$ is absolutely summable, then, for all k, $|d_k|^2 \leq K$ where K is a constant. Now, for all k, we have

$$|c_k(\lambda)| = \left|\sum_{s=1}^{p}\left[\sum_{l=1}^{m}M_{l,s}\sum_{j=1}^{m}M_{j,k}\,e^{i\,(j-l)\lambda}\right]\right|,$$
$$\leq \sum_{s=1}^{p}\left[\sum_{l=1}^{m}|M_{l,s}|\sum_{j=1}^{m}|M_{j,k}|\right].$$

The columns of $\mathbf{M}$ are absolutely summable. As p is fixed and $p \ll N$, with K a constant, we have $|c_k(\lambda)| \leq K$. The coefficients of the matrix $\boldsymbol{\Gamma}$ being absolutely summable, for all k, we have find a constant K_1 such that $|d_k| \leq K_1$. By defining w as a constant small enough and $\mu = \dfrac{w}{\sqrt{N}}$ such that $\mu^2 \, |c_k|^2 \, |d_k|^2 \underset{N \longrightarrow \infty}{\longrightarrow} 0$, then, for all $x > 0$ and A a constant, we have the following inequality:

$$\mathbb{P}[\hat{\gamma}_m^{sign}(\lambda) > x] \leq \exp\left(-N^{1/2} \, w \, x - A\right).$$

As for $\gamma_m^{cross}(\lambda)$, it remains to prove than $\max_{i \in \mathcal{I}} \sup_{\lambda \in [\lambda_i \ \lambda_{i+1}]} |\hat{\gamma}_m^{sign}(\lambda) - \hat{\gamma}_m^{sign}(\lambda_i)| \xrightarrow{a.s.} 0$ and this will left to the reader as it is the same as the proof of χ_1. We have proven than $\mathbb{P}\left[\sup_{\lambda \in [0 \ 2\pi)} \hat{\gamma}_m^{sign}(\lambda) > x\right] \underset{N \longrightarrow \infty}{\longrightarrow} 0$. As the right term of (21) tends to zero when N is tends to infinity, the proof of Theorem 1 is completed.

B. Proof of Theorem 3

The proof follow the same idea. With the notation $\check{\mathbf{C}}_{FP} = \mathcal{T}(\hat{\mathbf{C}}_{FP})$ where $\mathcal{T}$ is the Toeplitz operator defined in the introduction, the equation to prove becomes:

$$\left\|\mathcal{T}(\hat{\mathbf{C}}_{FP}) - \mathbb{E}\left[v(\tau\gamma)\,\tau\right]\mathbf{C}\right\| \xrightarrow{a.s.} 0. \quad (25)$$

This equation can be split as:

$$\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right) - \mathbb{E}\left[v(\tau\gamma)\,\tau\right]\mathbf{C}\right\|$$
$$\leq \left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\right\| + \left\|\mathcal{T}\left(\hat{\mathbf{S}}\right) - \mathbb{E}\left[v(\tau\,\gamma)\,\tau\right]\mathbf{C}\right\|.$$

Let us considering the following notations:

- $\hat{\mathbf{S}}$ the matrix such as $\left\|\check{\boldsymbol{\Sigma}} - \hat{\mathbf{S}}\right\| \xrightarrow{a.s.} 0$, as Theorem 3 has stated. As a reminder, $\hat{\mathbf{S}}$ is the matrix defined by:

$$\hat{\mathbf{S}} = \frac{1}{N}\sum_{i=1}^{N} v\left(\tau_i\,\gamma\right)\mathbf{y}_{wi}\,\mathbf{y}_{wi}^H,$$

where γ is the unique solution (if defined) of:

$$1 = \frac{1}{N}\sum_{i=1}^{N}\frac{\psi(\tau_i\,\gamma)}{1 + c\,\psi(\tau_i\,\gamma)},$$

where $g : x \mapsto \dfrac{x}{1 - c\,\phi(x)}$, $v : x \mapsto u \circ g^{-1}(x)$ and $\psi : x \mapsto x\,v(x)$.

- If $\mathbf{A} = \mathcal{T}\left((a_0, \ldots, a_{m-1})^T\right)$ is a Toeplitz matrix $(a_{-k} = a_k^*)$, we can define the spectral density as:

$$\gamma^{\mathbf{A}}(\lambda) \triangleq \sum_{k=1-m}^{m-1} a_k\,e^{i\,k\,\lambda}.$$

Finally, we denote by $\hat{\gamma}^{\mathbf{A}}(\lambda)$ the estimated spectral density of Toeplitz matrix $\mathbf{A}$.

To prove the consistency, we will decompose, as for Theorem 1, the equation (26) in two parts. As matrices $\mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right)$

and $\mathbf{C}$ are Toeplitz, it follows through (17):

$$\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right) - \mathbb{E}[v(\tau\,\gamma)\,\tau]\,\mathbf{C}\right\|$$
$$\leq \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \gamma^{\mathbb{E}[v(\tau\,\gamma)\,\tau]\,\mathbf{C}}(\lambda)\right| + \left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}}\right)\right\|$$
$$\leq \chi_1 + \chi_2, \quad (26)$$

where $\chi_1 = \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \gamma^{\mathbb{E}[v(\tau\,\gamma)\,\tau]\,\mathbf{C}}(\lambda)\right|$ and $\chi_2 = \left\|\mathcal{T}(\hat{\mathbf{C}}_{FP} - \hat{\mathbf{S}})\right\|$.

1) Part 1: convergence of $\chi_1 = \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \gamma^{\mathbb{E}[v(\tau\,\gamma)\,\tau]\,\mathbf{C}}(\lambda)\right|$: We will split χ_1 into two sub-terms:

$$\sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \gamma^{\mathbb{E}[v(\tau\,\gamma)\,\tau]\,\mathbf{C}}(\lambda)\right|$$
$$\leq \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right]\right|$$
$$+ \sup_{\lambda \in [0,2\pi)} \left|\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] - \gamma^{\mathbb{E}[v(\tau\,\gamma)\,\tau]\,\mathbf{C}}(\lambda)\right|,$$
$$\leq \chi_{11} + \chi_{12},$$

where $\chi_{11} = \sup_{\lambda \in [0,2\pi)} \left|\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) - \mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right]\right|$ and $\chi_{12} = \sup_{\lambda \in [0,2\pi)} \left|\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] - \gamma^{\mathbb{E}[v(\tau\,\gamma)\,\tau]\,\mathbf{C}}(\lambda)\right|$.

Part 1.1: convergence of χ_{11}: We will need the following lemma:

Lemma 11.
$$\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) = \mathbf{d}_m^H(\lambda)\,\hat{\mathbf{S}}\,\mathbf{d}_m(\lambda), \quad (27)$$

and:

$$\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] = \mathbb{E}\left[v(\tau\,\gamma)\,\tau\right]\,\mathbf{d}_m^H(\lambda)\,\mathbf{I}_m\,\mathbf{d}_m(\lambda), \quad (28)$$

where $\mathbf{d}_m(\lambda) = \dfrac{1}{\sqrt{m}}\left[1, e^{-i\,\lambda}, \ldots, e^{-i\,(m-1)\,\lambda}\right]^T$.

Proof: This is the same idea than for Lemma 5. First, we can write:

$$\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda) = \sum_{k=1-m}^{m-1} \check{s}_k\,e^{i\,k\,\lambda},$$
$$= \frac{1}{m\,N}\sum_{k=1-m}^{m-1} e^{i\,k\,\lambda}\sum_{j=0}^{m-1}\sum_{n=0}^{N-1} \hat{s}_{j,n}\,\hat{s}_{j+k,n}^{\star}\,\mathbb{1}_{0 \leq j+k < m},$$
$$= \frac{1}{m\,N}\sum_{l,l'=0}^{m-1} e^{-i\,(l'-l)\,\lambda}\sum_{n=0}^{N-1} \hat{s}_{l,n}\,\hat{s}_{l',n}^{\star} = \mathbf{d}_m^H(\lambda)\,\hat{\mathbf{S}}\,\mathbf{d}_m(\lambda).$$

The first part of the Lemma is then proven. Concerning $\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right]$, we can define $\mathbf{D}$ as the diagonal matrix containing the $\{v(\tau_i\,\gamma)\}_{i \in [\![0, N-1]\!]}$. We obtain:

$$\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] = \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\hat{\mathbf{S}}\right]\,\mathbf{d}_m(\lambda),$$
$$= \mathbf{d}_m^H(\lambda)\,\mathbb{E}\left[\frac{\mathbf{Y}_w\,\mathbf{D}\,\mathbf{Y}_w^H}{N}\right]\,\mathbf{d}_m(\lambda).$$

Then expliciting each element of $\mathbb{E}\left[\mathbf{Y}_w\,\mathbf{D}\,\mathbf{Y}_w^H\right]$ leads to:

$$\left(\mathbb{E}\left[\mathbf{Y}_w\,\mathbf{D}\,\mathbf{Y}_w^H\right]\right)_{i,j} = \mathbb{E}\left[\sum_{n=0}^{N-1} v(\tau_n\,\gamma)\,y_{w\,i,n}\,y_{w\,j,n}^{\star}\right],$$

$$= \sum_{n=0}^{N-1}\mathbb{E}\left[v(\tau_n\,\gamma)\,\tau_n\right] = N\,\mathbb{E}\left[v(\tau_n\,\gamma)\,\tau_n\right].$$

We obtain the following result: $\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] = \mathbb{E}\left[v(\tau\,\gamma)\,\tau\right]\mathbf{d}_m^H(\lambda)\,\mathbf{I}_m\,\mathbf{d}_m(\lambda)$.

∎

The rest of the proof for χ_{11} is the same as for Theorem 1 $\hat{\gamma}^{noise}$, but with $\mathbf{T}$ containing the $\{\tau_i\}_i$ on its diagonal, we will have $\|\mathbf{T}\|_\infty\,\|\mathbf{D}\|_\infty$ instead of $\|\mathbf{T}\|_\infty$. We obtain so $\chi_{11}\xrightarrow{a.s.}0$ as $m\longrightarrow\infty$.

Part 1.2: convergence of χ_{12}: Lemma 11 and (27) give us

$$\mathbb{E}\left[\hat{\gamma}^{\hat{\mathbf{S}}}(\lambda)\right] = \mathbb{E}\left[v(\tau\,\gamma)\,\tau\right]\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda).$$

and $\mathbb{E}\left[v(\tau\,\gamma)\,\tau\right]\gamma^{\mathbf{C}}(\lambda) = \mathbb{E}\left[v(\tau\,\gamma)\,\tau\right]\mathbf{d}_m^H(\lambda)\,\mathbf{C}\,\mathbf{d}_m(\lambda)$. This yields $\chi_{12}=0$.

2) Part 2: convergence of $\chi_2 = \left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\right\|$: It is proven, in [43] that $\left\|\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right\|\xrightarrow{a.s.}0$. Let $\mathbf{J}$ be a matrix such that $(\mathbf{J})_{j-i=1}=1$ and 0 elsewhere. $\mathbf{J}^k$ contains 1 only on the kth diagonal. As before, thanks to (17), we have:

$$\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\right\| \leq \sup_{\lambda\in[0,2\pi)}\left|\sum_{k=1-m}^{m-1}\left(\check{f}p_k-\check{s}_k\right)e^{i\,k\,\lambda}\right|.$$

Let us define $\mathcal{T}\left(\hat{\mathbf{C}}_{FP}\right) = \mathcal{L}\left(\left(\check{f}p_0,\ldots,\check{f}p_{m-1}\right)^T\right)$ with $\check{f}p_{-k}=\check{f}p_k^*$ and $\mathcal{T}\left(\hat{\mathbf{S}}\right) = \mathcal{L}\left(\left(\check{s}_0,\ldots,\check{s}_{m-1}\right)^T\right)$ with $\check{s}_{-k}=\check{s}_k^*$. We have:

$$\sup_{\lambda\in[0,2\pi)}\left|\sum_{k=1-m}^{m-1}\left(\check{f}p_k-\check{s}_k\right)e^{i\,k\,\lambda}\right|$$

$$= \sup_{\lambda\in[0,2\pi)}\left|\sum_{k=1-m}^{m-1}\frac{1}{m}\sum_{p-1}^{m}\left(\check{f}p_k-\check{s}_k\right)e^{i\,k\,\lambda}\mathbb{1}_{0\leq p+k\leq m}\right|,$$

$$= \sup_{\lambda\in[0,2\pi)}\left|\operatorname{Tr}\left(\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\frac{1}{m}\sum_{k=1-m}^{m-1}\left(\mathbf{J}^T\right)^k e^{i\,k\,\lambda}\right)\right|.$$

Moreover $\dfrac{1}{m}\displaystyle\sum_{k=1-m}^{m-1}\left(\mathbf{J}^T\right)^k e^{i\,k\,\lambda} = \mathbf{d}_m(\lambda)\,\mathbf{d}_m^H(\lambda)$. This leads to:

$$\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\right\|$$
$$\leq \sup_{\lambda\in[0,2\pi)}\left|\operatorname{Tr}\left(\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\mathbf{d}_m(\lambda)\,\mathbf{d}_m^H(\lambda)\right)\right|$$
$$= \sup_{\lambda\in[0,2\pi)}\left|\mathbf{d}_m^H(\lambda)\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\mathbf{d}_m(\lambda)\right|.$$

For any vector $\mathbf{x}$, the last equation becomes:

$$\sup_{\lambda\in[0,2\pi)}\left|\mathbf{d}_m^H(\lambda)\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\mathbf{d}_m(\lambda)\right|$$

$$\leq \sup_{\|\mathbf{x}\|_2=1}\left|\mathbf{x}^H\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\mathbf{x}\right|,$$

$$\leq \sup_{\|\mathbf{x}\|_2=1}\left\|\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\mathbf{x}\right\|_2 \leq \left\|\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right\|.$$

Finally, we obtain:

$$\left\|\mathcal{T}\left(\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right)\right\| \leq \left\|\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right\|.$$

As $\left\|\hat{\mathbf{C}}_{FP}-\hat{\mathbf{S}}\right\|\xrightarrow{a.s.}0$ then $\chi_2\xrightarrow{a.s.}0$ and the proof of Theorem 3 is completed.

APPENDIX B
PROOF OF THEOREM 2

As the proof is the same for $\check{\boldsymbol{\Sigma}}_{SCM}$ and $\check{\boldsymbol{\Sigma}}_{FP}$, let $\check{\boldsymbol{\Sigma}}$ denote one or the other of these matrices.

From the equations (7) and (15), as $\check{\mathbf{y}}_{wi}=\check{\mathbf{C}}^{-1/2}\,\mathbf{y}_i$, $\check{\boldsymbol{\Sigma}}$ is the unique solution of:

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\mathbf{y}_i^H\,\check{\mathbf{C}}^{-1/2}\,\boldsymbol{\Sigma}^{-1}\,\check{\mathbf{C}}^{-1/2}\,\mathbf{y}_i\right)\times$$
$$\check{\mathbf{C}}^{-1/2}\,\mathbf{y}_i\,\mathbf{y}_i^H\,\check{\mathbf{C}}^{-1/2}.$$

Rewriting this equation with the $\{\check{\mathbf{y}}_{wi}\}_i$

$$\mathbf{C}^{-1/2}\check{\mathbf{C}}^{1/2}\,\boldsymbol{\Sigma}\,\check{\mathbf{C}}^{1/2}\,\mathbf{C}^{-1/2}$$
$$= \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\check{\mathbf{y}}_{wi}^H\left(\mathbf{C}^{-1/2}\,\check{\mathbf{C}}^{1/2}\,\boldsymbol{\Sigma}\,\check{\mathbf{C}}^{1/2}\,\mathbf{C}^{-1/2}\right)^{-1}\times\right.$$
$$\left.\check{\mathbf{y}}_{wi}\right)\check{\mathbf{y}}_{wi}\,\check{\mathbf{y}}_{wi}^H,$$

we obtain the following relationship between $\check{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}$:

$$\check{\boldsymbol{\Sigma}} = \check{\mathbf{C}}^{-1/2}\,\mathbf{C}^{1/2}\,\hat{\boldsymbol{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}. \tag{29}$$

Then, equation (11) can be rewritten as

$$\left\|\check{\boldsymbol{\Sigma}}-\hat{\mathbf{S}}\right\| \leq \left\|\check{\boldsymbol{\Sigma}}-\hat{\boldsymbol{\Sigma}}\right\| + \left\|\hat{\boldsymbol{\Sigma}}-\hat{\mathbf{S}}\right\|. \tag{30}$$

Concerning the second term of the right hand side of (30), it is proven in [27] that the matrix $\hat{\mathbf{S}}$ given by (9) is such that

$$\left\|\hat{\boldsymbol{\Sigma}}-\hat{\mathbf{S}}\right\|\xrightarrow{a.s.}0. \tag{31}$$

With (29), the first term of right hand side of (30) can be rewritten as:

$$\left\|\check{\boldsymbol{\Sigma}}-\hat{\boldsymbol{\Sigma}}\right\| \leq \left\|\check{\mathbf{C}}^{-1/2}\,\mathbf{C}^{1/2}\,\hat{\boldsymbol{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}-\hat{\boldsymbol{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}\right\|$$
$$+ \left\|\hat{\boldsymbol{\Sigma}}\,\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}-\hat{\boldsymbol{\Sigma}}\right\|. \tag{32}$$

After left and right factorizations, we obtain:

$$\left\|\check{\boldsymbol{\Sigma}}-\hat{\boldsymbol{\Sigma}}\right\| \leq \left\|\check{\mathbf{C}}^{-1/2}\,\mathbf{C}^{1/2}-\mathbf{I}_m\right\|\left\|\hat{\boldsymbol{\Sigma}}\right\|\left(\left\|\mathbf{C}^{1/2}\,\check{\mathbf{C}}^{-1/2}\right\|+1\right).$$

As $\|\mathbf{C}\|$ has a bounded support, $\|\check{\mathbf{C}}\|$ is bounded too since its eigenvalues support converges almost surely toward the true distribution. Moreover, Theorem 1 and Theorem 2 have proved the consistency $\left\|\mathbf{C}-\check{\mathbf{C}}\right\|\xrightarrow{a.s.}0$. This ensures the proof.

REFERENCES

[1] J. Vinogradova, R. Couillet, and W. Hachem, "A new method for source detection, power estimation, and localization in large sensor networks under noise with unknown statistics," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3943–3946.

[2] N. Arkind and B. Nadler, "Parametric joint detection-estimation of the number of sources in array processing," in *2010 IEEE Sensor Array and Multichannel Signal Processing Workshop*, Oct 2010, pp. 269–272.

[3] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 590–600, March 1992.

[4] B. Nadler, F. Penna, and R. Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise level," in *2011 IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–5.

[5] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec 1974.

[6] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sept 1978.

[7] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[8] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 34, no. 3, pp. 276–280, March 1986.

[9] K. Abed-Meraim, J. F. Cardoso, A. Y. Gorokhov, P. Loubaton, and E. Moulines, "On subspace methods for blind identification of single-input multiple-output FIR systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 42–55, Jan 1997.

[10] H. Liu and G. Xu, "A subspace method for signature waveform estimation in synchronous CDMA systems," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1346–1354, Oct 1996.

[11] Z. Bai and J. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *Annals of probability*, 1998.

[12] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate analysis*, 1995.

[13] K. Cawse, A. Robin, and M. Sears, "The effect of noise whitening on methods for determining the intrinsic dimension of a hyperspectral image," in *2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2011, pp. 1–4.

[14] A. Combernoux, F. Pascal, M. Lesturgie, and G. Ginolhac, "Performances of low rank detectors based on random matrix theory with application to STAP," in *2014 International Radar Conference*, Oct 2014, pp. 1–4.

[15] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3930–3941, Oct 2009.

[16] N. Farsi, B. Escrig, and A. Hamza, "Estimation of the number of signals based on a sequence of hypothesis test and Random Matrix Theory," in *2015 Third World Conference on Complex Systems (WCCS)*, Nov 2015, pp. 1–5.

[17] R. Couillet, F. Pascal, and J. W. Silverstein, "A joint robust estimation and random matrix framework with application to array processing," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 6561–6565.

[18] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[19] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3930–3941, Oct 2009.

[20] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet, "A subspace estimator for fixed rank perturbations of large random matrices," *Journal of Multivariate Analysis*, vol. 114, pp. 427–447, 2013.

[21] F. Pascal and A. Kammoun, "Théorie des matrices aléatoires robustes et applications à la détection radar," *Traitement du Signal*, 2016.

[22] K. Cawse, M. Sears, A. Robin, S. B. Damelin, K. Wessels, F. van den Bergh, and R. Mathieu, "Using random matrix theory to determine the number of endmembers in a hyperspectral image," in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, June 2010, pp. 1–4.

[23] J. Vinogradova, R. Couillet, and W. Hachem, "Statistical inference in large antenna arrays under unknown noise pattern," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5633–5645, Nov 2013.

[24] E. Terreaux, J. P. Ovarlez, and F. Pascal, "Anomaly detection and estimation in hyperspectral imaging using random matrix theory tools," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2015, pp. 169–172.

[25] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Robust covariance matrix estimation in heterogeneous low rank context," *IEEE Transactions on Signal Processing*, vol. 64, no. 22, pp. 5794–5806, Nov 2016.

[26] A. Halimi, P. Honeine, M. Kharouf, C. Richard, and J. Y. Tourneret, "Estimating the intrinsic dimension of hyperspectral images using a noise-whitened eigengap approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3811–3821, July 2016.

[27] R. Couillet, "Robust spiked random matrices and a robust G-MUSIC estimator," *Journal of Multivariate Analysis*, vol. 140, pp. 139 – 161, 2015.

[28] J. P. Ovarlez, S. K. Pang, V. Pascal, F.and Achard, and T. K. Ng, "Robust detection using the SIRV background modelling for hyperspectral imaging," in *2011 IEEE International Geoscience and Remote Sensing Symposium*, July 2011, pp. 4316–4319.

[29] F. Gini, "Sub-optimum coherent radar detection in a mixture of K-distributed and Gaussian clutter," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 144, no. 1, pp. 39–48, Feb 1997.

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

[31] J. Vinogradova, R. Couillet, and W. Hachem, "Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4903–4913, Sept 2015.

[32] P. Vallet and P. Loubaton, "Toeplitz rectification and DOA estimation with music," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2237–2241.

[33] R. Couillet, M. S. Greco, J. P. Ovarlez, and F. Pascal, "RMT for whitening space correlation and applications to radar detection," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2015, pp. 149–152.

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

[35] 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.

[36] 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, 1970.

[37] K. Yao, "A representation theorem and its applications to spherically-invariant random processes," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 600–608, Sept 1973.

[38] 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.

[39] 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.

[40] P. Loubaton, "On the almost sure location of the singular values of certain gaussian block-hankel large random matrices," *Journal of Theorical Probability*, 2016.

[41] R. database. Hyperspectral remote sensing scenes. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[42] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[43] 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*, 2015.