ESTIMATION OF LARGE TOEPLITZ COVARIANCE MATRICES AND APPLICATION TO SOURCE DETECTION

Julia Vinogradova[1]*, Romain Couillet*[2]*, and Walid Hachem*[1]

[1]CNRS LTCI, Télécom ParisTech, France
[2]Telecommunication departement, Supélec, France

ABSTRACT

In this paper, performance results of two types of Toeplitz covariance matrix estimators are provided. Concentration inequalities for the spectral norm for both estimators have been derived showing exponential convergence of the error to zero. It is shown that the same rates of convergence are obtained in the case where the aggregated matrix of time samples is corrupted by a rank one matrix. As an application based on this model, source detection by a large dimensional sensor array with temporally correlated noise is studied.

Index Terms— Toeplitz covariance matrix, concentration inequalities, correlated noise, source detection.

1. INTRODUCTION

Stationary processes are used in many fields of signal processing. A fundamental task is to estimate the covariance matrix of these processes. Let $(v_t)_{t\in\mathbb{Z}}$ be a complex circularly symmetric Gaussian stationary process with zero mean and covariance function $(r_k)_{k\in\mathbb{Z}}$ with $r_k = \mathrm{E}[v_{t+k}v_t^*]$ and $r_k \to 0$ as $k \to \infty$. Consider N independent realizations $(v_{1,t})_t, ..., (v_{N,t})_t$ of $(v_t)_{t\in\mathbb{Z}}$ over the time window $t \in \{0, \dots, T-1\}$, and stack the observations in a matrix $V_T = [v_{n,t}]_{n,t=0}^{N-1,T-1}$. We can write $V_T = W_T R_T^{1/2}$, where $W_T \in \mathbb{C}^{N\times T}$ has independent $\mathcal{CN}(0,1)$ (standard circularly symmetric complex Gaussian) entries and $R_T^{1/2}$ is any square root of the Hermitian nonnegative definite Toeplitz $T \times T$ matrix defined by $R_T \triangleq [r_{i-j}]_{0\le i,j\le T-1}$. The objective is to estimate R_T from V_T.

Recently, from the increasing interest to large dimensional array processing, this estimation problem has drawn a renewed attention considering the high dimensional setting for which both N and T are large. Generally the estimation approaches are based on the biased and unbiased estimates $\hat{r}_{k,T}^b$ and $\hat{r}_{k,T}^u$ for r_k, respectively defined by

$$\hat{r}_{k,T}^b = \frac{1}{NT}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v_{n,t}^* \mathbb{1}_{0\le t+k\le T-1}$$

$$\hat{r}_{k,T}^u = \frac{1}{N(T-|k|)}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v_{n,t}^* \mathbb{1}_{0\le t+k\le T-1}$$

where $\mathbb{1}_A$ is the indicator function on the set A. Depending on the relative rate of growth of N and T, the estimates $\widehat{R}_T^b = [\hat{r}_{i-j,T}^b]_{0\le i,j\le T-1}$ and $\widehat{R}_T^u = [\hat{r}_{i-j,T}^u]_{0\le i,j\le T-1}$ may not be consistent. The estimation approaches developed during the last decade propose all to build banded or tapered versions of the estimated matrix $\widehat{R}_T$ by down-scaling estimates of entries sufficiently away from the diagonal [1, 2, 3, 4, 5]. These give rise for instance to the consistent estimate $\widehat{R}_{\gamma,T} = [[\widehat{R}_T]_{i,j}\mathbb{1}_{|i-j|\le\gamma}]$ for some well-chosen functions $\gamma(T)$ usually satisfying $\gamma(T) \to \infty$ and $\gamma(T)/T \to 0$. These methods however suffer from the following main limitations: (i) they assume the *a priori* knowledge of the rate of decrease of r_k (and restrict these rates to specific classes); (ii) the results are asymptotic in nature and do not provide explicit rules for selecting $\gamma(T)$ for practical finite values of N and T; (iii) the operations of banding and tapering do not guarantee the positive definiteness of the resulting covariance estimate.

In the present paper, the only assumption on r_k is that $\sum_{k=-\infty}^{\infty} |r_k| < \infty$. The non banded estimates of R_T are given by $\widehat{R}_T^b$ and $\widehat{R}_T^u$ and the consistence of these estimates is obtained thanks to the choice $N, T \to \infty$ with $N/T \to c \in (0, \infty)$. This setting is more practical in applications as long as both the finite values N and T are large and of the same order of magnitude. The contribution of this work consists in the establishment of concentration inequalities for the errors in spectral norm $\|R_T - \widehat{R}_T^b\|$ and $\|R_T - \widehat{R}_T^u\|$. The results are then generalized to the case where V_T is replaced by $V_T + P_T$ for a rank-one matrix P_T and we show that the concentration inequalities remain identical. As an application of the latter, we study a single source detection (modeled through P_T) by an array of N sensors embedded in a temporally correlated noise (modeled by V_T) performed in two steps. First, the matrix R_T is estimated from $V_T + P_T$ giving $\widehat{R}_T^b$ or $\widehat{R}_T^u$ which are both nonnegative definite with probability one. Then this es-

This work is funded in part by the French Ile-de-France region, DIM LSC fund, Digiteo project DESIR, and in part by the ANR-12-MONU-OOO3 DIONISOS.

timate is used as a whitening matrix, before applying a generalized likelihood ratio test (GLRT) procedure on the whitened observation.

The remainder of the article is organized as follows. The concentration inequalities for both biased and unbiased estimates are presented in Section 2. The extension to the rank-one perturbation model is provided in Section 3 and applied in the practical context of source detection in Section 4.

2. PERFORMANCE OF COVARIANCE MATRIX ESTIMATORS

2.1. Model and assumptions

Let $(r_k)_{k\in\mathbb{Z}}$ be a doubly infinite sequence of covariance coefficients. For any $T \in \mathbb{N}$, let R_T be a Hermitian nonnegative definite Toeplitz matrix

$$R_T = \mathcal{T}(r_{-(T-1)}, \dots, r_{T-1}) \triangleq \begin{bmatrix} r_0 & r_1 & \dots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \dots & r_{-1} & r_0 \end{bmatrix}.$$

Given $N = N(T) > 0$, consider the matrix model

$$V_T = [v_{n,t}]_{n,t=0}^{N-1,T-1} = W_T R_T^{1/2} \tag{1}$$

where $W_T = [w_{n,t}]_{n,t=0}^{N-1,T-1}$ has independent $\mathcal{CN}(0,1)$ entries and $R_T^{1/2}$ any square root of R_T. It is clear that $r_k = \mathbb{E}[v_{n,t+k}v_{n,t}^*]$ for any t, k, and $n \in \{0,\dots,N-1\}$.

To pursue, we need the following two assumptions.

Assumption 1. *The covariance coefficients r_k are absolutely summable and $r_0 \neq 0$.*

With this assumption, the covariance function

$$\Upsilon(\lambda) \triangleq \sum_{k=-\infty}^{\infty} r_k e^{-\imath k\lambda}, \quad \lambda \in [0, 2\pi)$$

is continuous on the interval $[0, 2\pi]$. From [6, Lemma 4.1], $\|R_T\| \leq \|\Upsilon\|_\infty$ with $\|X\|$ standing for the spectral norm for a matrix and Euclidean norm for a vector, and $\|\cdot\|_\infty$ is the sup norm of a function. Hence, Assumption 1 implies that $\sup_T \|R_T\| < \infty$.

We assume the following asymptotic regime denoted as "$T \to \infty$":

Assumption 2. *$T \to \infty$ and $N/T \to c > 0$.*

2.2. Main results

Our objective is to study the performance of two estimators of the covariance function considered in the literature and defined as

$$\hat{r}_{k,T}^b = \frac{1}{NT}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v_{n,t}^* \mathbb{1}_{0\leq t+k\leq T-1} \tag{2}$$

$$\hat{r}_{k,T}^u = \frac{1}{N(T-|k|)}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v_{n,t}^* \mathbb{1}_{0\leq t+k\leq T-1}. \tag{3}$$

Since $\mathbb{E}\hat{r}_{k,T}^b = (1-|k|/T)r_k$ and $\mathbb{E}\hat{r}_{k,T}^u = r_k$, the estimate $\hat{r}_{k,T}^b$ is biased while $\hat{r}_{k,T}^u$ is unbiased. Define also

$$\widehat{R}_T^b \triangleq \mathcal{T}(\hat{r}_{-(T-1),T}^b, \dots, \hat{r}_{(T-1),T}^b) \tag{4}$$

$$\widehat{R}_T^u \triangleq \mathcal{T}(\hat{r}_{-(T-1),T}^u, \dots, \hat{r}_{(T-1),T}^u). \tag{5}$$

The advantage of the estimate $\widehat{R}_T^u$ is the fact that it is asymptotically unbiased while the estimate $\widehat{R}_T^b$ is structurally nonnegative definite. The following theorems provide the results on the spectral behavior of these matrices under the form of concentration inequalities on the random variables $\|\widehat{R}_T^b - R_T\|$ and $\|\widehat{R}_T^u - R_T\|$ [7]:

Theorem 1. *Let Assumptions 1 and 2 hold true and let $\widehat{R}_T^b$ be defined as in (4). Then, for any $x > 0$,*

$$\mathbb{P}\left[\left\|\widehat{R}_T^b - R_T\right\| > x\right] \leq \exp\left(-cT\left(\frac{x}{\|\Upsilon\|_\infty} - \log\left(1+\frac{x}{\|\Upsilon\|_\infty}\right) + o(1)\right)\right)$$

where $o(1)$ is with respect to T and depends on x.

Proof. The proof is available in [7]. □

Theorem 2. *Let Assumptions 1 and 2 hold true and let $\widehat{R}_T^u$ be defined as in (5). Then, for any $x > 0$,*

$$\mathbb{P}\left[\left\|\widehat{R}_T^u - R_T\right\| > x\right] \leq \exp\left(-\frac{cTx^2}{4\|\Upsilon\|_\infty^2 \log T}(1+o(1))\right)$$

where $o(1)$ is with respect to T and depends on x.

Proof. A sketch of the proof is given in Section 5.

Remark 1. *The sketch of the proof is provided only for Theorem 2 since it presents more difficulties than the proof of Theorem 1.*

□

From these theorems, the error in spectral norm is bounded by an exponentially decreasing function of T. As consequence, obtained by the Borel-Cantelli lemma, $\|\widehat{R}_T^b - R_T\| \to 0$ and $\|\widehat{R}_T^u - R_T\| \to 0$ almost surely as $T \to \infty$. The slower rate of decrease of $T/\log(T)$ in the unbiased estimator exponent is due to the increased inaccuracy in the estimates of r_k for k close to $T-1$. Figure 1 shows an empirical evaluation

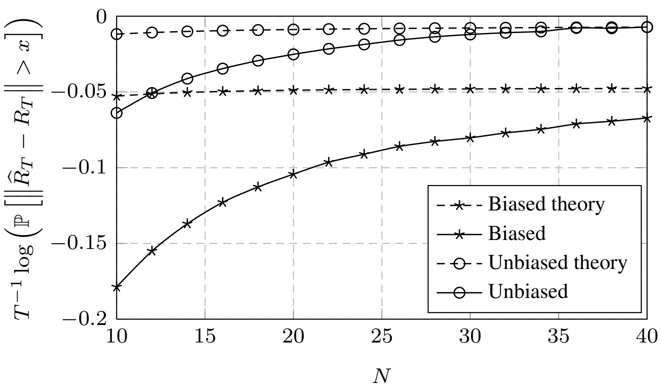

Fig. 1. Error probability of the spectral norm for $x = 2$, $c = 0.5$, $[R_T]_{k,l} = a^{|k-l|}$ with $a = 0.6$.

by Monte Carlo simulations of $\mathbb{P}[\|\widehat{R}_T - R_T\| > x]$ (Biased and Unbiased) with $\widehat{R}_T \in \{\widehat{R}_T^b, \widehat{R}_T^u\}$ compared against the theoretical exponential bounds of Theorems 1 and 2 (Biased theory and Unbiased theory). We observe that the rates obtained in Theorems 1 and 2 converge asymptotically to optimal.

3. COVARIANCE MATRIX ESTIMATION UNDER THE "SIGNAL PLUS NOISE" MODEL

Consider now the following model:

$$Y_T = [y_{n,t}]_{n,t=0}^{N-1,T-1} = P_T + V_T \tag{6}$$

where the $N \times T$ matrix V_T is unchanged and where P_T satisfies the following assumption:

Assumption 3. *$P_T \triangleq \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \Gamma_T^{1/2}$ where $\boldsymbol{h}_T \in \mathbb{C}^N$ is a deterministic vector such that $\sup_T \|\boldsymbol{h}_T\| < \infty$, the vector $\boldsymbol{s}_T = (s_0, \ldots, s_{T-1})^{\mathsf{T}} \in \mathbb{C}^T$ is a random vector independent of W_T with distribution $\mathcal{CN}(0, I_T)$, and $\Gamma_T^{1/2}$ satisfies $(\Gamma_T^{1/2})^{\mathsf{H}} \Gamma_T^{1/2} = \Gamma_T$ where $\Gamma_T = [\gamma_{ij}]_{i,j=0}^{T-1}$ is a $T \times T$ Hermitian nonnegative matrix with $\sup_T \|\Gamma_T\| < \infty$.*

The model can be seen as a Gaussian spatially white and temporally correlated noise with stationary temporal correlations perturbed by a rank one signal which can be also temporally correlated. The aim is still to estimate R_T from Y_T. We obtain the same expressions for the estimates given by (2) or (3) with only difference that the samples $v_{n,t}$ are replaced by the samples $y_{n,t}$. We have the following theorem proved in [7]:

Theorem 3. *Consider the model (6) and let Assumptions 1–3 hold true. Consider respectively the estimates*

$$\hat{r}_{k,T}^{bp} = \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} y_{n,t+k} y_{n,t}^* \mathbb{1}_{0 \le t+k \le T-1}$$

and

$$\hat{r}_{k,T}^{up} = \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} y_{n,t+k} y_{n,t}^* \mathbb{1}_{0 \le t+k \le T-1}.$$

Let $\widehat{R}_T^{bp} = \mathcal{T}(\hat{r}_{(1-T),T}^{bp}, \ldots, \hat{r}_{(T-1),T}^{bp})$ and $\widehat{R}_T^{up} = \mathcal{T}(\hat{r}_{(1-T),T}^{up}, \ldots, \hat{r}_{(T-1),T}^{up})$ be the Toeplitz matrices built from these estimates. Then for any $x > 0$, the following inequalities hold true:

$$\mathbb{P}\left[\left\|\widehat{R}_T^{bp} - R_T\right\| > x\right] \le \exp\left(-cT\left(\frac{x}{\|\boldsymbol{\Upsilon}\|_\infty} - \log\left(1 + \frac{x}{\|\boldsymbol{\Upsilon}\|_\infty}\right) + o(1)\right)\right)$$

and

$$\mathbb{P}\left[\left\|\widehat{R}_T^{up} - R_T\right\| > x\right] \le \exp\left(-\frac{cTx^2}{4\|\boldsymbol{\Upsilon}\|_\infty^2 \log T}(1 + o(1))\right).$$

Note that these results are similar to those presented in Section 2. The small rank perturbation P_T of the noise term V_T does not affect too much the estimators of R_T.

4. APPLICATION TO SOURCE DETECTION

Consider a sensor network composed of N sensors with zero (hypothesis H_0) or one (hypothesis H_1) source signal. The stacked signal matrix $Y_T = [y_0, \ldots, y_{T-1}] \in \mathbb{C}^{N \times T}$ observed during a time window of size T is modeled as

$$Y_T = \begin{cases} V_T & , H_0 \\ \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} + V_T & , H_1 \end{cases} \tag{7}$$

where $s_T^{\mathsf{H}} = [s_0^*, \ldots, s_{T-1}^*]$ are (hypothetical) independent $\mathcal{CN}(0,1)$ signals transmitted through the constant channel $\boldsymbol{h}_T \in \mathbb{C}^N$, and $V_T = W_T R_T^{1/2} \in \mathbb{C}^{N \times T}$ models a stationary noise matrix as in (1).

Most detection methods assume that a training pure noise sequence is available at the receiver. Here, without such an assumption, we proceed to an online signal detection test based on Y_T, by exploiting the consistence established in Theorem 3. First R_T is estimated by $\widehat{R}_T \in \{\widehat{R}_T^{bp}, \widehat{R}_T^{up}\}$, which is then used as a whitening matrix for Y_T:

$$Y_T \widehat{R}_T^{-1/2} = \begin{cases} W_T R_T^{1/2} \widehat{R}_T^{-1/2} & , H_0 \\ \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \widehat{R}_T^{-1/2} + W_T R_T^{1/2} \widehat{R}_T^{-1/2} & , H_1. \end{cases} \tag{8}$$

Since $\|R_T \widehat{R}_T^{-1} - I_T\| \to 0$ almost surely (by Theorem 3 as long as $\inf_{\lambda \in [0,2\pi)} \boldsymbol{\Upsilon}(\lambda) > 0$), for T large, the decision on the hypotheses (8) can be handled by the generalized likelihood ratio test (GLRT) [8] by approximating $W_T R_T \widehat{R}_T^{-1/2}$ as a purely white noise. We then have the following result [7]:

Theorem 4. *Let $\widehat{R}_T$ be any of $\widehat{R}_T^{bp}$ or $\widehat{R}_T^{up}$ strictly defined in Theorem 3 for Y_T now following model* (7). *Assume* $\inf_{\lambda\in[0,2\pi)} \boldsymbol{\Upsilon}(\lambda) > 0$ *and* $\liminf_T \|\boldsymbol{h}_T\|^2 \operatorname{Tr}\left(R_T^{-1}\right)/T \geq \sqrt{c}$ *and define the test*

$$\alpha = \frac{N\left\|Y_T\widehat{R}_T^{-1}Y_T^{\mathsf{H}}\right\|}{\operatorname{Tr}\left(Y_T\widehat{R}_T^{-1}Y_T^{\mathsf{H}}\right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma \tag{9}$$

where $\gamma \in \mathbb{R}^+$ *satisfies* $\gamma > (1+\sqrt{c})^2$. *Then, as* $T \to \infty$,

$$\mathbb{P}\left[\alpha \geq \gamma\right] \to \begin{cases} 0 & , H_0 \\ 1 & , H_1. \end{cases}$$

Recall from [8] that the decision threshold $(1+\sqrt{c})^2$ corresponds to the almost sure limiting largest eigenvalue of $\frac{1}{T}W_TW_T^{\mathsf{H}}$, that is the right-edge of the support of the Marčenko–Pastur law.

In the following we provide the simulations results for the performance of the test (9) with the noise modeled as an autoregressive process of order 1 with parameter a, *i.e.* $[R_T]_{k,l} = a^{|k-l|}$. The channel is written as a steering vector $\boldsymbol{h}_T = \sqrt{p/T}[1,\ldots,e^{2i\pi\theta(T-1)}]$ with $\theta = 10°$ and p a power parameter.

Figure 2 depicts the detection error $1-\mathbb{P}[\alpha \geq \gamma|H_1]$ of the test (9) for a false alarm rate (FAR) $\mathbb{P}[\alpha \geq \gamma|H_0] = 0.05$ under $\widehat{R}_T = \widehat{R}_T^{bp}$ (Biased) or $\widehat{R}_T = \widehat{R}_T^{up}$ (Unbiased) compared against the estimator that assumes R_T perfectly known (Oracle), *i.e.* that sets $\widehat{R}_T = R_T$ in (9), and against the GLRT test that wrongly assumes temporally white noise (White), *i.e.* that sets $\widehat{R}_T = I_T$ in (9). In the same setting as Figure 2, we now fix $N = 20, T = N/c = 40$ and plot the power of the test (9) versus signal-to-noise ratio (SNR) in Figure 3. The results are compared to the methods which estimate R_T from a pure noise sequence called Biased PN (pure noise) and Unbiased PN. We see that the proposed online method results are close to the method with PN estimation of R_T. Moreover, both figures suggest a close match in performance between Oracle and Biased, while Unbiased shows weaker performance. The gap between Biased and Unbiased confirms the theoretical conclusions.

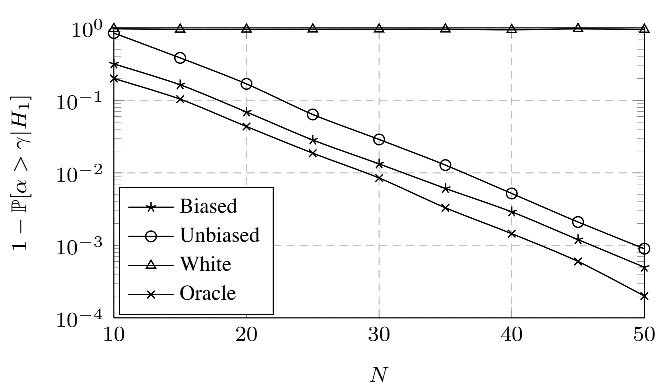

Fig. 2. Detection error versus N with FAR= 0.05, $p = 1$, SNR= 0 dB, $c = 0.5$, and $a = 0.6$.

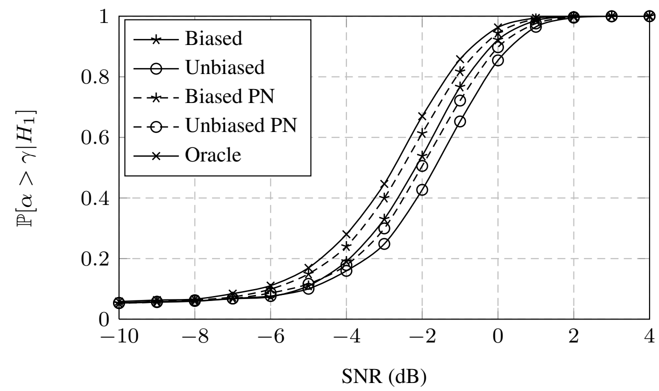

Fig. 3. Power of detection tests versus SNR (dB) with FAR= 0.05, $N = 20$, $c = 0.5$, and $a = 0.6$.

5. SKETCH OF THE PROOF OF THEOREM 2

Define

$$\widehat{\Upsilon}_T^u(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} \hat{r}_{k,T}^u e^{\imath k\lambda} \quad \text{and} \quad \Upsilon_T(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} r_k e^{\imath k\lambda}.$$

From [6, Lemma 4.1] for all $x > 0$

$$\mathbb{P}\left[\left\|\widehat{R}_T^u - R_T\right\| > x\right] \leq \mathbb{P}\left[\sup_{\lambda\in[0,2\pi)}\left|\widehat{\Upsilon}_T^u(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)\right| > x\right]$$

where $\Upsilon_T(\lambda) = \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)$ since the estimates $\hat{r}_{k,T}^u$ are unbiased. In order to control the term $\sup_{\lambda\in[0,2\pi)}\left|\widehat{\Upsilon}_T^u(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)\right|$ we proceed as follows. Choosing $\beta > 2$, let $\lambda_i = 2\pi\frac{i}{\lfloor T^\beta\rfloor}$, $i \in \mathcal{I}$, be a regular discretization of the interval $[0, 2\pi]$ with $\mathcal{I} = \{0,\ldots,\lfloor T^\beta\rfloor - 1\}$. We have

$$\begin{aligned}
&\sup_{\lambda\in[0,2\pi)}\left|\widehat{\Upsilon}_T^u(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)\right| \\
&\leq \max_{i\in\mathcal{I}}\sup_{\lambda\in[\lambda_i,\lambda_{i+1}]}\left|\widehat{\Upsilon}_T^u(\lambda) - \widehat{\Upsilon}_T^u(\lambda_i)\right| + \max_{i\in\mathcal{I}}\left|\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right| \\
&\quad + \max_{i\in\mathcal{I}}\sup_{\lambda\in[\lambda_i,\lambda_{i+1}]}\left|\mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)\right| \triangleq \chi_1 + \chi_2 + \chi_3.
\end{aligned}$$

The concentration inequalities on the random terms χ_1 and χ_2 and a bound on the deterministic term χ_3 are given by the following lemmas:

Lemma 1. *There exists a constant $C > 0$ such that if T is large enough, the following inequality holds:*

$$\mathbb{P}\left[\chi_1 > x\right] \leq \exp\left(-cT^2\left(\frac{xT^{\beta-2}}{C\sqrt{\log T}} - \log\frac{xT^{\beta-2}}{C\sqrt{\log T}} - 1\right)\right).$$

Lemma 2. *The following inequality holds:*

$$\mathbb{P}\left(\chi_2 > x\right) \leq \exp\left(-\frac{cx^2T}{4\left\|\boldsymbol{\Upsilon}\right\|_\infty^2\log T}(1+o(1))\right).$$

Lemma 3. $\chi_3 \le CT^{-\beta+2}\sqrt{\log T}$.

The proof of Lemma 1 is based on the use of a Lipschitz property of some function of λ. Observe that for $\beta > 2$, as $T \to \infty$, χ_1 is neglected as well as the deterministic term χ_3. We provide hereafter the proof of Lemma 2 since it sets the final rate of convergence. The proof is based on three main steps. The first step is the following result:

Lemma 4. *The following fact holds:*

$$\widehat{\Upsilon}^u_T(\lambda) = d_T(\lambda)^{\mathsf{H}}\left(\frac{V_T^{\mathsf{H}}V_T}{N} \odot B_T\right) d_T(\lambda)$$
$$\mathbb{E}\widehat{\Upsilon}^u_T(\lambda) = d_T(\lambda)^{\mathsf{H}}\left(R_T \odot B_T\right) d_T(\lambda)$$

where $d_T(\lambda) = 1/\sqrt{T}\left[1, e^{-\imath\lambda}, \ldots, e^{-\imath(T-1)\lambda}\right]^{\mathsf{T}}$, $\odot$ *is the Hadamard product of matrices, and*

$$B_T \triangleq \left[\frac{T}{T-|i-j|}\right]_{0\le i,j\le T-1}$$

satisfying $\|B_T\| \le \sqrt{2}T(\sqrt{\log T} + C)$ *with the constant* C *independent of* T.

The main difficulty of this proof relies on the fact that B_T has an unbounded spectral norm. The second step is the application of the following result used to obtain a more tractable expression of $\widehat{\Upsilon}^u_T(\lambda)$:

Lemma 5. *Let* x, $y \in \mathbb{C}^m$ *and matrices* $A, B \in \mathbb{C}^{m\times m}$. *Then*

$$x^{\mathsf{H}}(A \odot B)y = \mathrm{Tr}(D_x^{\mathsf{H}} A D_y B^{\mathsf{T}})$$

where $D_x = \mathrm{diag}(x)$ *and* $D_y = \mathrm{diag}(y)$.

From Lemma 4 and Lemma 5, after some simple manipulations, we get

$$\widehat{\Upsilon}^u_T(\lambda) = \frac{1}{N}\sum_{n=0}^{N-1} w_n^{\mathsf{H}} Q_T(\lambda) w_n$$

where $Q_T(\lambda) \triangleq R_T^{1/2} D_T(\lambda) B_T D_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}}$ with $D_T(\lambda) \triangleq \mathrm{diag}(d_T(\lambda))$ and $W_T = [w_0^{\mathsf{H}}, \ldots, w_{N-1}^{\mathsf{H}}]$. Now we deal with the matrix $Q(\lambda)$ with a smaller spectral norm (in comparison to that of B_T) bounded by $\sqrt{2}\|\boldsymbol{\Upsilon}\|_\infty (\log T)^{1/2} + C$ where the constant C is independent of λ. The third step of the proof consists in writing the eigenvalue decomposition of $Q_T(\lambda_i) = U_T \Sigma_T U_T^{\mathsf{H}}$ with $\Sigma_T = \mathrm{diag}(\sigma_0, \ldots, \sigma_{T-1})$. Since U_T is unitary and W_T has independent $\mathcal{CN}(0,1)$ elements, we get:

$$\widehat{\Upsilon}^u_T(\lambda) \stackrel{\mathcal{L}}{=} \frac{1}{N}\sum_{n=0}^{N-1}\sum_{t=0}^{T-1} |w_{n,t}|^2 \sigma_t$$

where $\stackrel{\mathcal{L}}{=}$ denotes equality in law. After applying the union bound, the rest of the proof is based on dealing with the term $\mathbb{P}\left(\widehat{\Upsilon}^u_T(\lambda_i) - \mathbb{E}\widehat{\Upsilon}^u_T(\lambda_i) > x\right)$ (for further details see [7]). From the Markov inequality and the Chernoff bound, for all $0 \le \tau < \inf_t \frac{N}{\sigma_t}$ and $x > 0$, we obtain:

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\Upsilon}^u_T(\lambda_i) - \mathbb{E}\widehat{\Upsilon}^u_T(\lambda_i) > x\right) \\
&\le \exp\Big(-\tau\Big(x + \sum_{t=0}^{T-1}\sigma_t\Big) - N\sum_{t=0}^{T-1}\log\Big(1 - \frac{\sigma_t\tau}{N}\Big)\Big) \\
&\le \exp\Big(-N\Big(\frac{\tau x}{N} - \frac{\tau^2}{2N^2}\sum_{t=0}^{T-1}\sigma_t^2\Big)\Big)\exp\Big(N\sum_{t=0}^{T-1}\Big|R_3\Big(\frac{\sigma_t\tau}{N}\Big)\Big|\Big)
\end{aligned}$$

where $\log(1-x) = -x - \frac{x^2}{2} + R_3(x)$ with $|R_3(x)| \le \frac{|x|^3}{3(1-\epsilon)^3}$ when $|x| < \epsilon < 1$. We shall manage this expression and to control the term $\exp(N\sum|R_3(\cdot)|)$. Choosing $\tau = \frac{axT}{\log T}$, $a > 0$ we get the result.

6. REFERENCES

[1] W. B. Wu and M. Pourahmadi, "Banding sample covariance matrices of stationary processes," *Statist. Sinica*, vol. 19, pp. 1755–1768, 2009.

[2] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Statist.*, vol. 36, pp. 199–227, 2008.

[3] H. Xiao and W. B. Wu, "Covariance matrix estimation for stationary time series," *Ann. Statist.*, vol. 40, pp. 466–493, 2012.

[4] T. T. Cai, C.-H. Zhang, and H. H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *Ann. Statist.*, vol. 38, pp. 2118–2144, 2010.

[5] T. T. Cai, Z. Ren, and H. H. Zhou, "Optimal rates of convergence for estimating Toeplitz covariance matrices," *Prob. Th. and Related Fields*, vol. 156, pp. 101–143, 2013.

[6] R. M. Gray, *Toeplitz and circulant matrices: A review*, Now Pub., 2006.

[7] J. Vinogradova, R. Couillet, and W. Hachem, "Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection," *submitted to IEEE Trans. Sig. Processing*, [Online] arXiv/1403.1243.

[8] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of statistical tests for single-source detection using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2400–2419, 2011.