Ratio-consistent estimation for long range dependent Toeplitz covariance with application to matrix data whitening[*]

Peng Tian and Jianfeng Yao

Department of Civil Engineering,
The University of Hong Kong
e-mail: tianpeng83@gmail.com

School of Data Science,
The Chinese University of Hong Kong (Shenzhen)
e-mail: jeffyao@cuhk.edu.cn

Abstract: We consider a data matrix $\mathbf{X} := \mathbf{C}_N^{1/2}\mathbf{Z}\mathbf{R}_M^{1/2}$ from a multivariate stationary process with a separable covariance function, where $\mathbf{C}_N$ is a $N \times N$ positive semi-definite matrix, $\mathbf{Z}$ a $N \times M$ random matrix of uncorrelated standardized white noise, and $\mathbf{R}_M$ a $M \times M$ Toeplitz matrix. Under the assumption of long range dependence (LRD), we re-examine the consistency of two toeplitzified estimators $\hat{\mathbf{R}}_M$ (unbiased) and $\hat{\mathbf{R}}_M^b$ (biased) for $\mathbf{R}_M$, which are known to be norm consistent with $\mathbf{R}_M$ when the process is short range dependent (SRD). However in the LRD case, some simulations suggest that the norm consistency does not hold in general for both estimators. Instead, a weaker *ratio consistency* is established for the unbiased estimator $\hat{\mathbf{R}}_M$, and a further weaker *ratio LSD consistency* is established for the biased estimator $\hat{\mathbf{R}}_M^b$. The main result leads to a consistent whitening procedure on the original data matrix $\mathbf{X}$, which is further applied to two real world questions, one is a signal detection problem, and the other is PCA on the space covariance $\mathbf{C}_N$ to achieve a noise reduction and data compression.

MSC2020 subject classifications: Primary 62M15; secondary 62H10, 15B52.
Keywords and phrases: Separable sample covariance matrix, long range dependence, whitening, Toeplitz matrix, high-dimensional PCA.

1. Introduction

Consider a random data matrix of the form

$$\mathbf{X} = \mathbf{C}_N^{1/2}\mathbf{Z}\mathbf{R}_M^{1/2} \qquad (1.1)$$

---

where $\mathbf{C}_N = (C_{N,i,j})$ and $\mathbf{R}_M = (R_{M,i,j})$ are $N \times N$ and $M \times M$ positive semi-definite Hermitian matrices, respectively, and $\mathbf{Z}$ is a white noise array of size $N \times M$. The matrix $\mathbf{X} = (X_{i,j})$ has a so-called separable covariance function, that is, $\text{Cov}(X_{i,j}, X_{i',j'}) = C_{N,i,i'}R_{M,j,j'}$. In other words, $\mathbf{C}_N$ and $\mathbf{R}_M$ represent the covariance function between rows and columns of $\mathbf{X}$, respectively. In this paper, by mimicking a stationary time series structure across the column vectors, we assume that $R_M = (r_{i-j})$ is a Toeplitz covariance matrix. Such data matrices appear in many applications, for example, as the noise part of a signal-plus-noise models in signal processing problems [27, 4, 23], or portfolio optimization problems [12], or as a stand-alone model in [26]. We note that if $\mathbf{C}_N$ is identity, the rows of $\mathbf{X}$ can be interpreted as i.i.d copies of a section of stationary process considered in [17, 16, 24], and if $\mathbf{C}_N$ is diagonal with i.i.d random entries independent of $\mathbf{Z}$, the rows of $\mathbf{X}$ can be interpreted as some i.i.d elliptically symmetric random vectors with Toeplitz scatter matrix $\mathbf{R}_M$, which were considered in e.g. [23, 12].

In the large dimensional context where both M and N are large, the estimation of $\mathbf{R}_M$ or $\mathbf{C}_N$ from the observable data $\mathbf{X}$ is a challenging question. Consider the following "sample covariance matrices"

$$\mathbf{S}_X := \frac{1}{M}\mathbf{X}\mathbf{X}^*, \quad \underline{\mathbf{S}}_X := \frac{1}{N}\mathbf{X}^*\mathbf{X}. \tag{1.2}$$

Then direct calculation shows that $\mathbb{E}(\mathbf{S}_X) = (M^{-1}\,\text{tr}\,\mathbf{R}_M)\mathbf{C}_N$ and $\mathbb{E}(\underline{\mathbf{S}}_X) = (N^{-1}\,\text{tr}\,\mathbf{C}_N)\mathbf{R}_M$. In other words, $\mathbf{S}_X$ and $\underline{\mathbf{S}}_X$ are unbiased (up to scalar factors) estimators of the matrices $\mathbf{C}_N$ and $\mathbf{R}_M$, respectively. However, the results from Random matrix theory (RMT) show that neither of them is consistent in the large dimensional regime, see e.g. [15, 22, 31].

Taking into account the Toeplitz structure of $\mathbf{R}_M$, it is possible to construct better estimators for $\mathbf{R}_M$. In [27], the authors considered two estimators

$$\hat{\mathbf{R}}_M := (\hat{r}_{i-j})_{1 \leq i,j \leq M}, \quad \hat{\mathbf{R}}_M^b := (\hat{r}_{i-j}^b)_{1 \leq i,j \leq M}, \tag{1.3}$$

where

$$\hat{r}_k := \frac{1}{M-|k|}\sum S_{i+k,i}\mathbb{1}_{\{1 \leq i, i+k \leq M\}}, \quad \hat{r}_k^b := \frac{1}{M}\sum S_{i+k,i}\mathbb{1}_{\{1 \leq i, i+k \leq M\}} \tag{1.4}$$

with $S_{i,j}$ the entries of $\underline{\mathbf{S}}_X$. Let $\xi_N = \text{tr}\,\mathbf{C}_N/N$. Note that $\hat{\mathbf{R}}_M$ is an unbiased estimator of $\xi_N\mathbf{R}_M$, whereas $\hat{\mathbf{R}}_M^b$ is biased. It is proved in [27, 4] that, if the entries of $\mathbf{Z}$ are i.i.d standard complex Gaussian and if the sequence $(r_k)_{k \in \mathbb{Z}}$ in $\mathbf{R}_M$ is absolutely summable, both estimators are (spectral) norm consistent, that is,

$$\left\|\hat{\mathbf{R}}_M - \xi_N\mathbf{R}_M\right\| \xrightarrow{\text{a.s}} 0, \quad \left\|\hat{\mathbf{R}}_M^b - \xi_N\mathbf{R}_M\right\| \xrightarrow{\text{a.s}} 0 \tag{1.5}$$

as $M, N \to \infty$ with $N/M \to c \in (0, \infty)$. Such consistent estimators can be used to whiten the correlation between the columns of $\mathbf{X}$ in order to facilitate the inference on $\mathbf{C}_N$, as done in [27] and other articles.

The absolute summability of the sequence $(r_k)_{k\in\mathbb{Z}}$ means that the columns of $\mathbf{X}$ are short range dependent (SRD), and this is crucial to the norm consistency (1.5). As an opposite scenario, the phenomenon of long range dependence (LRD) has been frequently observed in various fields like engineering and economic processes (see [9, 21, 7, 19] and the references therein). In this paper, we parameterize the LRD by $a \in (0,1)$, and study the consistency properties of $\hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_M^b$ in the LRD case. Our main results are

(i) The norm consistency (1.5) for the unbiased estimator $\hat{\mathbf{R}}_M$ continues to hold in the LRD case for the lower half $a \in (0,1/2)$ (the pro-SRD side), but simulation studies suggest that it may not hold for the upper half $a \in (1/2,1)$, see §4.1.

(ii) The unbiased estimator $\hat{\mathbf{R}}_M$ is ratio consistent in the sense that

$$\left\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2} - \xi_N\mathbf{I}\right\| \xrightarrow{\text{a.s.}} 0, \tag{1.6}$$

and consequently,

$$\left\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M^{1/2} - \sqrt{\xi_N}\mathbf{I}\right\| \xrightarrow{\text{a.s}} 0; \tag{1.7}$$

(iii) The biased estimator $\hat{\mathbf{R}}_M^b$ is not ratio consistent. We will prove that under the same conditions with $\xi_N = 1$, we have almost surely

$$\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M^b\mathbf{R}_M^{-1/2} - \mathbf{I}\| \not\to 0. \tag{1.8}$$

As a corollary, $\hat{\mathbf{R}}_M^b$ is not norm consistent, that is, almost surely,

$$\|\hat{\mathbf{R}}_M^b - \mathbf{R}_M\| \not\to 0. \tag{1.9}$$

(iv) A weaker ratio LSD consistency holds for $\hat{\mathbf{R}}_M^b$ in the sense that the empirical spectral distribution (ESD) of $\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M^b\mathbf{R}_M^{-1/2}$ converges to δ_1 (assuming $\xi_N \to 1$).

An immediate application of (ii) is to whiten the correlation $\mathbf{R}_M^{1/2}$ in the data $\mathbf{X}$ by multiplication of $\hat{\mathbf{R}}_M^{-1/2}$. Let

$$\mathbf{Y}_w := \mathbf{X}\hat{\mathbf{R}}_M^{-1/2}, \quad \text{and} \quad \mathbf{Y}_{Rid} := \xi_N^{-1/2}\mathbf{C}_N^{1/2}\mathbf{Z},$$

then by (1.7), under some proper conditions, we will have

$$\|\mathbf{Y}_w - \mathbf{Y}_{Rid}\| \xrightarrow{\text{a.s}} 0.$$

Note that the matrix $\mathbf{Y}_{Rid}$ is not observable but can be approximated by the whitened data matrix $\mathbf{Y}_w$, with which the inference on $\mathbf{C}_N$ becomes much easier, knowing that RMT contains many inference methods on $\mathbf{C}_N$ through the sample covariance matrix

$$\mathbf{S}_{Rid} := \frac{1}{M}\mathbf{Y}_{Rid}\mathbf{Y}_{Rid}^*,$$

see e.g. [30] and [13]. In this paper, we apply this result in a signal plus noise model to detect signals, reduce noise and compress the data using PCA. Some other applications based on the ratio consistency of $\hat{\mathbf{R}}_M$, such as the prediction of a multivariate separable time series, may be discussed in future works.

The result (iii) is striking because in the SRD case, $\hat{\mathbf{R}}_M^b$ is a better estimator for $\mathbf{R}_M$ than $\hat{\mathbf{R}}_M$ with a smaller variance. In the LRD case however, the bias is no longer negligible. In §4.2 we will illustrate by numeric simulations that this inconsistency may invalidate some subsequent whitening procedures. However, $\hat{\mathbf{R}}_M^b$ has a weaker ratio LSD consistency (iv) which may still be useful.

It is worthy of noticing that the entries $\hat{r}_k$ or $\hat{r}_k^b$, or some other banded/tapered estimators based on them are often used to estimate r_k or the periodogram. Many results have been established in this aspect, see e.g. [5, 28, 14, 29]. For matrix estimators, Ing et al. [11] established the norm consistency of an estimator for $\mathbf{R}_M^{-1}$ under some special LSD conditions. However, to the authors' best knowledge, the ratio consistency of matrix estimators for the purpose of whitening in the presence of LRD has not yet been established before.

We now describe some important technical innovations introduced in this paper as compared to the existing literature for the SRD case. The general structure of the proof of our main theorem, Theorem 2.2 follows [27]. But unlike the reference where the white noise matrix $\mathbf{Z}$ has i.i.d complex Gaussian entries, we also allow the rows $\mathbf{z}_i$ of $\mathbf{Z}$ to be uniformly distributed on a centered sphere in $\mathbb{R}^M$ or $\mathbb{C}^M$. This setting permits the rows of $\mathbf{X}$ to have more general elliptical distributions. Thus the M columns of $\mathbf{Z}$ are uncorrelated but dependent and new tools are needed in various moment estimation involving these noise variables such as

$$\mathbb{E}e^{\tau \sum_m \sigma_m |Z_{1,m}|^2}$$

where $Z_{1,m}, m = 1, \ldots, M$ are the elements of the first row.

In the proof of Theorem 2.2, an accurate upper bound for a trace of non commutative product of several Toeplitz matrices (of type $\text{tr}(\mathbf{D}_M(\theta)\mathbf{R}_M \mathbf{D}_M(\theta)^*\mathbf{B}_M)^2$, where $\mathbf{R}_M, \mathbf{B}_M$ are Toeplitz, and $\mathbf{D}_M(\theta)$ is diagonal depending on $\theta \in (0, 2\pi)$) is crucial. In [27], a global bound was obtained using linear algebraic method thanks to the boundedness of $\|\mathbf{R}_M\|$. In this paper, since $\|\mathbf{R}_M\| \to \infty$, the global bound is not sufficient. We rewrite the same trace as a two dimensional harmonic integral and then estimate its bound in terms of M and the underlying spectral density $f(\theta)$. This is not trivial since f has a power singularity at 0.

Notations

Matrices are denoted by bold capital characters, row or column vectors are denoted by bold characters. For $x \in \mathbb{R}$, δ_x denotes the Dirac measure at x. For a Hermitian $N \times N$ matrix $\mathbf{S}$, its eigenvalues are denoted as $\lambda_1(\mathbf{S}) \geq \cdots \geq \lambda_N(\mathbf{S})$, and $\mu^{\mathbf{S}} := N^{-1} \sum_{i=1}^N \delta_{\lambda_i(\mathbf{S})}$ denotes the ESD of $\mathbf{S}$. The largest and smallest eigenvalues of $\mathbf{S}$ are also denoted by $\lambda_{\max}(\mathbf{S})$ and $\lambda_{\min}(\mathbf{S})$, respectively. For a matrix $\mathbf{A}$, $A_{i,j}$ stands for its ith row and jth column element, $\|\mathbf{A}\|$ and

$\|\mathbf{A}\|_F$ its operator norm and Frobenius norm, respectively, and $\mathbf{A}^*$ its conjugate transpose. The spectrum of a square matrix $\mathbf{A}$ is denoted by $\mathrm{Spec}(\mathbf{A})$. For a function f, $\|f\|_1$ and $\|f\|_\infty$ stand for its L^1 and L^∞ norm, respectively. If a, b are two elements of a Hilbert space, we denote their inner product as $\langle a, b \rangle$. The symbol K denotes a constant which may take different values from one place to another. If several constants are needed in one expression, we will denote them by $K_1, K_2, \ldots$. For two sequences of positive numbers a_n and b_n, $a_n \lesssim (\gtrsim) b_n$ means that there exists a constant $K > 0$ such that $a_n \leq (\geq) K b_n$ for all n, and $a_n \asymp b_n$ means that there exist constants $0 < K_1 < K_2$ such that $K_1 b_n \leq a_n \leq K_2 b_n$ for all n. The underlying constants may depend on the spectral density f defined in the assumption **A5** below, but do not depend on any other variables in this article. The notation $a_n \gg b_n$ (resp. $a_n \ll b_n$) means that $a_n / b_n \to \infty$ (resp. $a_n / b_n \to 0$).

Organization

In §2, we state the main results. In §3, we develop applications to signal detection and high-dimensional PCA for the data matrix $\mathbf{X}$. In §4, we provide in 4.1 numeric simulations to show the norm inconsistency of $\hat{\mathbf{R}}_M$, and to illustrate of the ratio inconsistency of $\hat{\mathbf{R}}_M^b$ with its impact to the whitening procedure in 4.2. We prove our main results Theorem 2.2 and Proposition 2.4 in §5 and §6, respectively. The other proofs are given in the appendices.

2. Main results

2.1. Model and assumptions

We consider the random data matrix $\mathbf{X}$ in (1.1) with the following assumptions.

A1 The rows $\mathbf{z}_i$ of $\mathbf{Z}$ are i.i.d real or complex random vectors, either standard normal distributed, or distributed as $\sqrt{M}\mathbf{u}$ where $\mathbf{u}$ follows the uniform (Haar) measure on the unit sphere in $\mathbb{R}^M$ or $\mathbb{C}^M$.

A2 The matrices $\mathbf{C}_N$ are nonnegative and diagonal, i.e. $\mathbf{C}_N = \mathrm{diag}(c_1, \ldots, c_N)$ where $c_n \geq 0$ and c_n may also depend on N for $n = 1, \ldots, N$. Moreover there exist constants $C > 0$ and $\kappa > 0$ such that

$$\frac{1}{N} \mathrm{tr}\, \mathbf{C}_N^2 \leq C, \quad \|\mathbf{C}_N\| \leq \kappa \log M.$$

A3 The Toeplitz matrices $\mathbf{R}_M = (r_{i-j})_{i,j=1}^M$ have a positive spectral density $f \in L^1(-\pi, \pi)$ which is bounded in any set of the form $[-\pi, \pi] \backslash (-\delta, \delta)$ with $\delta > 0$.

A4 The spectral density f is bounded away from 0:

$$\underset{\theta \in (-\pi, \pi)}{\mathrm{ess\,inf}}\, f(\theta) > 0.$$

A5 The spectral density f is even and has the following asymptotic behavior near 0:

$$f(x) = \frac{L(|x|^{-1})}{|x|^a}$$

for $x \in [-\pi, \pi] \backslash \{0\}$ where $a \in (0, 1)$ and L defined in $[\pi^{-1}, \infty)$ is a slowly varying function at ∞.

The assumptions on $\mathbf{Z}$ and $\mathbf{C}_N$ allow the matrix $\mathbf{X}$ to cover two types of models. When the matrix $\mathbf{Z}$ has i.i.d standard real (resp. complex) Gaussian entries and $\mathbf{C}_N$ is a real symmetric (resp. complex Hermitian) deterministic matrix, each column of $\mathbf{X}$ is distributed as $\mathcal{N}(0, \mathbf{C}_N)$ and the correlation between two columns $\mathbf{x}_i, \mathbf{x}_j$ is $\gamma(i-j)\mathbf{C}_N$, which is the product of a scalar $\gamma(i-j)$ depending only on the difference of their indices $i - j$ and a fixed matrix $\mathbf{C}_N$. Then $\mathbf{X}$ represents a N-dimensional stationary Gaussian process with a separable correlation structure. Note that in this case, the multivariate process $\mathbf{X}$ is a linear transform by $\mathbf{C}_N^{1/2}$ of N i.i.d samples of a univariate stationary process. By the orthogonal (resp. unitary) invariance of the columns in $\mathbf{X}$, we can assume that $\mathbf{C}_N$ is diagonal without modifying the distribution of $\mathbf{S}_X, \hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_M^b$ defined in (1.2)–(1.4). When the rows $\mathbf{z}_i$ are distributed as $\sqrt{M}\mathbf{u}$ where $\mathbf{u}^\top$ follows the uniform (Haar) measure on the unit sphere in $\mathbb{R}^M$ or $\mathbb{C}^M$, and $\mathbf{C}_N = \mathrm{diag}(\nu_1, \ldots, \nu_N)$ with ν_i some i.i.d nonnegative random variables, independent of $\mathbf{z}_i$, we can write the rows of $\mathbf{X}$ as

$$\mathbf{x}_i = \sqrt{\nu_i}\mathbf{z}_i\mathbf{R}_M^{1/2}.$$

Then $\mathbf{x}_i^\top$ has an elliptical distribution, and the data matrix $\mathbf{X}$ represents a set of i.i.d samples of elliptical random vectors with Toeplitz scatter matrix $\mathbf{R}_M$. Because $\mathbf{C}_N$ is independent of $\mathbf{Z}$, we can treat $\mathbf{C}_N$ as deterministic by standard conditioning arguments, and our results are still applicable.

Recall that the spectral density of a sequence of Toeplitz matrices $\mathbf{R}_M = (r_{i-j})$ is a function $f \in L^1(-\pi, \pi)$ whose Fourier coefficients are r_k:

$$r_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-\mathbf{i}kx}\,\mathrm{d}x.$$

If f is real, then $\mathbf{R}_M$ is Hermitian; if f is positive, then $\mathbf{R}_M$ is positive definite; if f is positive and even, then $\mathbf{R}_M$ is real symmetric and positive definite. We will consider f as a 2π-periodic function so that $f(x)$ is well defined by periodicity for all real x. Note that the assumption **A4** ensures that the smallest eigenvalue of $\mathbf{R}_M$ is positive and bounded away from 0, thus $\mathbf{R}_M$ is invertible for all M with $\|\mathbf{R}_M^{-1}\|$ bounded. If $\mathbf{R}_M$ is the autocovariance matrix of a stationary process and satisfies **A3**, **A5**, then the process is LRD by [18, Definition 2.1.5 (Condition IV)].

Recall the definition of $\mathbf{S}_X$ in (1.2), the two estimators $\hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_M^b$ in (1.3)–(1.4), and $\xi_N = N^{-1}\,\mathrm{tr}\,\mathbf{C}_N$.

2.2. Consistency properties of the unbiased estimator $\hat{\mathbf{R}}_M$

In this subsection we study the consistency of the unbiased estimator $\hat{\mathbf{R}}_M$. We first point out that using a norm bound of $\mathbf{R}_M$ given in Lemma 5.6 below, and simply adapting the proof of [27, Theorem 2], one can prove the following large deviation bound.

Proposition 2.1. *Assume that **A1**, **A3** and **A5** hold with $\mathbf{Z}$ having i.i.d complex Gaussian entries, $\mathbf{C}_N$ bounded in spectral norm, and $0 < a < 1/2$. Also assume that $N/M \to c \in (0, \infty)$. Then there exists a constant $K > 0$, such that for any fixed $x > 0$, we have*

$$\mathbb{P}\left(\left\|\hat{\mathbf{R}}_M - \xi_N \mathbf{R}_M\right\| > x\right) \leq \exp\left(-\frac{KcM^{1-2a}x^2}{\|\mathbf{C}_N\|^2 L^2(M)\log M}(1 + o(1))\right). \quad (2.1)$$

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

Proof. By checking carefully the proof of [27, Theorem 2], we note that we can adapt it by replacing all the occurrences of the infinite-norm of the spectral density ($\|\mathbf{\Upsilon}\|_\infty$ in [27]) with the spectral norm of the matrix $\|\mathbf{R}_M\|$, and also re-analyzing the contribution of the term $\|\mathbf{R}_M\|$ where this term was previously bounded in [27]. We also need to consider the contribution of $\mathbf{C}_N$, which will introduce the factor $\|\mathbf{C}_N\|^2$ in the denominator of the exponential bound there. Therefore, we have the following estimate:

$$\mathbb{P}\left(\left\|\hat{\mathbf{R}}_M - \xi_N \mathbf{R}_M\right\| > x\right) \leq \exp\left(-\frac{KNx^2}{\|\mathbf{C}_N\|^2\|\mathbf{R}_M\|^2\log M}(1 + o(1))\right). \quad (2.2)$$

Taking into account the bound $\|\mathbf{R}_M\| \asymp M^a L(M)$ given in Lemma 5.6 below, the result follows. The details are omitted. $\quad\square$

This proposition implies that in some LRD cases, where $\mathbf{R}_M$ satisfies **A5** with $0 < a < 1/2$, the unbiased estimator $\hat{\mathbf{R}}_M$ is still norm consistent. However, this result has the following defects. Firstly, the result does not cover the case $1/2 \leq a < 1$ (In fact we conjecture that in this case the norm consistency does not hold. This will be supported by simulations with an heuristic argument in §4.1). However in some applications such as whitening, the convergence of the ratio $\mathbf{R}_M^{1/2}\hat{\mathbf{R}}_M^{-1}\mathbf{R}_M^{1/2}$ to some scaled identity $\xi\mathbf{I}$ suffices. Secondly, even for the norm consistent case, we can see that the convergence rate ensured by this proposition is no better than $\sqrt{\log(M)}L(M)/M^{1/2-a}$, which depends on a and gets worse and worse when a approaches $1/2$. Meanwhile, since $\mathbf{R}_M^{-1}$ is merely bounded under **A4**, the proposition provides the same convergence rate for the ratio $\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2}$.

These facts motivate us to establish the following large deviation bound for the ratio.

Theorem 2.2. *Under the assumptions **A1**–**A5**, for any sequence x_M satisfying $M^{-\gamma} \lesssim x_M \lesssim C/\kappa$ for some $\gamma > 0$, there exist $K > 0$, such that for large enough*

M and any $N \geq 1$, we have

$$\mathbb{P}\left(\left\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2} - \xi_N\mathbf{I}\right\| > x_M\right) \leq 2\exp\left(-\frac{Nx_M^2}{CK\log^2 M} + \beta\log M\right).$$

(2.3)

where β is any integer larger than $2 + a + \gamma$.

Remark 2.1. The inequality (2.3) literally holds for large enough M and all $N \geq 1$. Meaningful asymptotic results can be obtained by considering $M, N \to \infty$ with different regimes. We leave the freedom of specifying a precise exploding speed of M, N to specific context of applications. For example, in order to get the almost sure convergence

$$\left\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2} - \xi_N\mathbf{I}\right\| \xrightarrow{\text{a.s}} 0,$$

(2.4)

we should assume $M \to \infty$ and $N = N(M) \gg \log^3 M$.

As another example, if the target is some precise convergence rate, we will consider small x_M such as $x_M = A(\log^{\frac{3}{2}} M)/\sqrt{N}$ with a large enough constant A. Then from (2.3), we have

$$\mathbb{P}\left(\left\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2} - \xi_N\mathbf{I}\right\| > x_M\right) \leq 2M^{-(\frac{A^2}{CK} - \beta)}.$$

(2.5)

So if we take $N = N(M) \lesssim M^{2\gamma}$ for some $\gamma > 0$ (which is required by the statement of Theorem 2.2) and a large enough A (to make the RHS of (2.5) summable), almost surely, for large enough M,

$$\left\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2} - \xi_N\mathbf{I}\right\| \leq \frac{A\log^{\frac{3}{2}} M}{\sqrt{N}}.$$

(2.6)

Of course, if we want this bound to vanish, we need $N \gg \log^3 M$.

From this theorem, in order to accurately estimate the autocovariance matrix $\mathbf{R}_M$, we need that the dimension N is also large enough. This is natural since in the Gaussian case, N is the number of i.i.d. copies of a univariate stationary process (the rows of $\mathbf{ZR}_M^{1/2}$) we used to construct $\mathbf{X}_M$ via the linear transform $\mathbf{C}_N^{1/2}$.

As a corollary of Theorem 2.2, when N, M are of the same order, the matrix $\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/2}$ is equivalent to $\sqrt{\xi_N}\mathbf{I}$. This fact is useful when we want to rebuild the uncorrelated data $\mathbf{Y} = \mathbf{C}_N^{1/2}\mathbf{Z}$ by whitening $\mathbf{R}_M^{1/2}$.

Corollary 2.3. *Under the same assumptions as in Theorem 2.2, assume moreover that ξ_N is lower bounded from 0. Then as $M, N \to \infty$ with $N \asymp M$, we have almost surely*

$$\left\|\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/2} - \sqrt{\xi_N}\mathbf{I}\right\| \to 0.$$

(2.7)

Using the ratio consistency of $\hat{\mathbf{R}}_M$, we develop a consistent whitening procedure on $\mathbf{X}$, which is further applied to signal detection, noise reduction and data compression. See §3.

When $\mathbf{C}_N$ is random and independent of $\mathbf{Z}$ such that $\mathbf{A2}$ is almost surely satisfied for large enough M, N, the almost sure convergence (2.4) also holds. For example, if $\mathbf{C}_N = \mathrm{diag}(\nu_1, \ldots, \nu_N)$ with $(\nu_i)_{i \in \mathbb{N}}$ a sequence of i.i.d sub-exponential (in the sense that $\mathbb{P}(|\nu_1| > t) \leq K_1 e^{-t/K_2}$ with some $K_1, K_2 > 0$ for any $t > 0$) positive random variables satisfying $\mathbb{E}|\nu_1|^2 = 1$, then

$$\frac{1}{N} \sum_{i=1}^{N} \nu_i^2 \xrightarrow[N \to \infty]{} 1,$$

and there exists $\kappa > 0$ such that almost surely for large enough N,

$$\max_{1 \leq i \leq N} \{\nu_i\} \leq \kappa \log N.$$

In this case (2.4) holds as $M, N \to \infty$ with $N \asymp M$.

2.3. Consistency properties of the biased estimator $\hat{\mathbf{R}}_M^b$

In the SRD case, the biased estimator $\hat{\mathbf{R}}_M^b$ has several advantages over $\hat{\mathbf{R}}_M$. Firstly, it is structurally positive semi-definite (see Lemma 3 of [27]). Secondly, it has smaller deviation from its expectation (the fluctuation rate is lower than $N^{-\alpha}$ for any $\alpha < 1$, see [4]), because the inaccuracy of the elements near the top-right and bottom-left corners is more reduced in $\hat{\mathbf{R}}_M^b$ than in $\hat{\mathbf{R}}_M$.

However in the LRD case, $\hat{\mathbf{R}}_M^b$ is no longer consistent with $\mathbf{R}_M$, even in the sense of ratio consistency. In fact, $\hat{\mathbf{R}}_M^b$ is ratio consistent with its expectation

$$\mathbf{R}_M^b := \mathbb{E}\hat{\mathbf{R}}_M^b = \left(\left(1 - \frac{|i-j|}{M}\right) r_{i-j} \right)_{i,j=1}^{M},$$

whose difference from $\mathbf{R}_M$ is no longer negligible in the LRD case. This is precisely established below.

Proposition 2.4. *Let $\mathbf{X}$ be defined in* (1.1) *with $\mathbf{A1}$–$\mathbf{A5}$ hold. Suppose that $\xi_N = N^{-1} \mathrm{tr}\, \mathbf{C}_N = 1$. Then as $M \to \infty$ with $N = N(M) \gg \log^3 M$, almost surely, $\hat{\mathbf{R}}_M^b$ is ratio consistent with $\mathbf{R}_M^b$:*

$$\|(\mathbf{R}_M^b)^{-1/2} \hat{\mathbf{R}}_M^b (\mathbf{R}_M^b)^{-1/2} - \mathbf{I}\| \to 0, \tag{2.8}$$

but not with $\mathbf{R}_M$:

$$\|\mathbf{R}_M^{-1/2} \hat{\mathbf{R}}_M^b \mathbf{R}_M^{-1/2} - \mathbf{I}\| \nrightarrow 0. \tag{2.9}$$

Note that the inconsistency (2.9) is a special phenomenon caused by LRD, because in the SRD case, as long as $\mathbf{A4}$ holds, $\|\mathbf{R}_M\|$ and $\|\mathbf{R}_M^{-1}\|$ are both bounded, then the "ratio consistency" and the "norm consistency" are equivalent.

The proof of Proposition 2.4 also shows that the inconsistency (2.9) caused by LRD affects not only the biased estimator $\hat{\mathbf{R}}_M^b$, but more generally a large class

of tapered estimators of $\mathbf{R}_M$. Analogous to $\hat{\mathbf{R}}_M^b$, we often taper the estimates of r_k for values of k close to $M-1$ in order to reduce the inaccuracy of these estimates. But in the LRD case, such tapering often modifies the asymptotic behavior of the largest eigenvalue of the resulting estimator for $\mathbf{R}_M$, which in turn destroys its ratio consistency, see the proof of Proposition 2.4.

Despite this ratio inconsistency of $\hat{\mathbf{R}}_M^b$ with $\mathbf{R}_M$, we find that only a small part of the eigenvalues of $\mathbf{R}_M^{-1}\hat{\mathbf{R}}_M^b$ deviate from 1. In fact, we will establish the ratio LSD consistency between the two matrices.

Proposition 2.5. *Let* $\mathbf{X}$ *be defined in* (1.1) *with* ***A1***–***A5*** *hold. Suppose that* $\xi_N \to 1$ *as* $N \to \infty$. *Then as* $M \to \infty$ *with* $N = N(M) \gg \log^3 M$, *almost surely the ESD of* $\mathbf{R}_M^{-1}\hat{\mathbf{R}}_M^b$ *converges weakly to* δ_1.

Thanks to the ratio LSD consistency of $\hat{\mathbf{R}}_M^b$ with $\mathbf{R}_M$, $\hat{\mathbf{R}}_M^b$ can still serve as a good approximation for $\mathbf{R}_M$ in certain circumstances, see §4.2.

We end this section by recapitulating the consistency properties of estimators $\hat{\mathbf{R}}_M$ and $\hat{\mathbf{R}}_M^b$ with $\mathbf{R}_M$, in the cases of SRD and LRD, respectively.

TABLE 1
Recapitulation of consistency properties

	SRD	LRD
$\hat{\mathbf{R}}_M$	norm consistent	norm consistent when $0 < a < 1/2$
		*norm inconsistent when $1/2 < a < 1$
	ratio consistent	ratio consistent
	ratio LSD consistent	ratio LSD consistent
$\hat{\mathbf{R}}_M^b$	norm consistent	norm inconsistent
	ratio consistent	ratio inconsistent
	ratio LSD consistent	ratio LSD consistent

* is only supported by numerical studies. See Section 4.1.

3. Applications to matrix data whitening

Suppose that we have observed a data matrix $\mathbf{X} = \mathbf{C}_N^{1/2}\mathbf{Z}\mathbf{R}_M^{1/2}$ with M, N large but of the same order, and we want to detect the spike eigenvalues and the associated eigenvectors of $\mathbf{C}_N$. If $\mathbf{R}_M$ is identity, the data matrix becomes $\mathbf{Y} = \mathbf{C}_N^{1/2}\mathbf{Z}$, and from RMT results, we can find the spike eigenvalues (usually the extreme ones in applications) of

$$\mathbf{S}_Y = \frac{1}{M}\mathbf{Y}\mathbf{Y}^* = \frac{1}{M}\mathbf{C}_N^{1/2}\mathbf{Z}\mathbf{Z}^*\mathbf{C}_N^{1/2},$$

and calculate the spike eigenvalues of $\mathbf{C}_N$ using the formula relating the spike eigenvalues of $\mathbf{C}_N$ and $\mathbf{S}_Y$, see [30, Chapter 11] for more details. However, for general matrix $\mathbf{R}_M$, especially when the underlying stationary process is LRD, the above method fails because the information relevant to $\mathbf{C}_N$ is mixed with the covariance matrix $\mathbf{R}_M$. In fact from [25, Corollary 2.1] we know that if **A3**,**A5** hold and if in addition, the LSD of $\mathbf{C}_N$ weakly converges, then for any

fixed $m \geq 1$, the m largest eigenvalues of $\mathbf{S}_X = M^{-1}\mathbf{X}\mathbf{X}^*$ are asymptotically equivalent to

$$\frac{\operatorname{tr}\mathbf{C}_N}{M}\{\lambda_1(\mathbf{R}_M), \ldots, \lambda_m(\mathbf{R}_M)\},$$

as $N, M \to \infty$ and $N/M \to c \in (0, \infty)$. Thus, even if $\mathbf{C}_N$ is bounded, the largest eigenvalues of $\mathbf{S}_X$ tend to infinity following those of $\mathbf{R}_M$, and only the summary statistic $\operatorname{tr}\mathbf{C}_N$ appears in their first order limits. It is thus very difficult, if not impossible, to estimate the number and locations of the spiked eigenvalues, and the associated eigenvectors of $\mathbf{C}_N$. This breaks down any PCA on the original data based on $\mathbf{C}_N$.

Using Theorem 2.2 and Corollary 2.3, we can whiten and remove the time correlation $\mathbf{R}_M^{1/2}$ from the data matrix $\mathbf{X}$ by multiplying it with $\hat{\mathbf{R}}_M^{-1/2}$. Consider the whitened data matrix

$$\mathbf{Y}_w := \mathbf{X}\hat{\mathbf{R}}_M^{-1/2}$$

and the whitened sample covariance matrix

$$\mathbf{S}_w := \frac{1}{M}\mathbf{Y}_w\mathbf{Y}_w^* = \frac{1}{M}\mathbf{C}_N^{1/2}\mathbf{Z}(\mathbf{R}_M^{1/2}\hat{\mathbf{R}}_M^{-1}\mathbf{R}_M^{1/2})\mathbf{Z}^*\mathbf{C}_N^{1/2}. \tag{3.1}$$

Consider also the corresponding matrices with $\mathbf{R}_M$ equaling to identity and $\mathbf{C}_N$ normalized by ξ_N:

$$\mathbf{Y}_{Rid} := \xi_N^{-1/2}\mathbf{Y}, \quad \mathbf{S}_{Rid} := \frac{1}{M}\mathbf{Y}_{Rid}\mathbf{Y}_{Rid}^* = \frac{1}{\xi_N}\underline{\mathbf{S}}_Y. \tag{3.2}$$

Our theory ensures that the impact of the covariance matrix $\mathbf{R}_M$ is properly removed from $\mathbf{X}$ so that the matrix $\mathbf{S}_w$ is close to $\mathbf{S}_{Rid}$ in spectral norm.

Proposition 3.1. *Let $\mathbf{A1}$–$\mathbf{A5}$ hold with $\mathbf{C}_N$ bounded in spectral norm and $\xi_N = \operatorname{tr}\mathbf{C}_N/N \geq \epsilon > 0$. Then as $N, M \to \infty$ with $N/M \to c \in (0, \infty)$, we have*

$$\|\mathbf{S}_w - \mathbf{S}_{Rid}\| \xrightarrow{\text{a.s}} 0.$$

The proposition ensures that many statistical methods for the standard covariance matrix $\mathbf{S}_{Rid} = \xi_N^{-1}\mathbf{S}_Y$ are applicable to the whitened matrix $\mathbf{S}_w$. In the following, we develop two statistical applications of this whitening procedure with the following deformed "signal plus noise" model.

Let $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M)$ with

$$\mathbf{y}_j = \mathbf{A}\mathbf{m}_j + \sigma\mathbf{n}_j, \tag{3.3}$$

where $\mathbf{A}$ is a $N \times p$ matrix with $p \geq 0$ a fixed integer, $\sigma > 0$ and $\mathbf{m}_j \sim \mathcal{CN}(0, \mathbf{I}_p)$, $\mathbf{n}_j \sim \mathcal{CN}(0, \mathbf{I}_N)$ are standard complex Gaussian vectors, independent of each other and across i. The vectors $\mathbf{y}_j$ represent antenna data, where $\mathbf{A}\mathbf{m}_j$ represent the signal and $\sigma\mathbf{n}_j$ the noise. A little calculation shows that the matrix $\mathbf{Y}_{Rid}$ can be written in the form

$$\mathbf{Y} = \mathbf{C}_N^{1/2}\mathbf{Z}, \tag{3.4}$$

where $\mathbf{C}_N = \mathrm{Cov}(\mathbf{y}_i) = \mathbf{A}\mathbf{A}^* + \sigma^2 \mathbf{I}_N$, and $\mathbf{Z}$ is a $N \times M$ matrix with i.i.d standard complex Gaussian entries. Note that in this case $\mathbf{C}_N$ are unitarily similar to the diagonal matrix

$$\sigma^2 \, \mathrm{diag}(\alpha_1, \ldots, \alpha_p, 1, \ldots, 1), \tag{3.5}$$

where $\alpha_i = 1 + \frac{\lambda_i(\mathbf{A}\mathbf{A}^*)}{\sigma^2} =: 1 + \beta_i$ are signal strengths, and $\beta_i := \frac{\lambda_i(\mathbf{A}\mathbf{A}^*)}{\sigma^2}$ are signal-to-noise ratios.

Now suppose that the data matrix $\mathbf{Y}$ is "polluted" during its transmission which takes the form of a LRD time series, and only the matrix

$$\mathbf{X} = \mathbf{Y}\mathbf{R}_M^{1/2} \tag{3.6}$$

is observable, where $\mathbf{R}_M$ is a Toeplitz matrix satisfying **A3**, **A4**, **A5**. For the ease of numerical simulations, we let the entries of $\mathbf{R}_M$ be

$$r_{i-j} = \frac{1}{(1 + |i - j|)^{1-a}}. \tag{3.7}$$

By [18, Proposition 2.2.14], $\mathbf{R}_M$ satisfies **A3** and **A5**. By [32, Theorem 1.5, Chapter V], $\mathbf{R}_M$ also satisfies **A4**. Indeed, if the diagonal entry r_0 is large enough such that $(r_n)_{n \geq 0}$ is convex, the spectral density of $\mathbf{R}_M$ is nonnegative. The minimal value of such r_0 is $2^a - 3^{a-1} < 1$ for $0 < a < 1$. Thus when we take $r_0 = 1$, the spectral density f is larger than $1 - 2^a + 3^{a-1} > 0$.

In §3.1, we detect the number of spikes in $\mathbf{C}_N$, and estimate the signal strength α_i. In §3.2, we proceed a PCA on the data matrix $\mathbf{X}$, or on the whitened data $\mathbf{Y}_w$ to reduce the noise, and the obtained matrices are of rank p, realizing a compression of the original data matrix $\mathbf{X}$.

3.1. Detection of the number of signals and estimation of their strengths α_i

As an immediate application of the asymptotic proximity between $\mathbf{S}_w$ and $\mathbf{S}_{Rid}$, we propose two algorithms, 1) to identify the number of spikes p from $\mathbf{S}_w$, and 2) to estimate the spikes $\alpha_1, \ldots, \alpha_p$. We assume that the spikes $\alpha_1, \ldots, \alpha_p$ are simple. The following proposition is the theoretical base of our algorithms. It is a corollary of Proposition 3.1 and [30, Theorem 2.9, Theorem 11.3].

Proposition 3.2. *Let the rows of* $\mathbf{Y}$ *be defined in* (3.3) *and assume that the conditions of Proposition 3.1 hold. Then the LSD of* $\mathbf{S}_w$ *is the Marčenko-Pastur distribution*

$$\mathbb{P}_{MP}(\mathrm{d}\lambda) := \frac{\sqrt{[(\lambda^+ - \lambda)(\lambda - \lambda^-)]_+}}{2\pi c \lambda} \mathbb{1}_{\lambda \in [\lambda^-, \lambda^+]} \, \mathrm{d}\lambda + \left(1 - c^{-1}\right) \delta_0(\mathrm{d}\lambda) \mathbb{1}_{\{c > 1\}}, \tag{3.8}$$

where $\lambda^\pm = (1 \pm \sqrt{c})^2$. *Furthermore, each* α_i *larger than* $1 + \sqrt{c}$ *produces a spiked eigenvalue of* $\mathbf{S}_w$, *respectively, converging to*

$$\lambda_i = \alpha_i + c\alpha_i/(\alpha_i - 1). \tag{3.9}$$

The proof will be omitted since it can be easily done by noticing that in this particular case, $\mathbf{S}_{Rid} = M^{-1}\tilde{\mathbf{C}}_N^{1/2}\mathbf{Z}^*\mathbf{Z}\tilde{\mathbf{C}}_N^{1/2}$, where $\tilde{\mathbf{C}}_N = \frac{\mathbf{C}_N}{\xi_N}$ is the normalized matrix unitarily similar to

$$\frac{\mathrm{diag}(\alpha_1, \ldots, \alpha_p, 1, \ldots, 1)}{\xi_N'}$$

with

$$\xi_N' := \frac{\sum_{i=1}^p \alpha_i + N - p}{N} \xrightarrow[N\to\infty]{} 1.$$

In order to estimate p, we calculate the eigenvalues of $\mathbf{S}_w$ denoted as

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_N.$$

Note that for current model setting, the ratio $\hat{\lambda}_j/\hat{\lambda}_{j+1}$ is larger than $1 + \varepsilon > 1$ for $j < p$, and is close to 1 for $j \geq p$. We thus choose three numbers $\gamma_1, \gamma_2, \gamma_3$ such that the event

$$\left\{\frac{\hat{\lambda}_{p+i}}{\hat{\lambda}_{p+i+1}} < \gamma_i, \, i = 1, 2, 3\right\}$$

occurs with probability no less than some threshold (e.g. 99.9%). We will determine γ_i by Monte-Carlo simulation. Then

$$\hat{p} := \inf\{k \geq 0 \, : \, \hat{\lambda}_{k+1}/\hat{\lambda}_{k+2} < \gamma_1, \, \hat{\lambda}_{k+2}/\hat{\lambda}_{k+3} < \gamma_2, \, \hat{\lambda}_{k+3}/\hat{\lambda}_{k+4} < \gamma_3\}$$

is our proposed estimate for p.

Remark 3.1. Theoretically, we can use

$$\hat{p}' := \inf\{k \geq 0 \, : \, \hat{\lambda}_{k+1}/\hat{\lambda}_{k+2} < \gamma_1\}$$

as an estimate of p. However even though α_i's are distinct, the corresponding sample eigenvalues $\hat{\lambda}_i$ can still get so close that the algorithm may terminate prematurely and $\hat{p}'$ tends to under-estimate the real p. The triple tests in $\hat{p}$ reinforce the robustness of the estimator against such situation.

Once we have estimated p, we can further estimate the values of $\alpha_1, \ldots, \alpha_p$. According to (3.9), a rough estimation is given by

$$\hat{\alpha}_i := \frac{(1 - c + \hat{\lambda}_i) + \sqrt{(1 - c + \hat{\lambda}_i)^2 - 4\hat{\lambda}_i}}{2}. \tag{3.10}$$

However, note that $\mathbf{S}_{Rid}$ is different from the standard spike model by a factor of ξ_N^{-1}. The estimation error may be considerable if there are many large spikes. We can estimate ξ_N in order to correct the error. Note that the exact asymptotic location of $\hat{\lambda}_{p+1}$ is at $\xi_N^{-1}(1 + \sqrt{c})^2$, so we let

$$\hat{\xi}_N := \frac{(1 + \sqrt{c})^2}{\hat{\lambda}_{\hat{p}+1}}.$$

And then the corrected estimation of α_i is

$$\hat{\alpha}_i := \frac{(1 - c + \hat{\lambda}_i \hat{\xi}_N) + \sqrt{(1 - c + \hat{\lambda}_i \hat{\xi}_N)^2 - 4\hat{\lambda}_i \hat{\xi}_N}}{2}. \tag{3.11}$$

We will see in the simulation that the accuracy of this estimator is satisfactory.

We do some numerical simulations to test the efficiency and robustness of the two estimation procedures. We take $M = 500, N = 833$ ($c = 0.6$) and $a = 0.7$. Although the number p and the spikes α_i are assumed to be fixed in the description of model, in order to add some challenge to the test, we pick p randomly following Poisson distribution with parameter 4, and then $\alpha_1, \ldots, \alpha_p$ are independently and uniformly positioned in the interval $[3, 10]$. Let $\sigma^2 = 1$. With this construction of $\mathbf{C}_N$, we know that if $p \geq 1$, almost surely the spikes $\alpha_1, \ldots, \alpha_p$ are simple.

Next we determine the three numbers $\gamma_1, \gamma_2, \gamma_3$ defined in the description of the algorithm. Independent samples are drawn 1000 times under the spike-free model (or "white" model), that is, $\mathbf{C}_N = \mathbf{I}$, and the ratios $\lambda_i(\mathbf{S}_w)/\lambda_{i+1}(\mathbf{S}_w)$ for $i = 1, 2, 3$ are recorded. Let γ_i be the largest value of these ratios for $i = 1, 2, 3$, respectively. By this method, we find

$$\gamma_1 = 1.04418, \ \gamma_2 = 1.0353, \ \gamma_3 = 1.0294.$$

Using the above configurations, we make 1000 independent realizations, and register the frequency (over the total 1000 realizations) of each case in Table 2. For all the realizations such that $\hat{p} = p \neq 0$, we calculate the estimates $\hat{\alpha}_i, i = 1, \ldots, \hat{p}$ using (3.11) and also their relative errors (RErr). We put the average RErr in Table 2.

TABLE 2
Accuracy of $\hat{p}$ and $\hat{\alpha}_i$

	Proportion		Proportion	Mean RErr of $\hat{\alpha}_i$
$\hat{p} = p$	99.8%	$p \neq 0$	97.5%	0.0031958
		$p = 0$	2.3%	
$\hat{p} > p$	0.2%	$p \neq 0$	0.2%	
		$p = 0$	0	
$\hat{p} < p$	0			

We can see that the estimator $\hat{p}$ has an accuracy of 99.8%, and the estimators $\hat{\alpha}_i$ are also accurate.

3.2. PCA on time-correlated data matrix

The PCA is a widely used method for noise reduction and data compression. Given the data matrix $\mathbf{X}$, our aim in this section is to reduce the row dimension N by removing the noise and preserving as much signal as possible contained in $\mathbf{Y}$.

We recall the main steps of PCA on the row vectors of $\mathbf{Y}$, if the original data $\mathbf{Y}$ is available.

1. Calculate the eigenvalues and the associated eigenvectors of the sample covariance matrix $\mathbf{S}_Y = M^{-1}\mathbf{YY}^*$.
2. Estimate the number of principal components (PC) p using the algorithm described in §3.1.
3. Let $\mathbf{v}_1, \ldots, \mathbf{v}_p$ be the eigenvectors of $\mathbf{S}_Y$ associated to the spikes. Then we get the PC of each vector $\mathbf{y}_k$ by projecting it into the subspace generated by $(\mathbf{v}_1, \ldots, \mathbf{v}_p)$, that is,

$$\mathbf{y}_{k,pc} := \mathbf{P}\mathbf{y}_k = \sum_{i=1}^{p} \langle \mathbf{y}_k, \mathbf{v}_i \rangle \mathbf{v}_i.$$

   The PC of the matrix data is then

$$\mathbf{Y}_{pc} := (\mathbf{y}_{1,pc}, \ldots, \mathbf{y}_{M,pc}) = \mathbf{P}\mathbf{Y},$$

   where

$$\mathbf{P} := \sum_{i=1}^{p} \mathbf{v}_i \mathbf{v}_i^*$$

   is the matrix of the orthogonal projection into the subspace generated by $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$. In this way we have removed the noise from the data.
4. We can further compress the data by expressing the vectors $\mathbf{y}_{k,pc}$ as p-dimensional vectors in the coordinate system $\mathbf{v}_1, \ldots, \mathbf{v}_p$, that is,

$$\mathbf{y}_{k,pc}|_{(\mathbf{v}_1, \ldots, \mathbf{v}_p)} = (\mathbf{v}_1^*, \ldots, \mathbf{v}_p^*)^\top \mathbf{y}_k.$$

   So the compressed data matrix from $\mathbf{Y}$ is

$$\mathbf{Y}_{pc}|_{(\mathbf{v}_1, \ldots, \mathbf{v}_p)} = (\mathbf{v}_1^*, \ldots, \mathbf{v}_p^*)^\top \mathbf{Y}.$$

   Note that after compression, the dimension of the matrix becomes $p \times M$. In this way we have reduced the signal dimension from N to p.

When only $\mathbf{X}$ is observed, our main results suggest that we can proceed the PCA on the whitened matrix $\mathbf{Y}_w$ to obtain the estimated PC of data. Note that in fact

$$\mathbf{S}_w = \frac{1}{M}\mathbf{Y}_w \mathbf{Y}_w^*,$$

so the PCA on $\mathbf{Y}_w$ can be done using the algorithm described above by replacing $\mathbf{Y}$ with $\mathbf{Y}_w$, and $\mathbf{S}_Y$ with $\mathbf{S}_w$, respectively. Let $\mathbf{y}_{w,k,pc}$ be the PC from the whitened data $\mathbf{y}_{w,k}$, and let $\mathbf{Y}_{w,pc} = (\mathbf{y}_{w,1,pc}, \ldots, \mathbf{y}_{w,M,pc})$. Then $\mathbf{Y}_{w,pc}$ and $\mathbf{Y}_{pc}$ are close in the following sense.

Proposition 3.3. *Let $\mathbf{Y}$ be defined in* (3.3) *and $\mathbf{X} = \mathbf{YR}_M^{1/2}$ with the conditions of Proposition 3.1 hold. Assume also that the distance between α_i and 1 are lower bounded. Then*

$$\frac{1}{M}\|\mathbf{Y}_{w,pc} - \mathbf{Y}_{pc}/\sqrt{\xi_N}\|_F^2 \xrightarrow{\text{a.s.}} 0. \tag{3.12}$$

Note that $\mathbf{Y}$ contains M columns (so there are M signal vectors). Then (3.12) means that the PC's from whitened data and those from the original data are in average close to each other.

We take $N = 1500$, $M = 2500$ ($c = 0.6$), $p = 3$, and a matrix $\mathbf{A}$ of dimension 1500×3. Although the matrix $\mathbf{A}$ in the model setting is deterministic, in our simulation program we have constructed it from three column Gaussian vectors of distributions $\mathcal{N}(0, \sigma_i^2 \mathbf{I})$ with $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.3$ respectively, in order to better approximate to the reality. Finally, the singular values of $\mathbf{A}$ are $5.21, 20.89, 42.49$. Let $\mathbf{Y}$ be a $N \times M$ random matrix whose rows are defined as in (3.3) with $\sigma^2 = 1$. Let $\mathbf{X} = \mathbf{Y} \mathbf{R}_M^{1/2}$ where $\mathbf{R}_M$ is defined in (3.7) with $a = 0.2$. Then we proceed the PCA on the whitened data $\mathbf{Y}_w$ as described above, and get $\mathbf{Y}_{w,pc}$. Since the data matrix $\mathbf{Y}$ is also available in such simulation experiments, we can also calculate $\mathbf{Y}_{pc}$, the PC of $\mathbf{Y}$, and compare the row vectors of the two matrices.

The comparison result is illustrated in Figure 1, where in 1(a), we draw the Euclidean norm of $\mathbf{y}_{w,k,pc} - \mathbf{y}_{k,pc} / \sqrt{\xi_N}$, and in 1(b), we draw the real part of cosine similarity of $\mathbf{y}_{w,k,pc}$ and $\mathbf{y}_{k,pc}$, for $k = 1, \ldots, M$. The cosine similarity of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$ is defined by

$$\cos(\mathbf{u}, \mathbf{v}) := \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|},$$

and its real part represents the cosine similarity of $\mathbf{u}, \mathbf{v}$ regarded as real vectors in $\mathbb{R}^{2N}$. From the simulation result we can see that the the PC vectors from whitened data are close to those from the original data. The norms of differences are under 0.5 for the majority part of the signal vectors, while the norms of the PC signal vectors themselves are around 10. The real cosine similarities are close to 1, which means that the directions of the two vectors under comparison are almost the same.

We repeat the same experiments with the same parameters except for $a = 0.95$. The results are shown in Figure 1, plots 1(c) and 1(d). We can see that although the LRD parameter a is quite close to 1 (the process has a quite long memory), the PCA has a comparable accuracy to the previous case.

Remark 3.2. Thanks to the separable structure $\mathbf{X} = \mathbf{Y} \mathbf{R}_M^{1/2}$, if we are not interested in the original data but just want to reduce the underlying noise for compression, we only need to project the columns of $\mathbf{X}$ into the subspace generated by $\mathbf{v}_{w,1}, \ldots, \mathbf{v}_{w,p}$, the eigenvectors of $\mathbf{S}_w$, and get
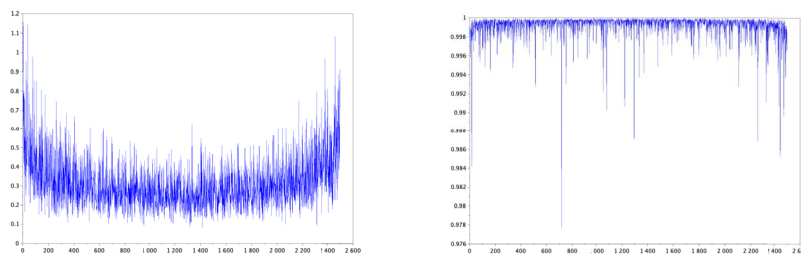
$$\mathbf{X}_{w,pc} = \left(\sum_{i=1}^p \mathbf{v}_{w,i} \mathbf{v}_{w,i}^* \right) \mathbf{Y} \mathbf{R}_M^{1/2},$$
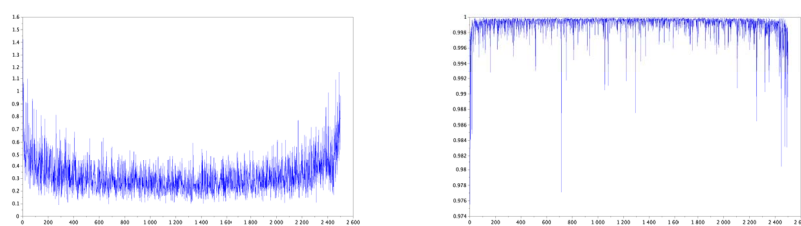
which is the PC of $\mathbf{X}$.

4. Numerical studies on inconsistency properties

In order to demonstrate the impact of new phenomena caused by LRD in the whitening procedure, and also to illustrate some of our conjectures, we present

(a) The Euclidean norm of $\mathbf{y}_{w,k,pc}$ − $\mathbf{y}_{k,pc}/\sqrt{\xi_N}$, with $a = 0.2$.

(b) The real part of cosine similarity of $\mathbf{y}_{w,k,pc}$ and $\mathbf{y}_{k,pc}$, with $a = 0.2$.

(c) The Euclidean norm of $\mathbf{y}_{w,k,pc}$ − $\mathbf{y}_{k,pc}/\sqrt{\xi_N}$, with $a = 0.95$

(d) The real part of cosine similarity of $\mathbf{y}_{w,k,pc}$ and $\mathbf{y}_{k,pc}$, with $a = 0.95$

FIG 1. *Comparison of PCA on the original and whitened data matrices.*

several numeric simulations in this section.

Throughout this section, we assume that

$$\mathbf{C}_N = \zeta_N^{-1} \operatorname{diag}(\alpha_1, \ldots, \alpha_p, 1, \ldots, 1) \tag{4.1}$$

where $p \geq 0$ is a fixed integer, $\alpha_i > 1$, $i = 1, \ldots, p$ are some fixed positive numbers, and the normalization $\zeta_N = N^{-1}(\alpha_1 + \cdots + \alpha_p + N - p)$ is such that $\operatorname{tr} \mathbf{C}_N = N$. When $p = 0$, $\mathbf{C}_N$ is identity. We assume also that the entries of $\mathbf{Z}$ are i.i.d standard real or complex Gaussian.

4.1. Norm inconsistency of $\hat{\mathbf{R}}_M$ when $1/2 < a < 1$

We take $\mathbf{C}_N = \mathbf{I}$, $\mathbf{Z}$ having i.i.d real standard Gaussian entries. In order to check the consistency of the unbiased estimate $\hat{\mathbf{R}}_M$ with $\mathbf{R}_M$ in spectral norm, we take $a = 0.9, 0.7, 0.5, 0.3, 0.1$ and $M = 250, 500, 1000, 2000$ with $N = 2M$. For each case we sample 500 independent realizations, and list the medians of $\|\hat{\mathbf{R}}_M - \mathbf{R}_M\|$ in Table 3.

We know that if $\|\hat{\mathbf{R}}_M - \mathbf{R}_M\| \to 0$ in probability, then the median must also converge to 0. However from Table 3 we can see that when $a > 0.5$, the median of $\|\hat{\mathbf{R}}_M - \mathbf{R}_M\|$ is large and increasing with M ($N = 2M$). When $a = 0.5$,

TABLE 3
Medians of $\|\hat{\mathbf{R}}_M - \mathbf{R}_M\|$.

a	$M = 250$	$M = 500$	$M = 1000$	$M = 2000$
0.9	7.0658	10.9346	12.7499	16.5067
0.7	3.2477	4.0702	4.6998	5.7822
0.5	1.8272	1.9393	1.9315	1.9579
0.3	1.1738	1.0211	1.0285	0.9318
0.1	0.7494	0.6552	0.5792	0.4873

(a) ESD of $(\hat{\mathbf{R}}_M^b)^{-1}\mathbf{R}_M$.

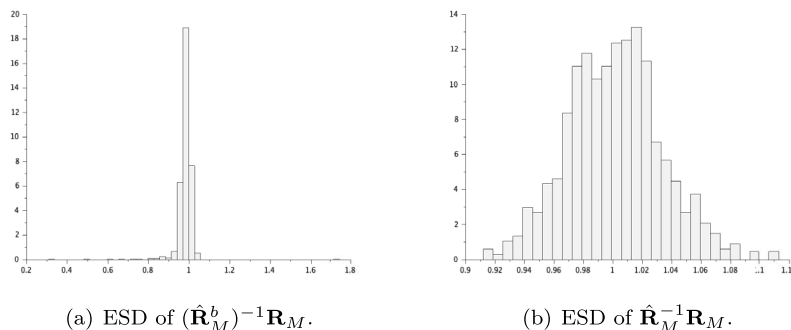(b) ESD of $\hat{\mathbf{R}}_M^{-1}\mathbf{R}_M$.

FIG 2. *Comparison of ESD between $(\hat{\mathbf{R}}_M^b)^{-1}\mathbf{R}_M$ and $\hat{\mathbf{R}}_M^{-1}\mathbf{R}_M$.*

which is the theoretical threshold of spectral norm consistency, the median of $\|\hat{\mathbf{R}}_M - \mathbf{R}_M\|$ seems oscillating, neither increasing nor decreasing. When $a < 0.5$, in which case we know that $\|\hat{\mathbf{R}}_M - \mathbf{R}_M\| \to 0$ almost surely (see (2.1)), the medians are relatively small and tend to decrease with M.

4.2. Ratio inconsistency of $\mathbf{R}_M^b$ and its impact on the whitening procedure

We have seen that a striking difference between the LRD and SRD situations is that the biased estimator $\mathbf{R}_M^b$ is not ratio consistent (Proposition 2.4), but instead, it is ratio LSD consistent (Proposition 2.5). That means, only a small number of eigenvalues of the ratio $\mathbf{R}_M^{-1}\hat{\mathbf{R}}_M^b$ deviate from 1. But how many deviating eigenvalues are there, and how does this affect the applications? We now study these questions via numerical experiments.

We take $M = 1000, N = 2000$ ($c = 2$), $\mathbf{C}_N = \mathbf{I}$, $\mathbf{Z}$ has real Gaussian entries, and $a = 0.9$ (the same configuration as the first row and second column of Table 3). We plot the histograms of the spectra of $(\hat{\mathbf{R}}_M^b)^{-1}\mathbf{R}_M$ and $\hat{\mathbf{R}}_M^{-1}\mathbf{R}_M$ in Figure 2. We note that the major part of eigenvalues of $(\hat{\mathbf{R}}_M^b)^{-1}\mathbf{R}_M$ are close to 1, but there are several extreme ones which are far away, the smallest at 0.3, and the largest at 1.75. In contrary, the spectrum of $\hat{\mathbf{R}}_M^{-1}\mathbf{R}_M$ spreads in the interval $[0.91, 1.12]$ much concentrated around 1.

Because of the ratio inconsistency of $\hat{\mathbf{R}}_M^b$, we may observe some extra "pseudo" spikes caused by the spikes of $\mathbf{R}_M^{1/2}(\hat{\mathbf{R}}_M^b)^{-1}\mathbf{R}_M^{1/2}$, if we replace $\hat{\mathbf{R}}_M$ with $\hat{\mathbf{R}}_M^b$ in
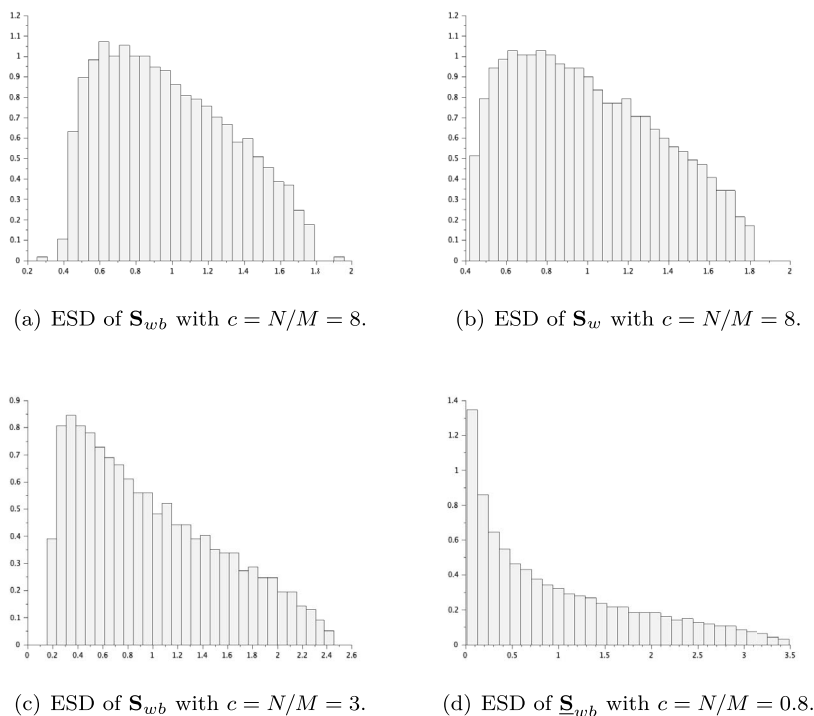
(a) ESD of $\mathbf{S}_{wb}$ with $c = N/M = 8$.

(b) ESD of $\mathbf{S}_w$ with $c = N/M = 8$.

(c) ESD of $\mathbf{S}_{wb}$ with $c = N/M = 3$.

(d) ESD of $\underline{\mathbf{S}}_{wb}$ with $c = N/M = 0.8$.

FIG 3. *Inconsistency of* $\mathbf{R}_M^b$ *may cause pseudo spikes under some conditions.*

the whitening procedure. Let

$$\underline{\mathbf{S}}_{wb} := \frac{1}{M}\mathbf{X}^*(\hat{\mathbf{R}}_M^b)^{-1}\mathbf{X} = \frac{1}{M}\mathbf{C}_N^{1/2}\mathbf{Z}^*(\mathbf{R}_M^{1/2}(\hat{\mathbf{R}}_M^b)^{-1}\mathbf{R}_M^{1/2})\mathbf{Z}\mathbf{C}_N^{1/2},$$

which parallels the matrix $\mathbf{S}_w$ in (3.1) with this replacement. In order to better illustrate the pseudo spikes, when $N > M$, we will plot the ESD of its dual sample covariance matrix

$$\mathbf{S}_{wb} := \frac{1}{N}(\hat{\mathbf{R}}_M^b)^{-1/2}\mathbf{X}\mathbf{C}_N\mathbf{X}^*(\hat{\mathbf{R}}_M^b)^{-1/2}.$$

We take $M = 1000, N = 8000, \mathbf{C}_N = \mathbf{I}$, and plot in Figure 3 the histogram of the ESD's of $\mathbf{S}_{wb}$ and also of the corresponding dual matrix $\mathbf{S}_w$ derived from $\mathbf{S}_w$. We can see that when $\mathbf{C}_N$ has no spikes, some unexpected spikes are observed in the ESD of $\mathbf{S}_{wb}$, whereas the corresponding $\mathbf{S}_w$ does not have this problem. Note also that this phenomenon occurs only with very large ratio $c = N/M$. When we take $N = 3000$ or $N = 800$ instead, the pseudo spikes disappear, see Figure 3(c) and 3(d).

Here is a heuristic explanation. From the ratio LSD consistency described in Proposition 2.5, the ratio inconsistency described in Proposition 2.4 and the

numeric simulations in this section, we can think of $(\mathbf{R}_M^b)^{-1/2}\mathbf{R}_M(\mathbf{R}_M^b)^{-1/2}$ asymptotically as a finite perturbation of identity. Thus the appearance or disappearance of pseudo spikes can be explained by the spikes of separable model

$$\mathbf{S}_{sep} := \frac{1}{M}\mathbf{C}_N^{1/2}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}^*\mathbf{C}_N^{1/2}$$

where $\boldsymbol{\Sigma}$ can be a positive deterministic Hermitian matrix. Largest eigenvalues of this matrix obeys a phase transition phenomenon as established in [6].

5. Proof of Theorem 2.2

5.1. Some preliminaries

As the matrices $\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2}$ and $\hat{\mathbf{R}}_M\mathbf{R}_M^{-1}$ have the same eigenvalues, we have

$$\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2} - \xi_N\mathbf{I}\| = \max_i\{|\lambda_i(\hat{\mathbf{R}}_M\mathbf{R}_M^{-1} - \xi_N\mathbf{I})|\}.$$

Then the idea of proof is to estimate the range of eigenvalues of the matrix $\hat{\mathbf{R}}_M\mathbf{R}_M^{-1}$.

The following lemma connects the spectrum of $\hat{\mathbf{R}}_M\mathbf{R}_M^{-1}$ with the ratio of spectral densities of the two Toeplitz matrices $\hat{\mathbf{R}}_M$ and $\mathbf{R}_M$. It was first proved in [10] and extended to integrable spectral densities in [20, Theorem 2.1].

Lemma 5.1. *Let $\mathbf{R}_{1,M}, \mathbf{R}_{2,M}$ be two $M \times M$ Toeplitz matrices generated by positive spectral densities $f_1, f_2 \in L^1(0, 2\pi)$, respectively. Then for any $M \geq 1$,*

$$\mathrm{Spec}(\mathbf{R}_{1,M}\mathbf{R}_{2,M}^{-1}) \subset \left[\underset{\theta\in[0,2\pi]}{\mathrm{ess\,inf}}\,\frac{f_1(\theta)}{f_2(\theta)}, \quad \underset{\theta\in[0,2\pi]}{\mathrm{ess\,sup}}\,\frac{f_1(\theta)}{f_2(\theta)}\right].$$

By this lemma, the spectral densities of the two Toeplitz matrices $\hat{\mathbf{R}}_M$ and $\mathbf{R}_M$ are important. We note that $\hat{\mathbf{R}}_M$ is random and depends on N, M, then so must be its spectral density, and the coefficients of orders higher than $M-1$ can be arbitrary. For each N and M, we define

$$\hat{f}_M(\theta) := \xi_N f(\theta) + \sum_{n=-M+1}^{M-1}(\hat{r}_n - \xi_N r_n)e^{\mathbf{i}n\theta} \tag{5.1}$$

where f is the spectral density of $\mathbf{R}_M$. Note that the Fourier coefficients of $\hat{f}_M$ are $\hat{r}_m$ for $-M+1 \leq m \leq M-1$, thus for this particular N and M, $\hat{f}_M$ is the spectral density of $\hat{\mathbf{R}}_M$, and by Lemma 5.1, the eigenvalues of $\hat{\mathbf{R}}_M\mathbf{R}_M^{-1}$ are in the interval

$$\left[\underset{\theta\in[0,2\pi]}{\mathrm{ess\,inf}}\,\frac{\hat{f}_M(\theta)}{f(\theta)}, \quad \underset{\theta\in[0,2\pi]}{\mathrm{ess\,sup}}\,\frac{\hat{f}_M(\theta)}{f(\theta)}\right].$$

Thus for any $x > 0$, we have

$$\mathbb{P}\left(\|\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2} - \xi_N\mathbf{I}\| > x\right) \leq \mathbb{P}\left(\operatorname*{ess\,sup}_{\theta \in [0,2\pi]}\left|\frac{\hat{f}_M(\theta)}{f_M(\theta)} - \xi_N\right| > x\right). \quad (5.2)$$

Let

$$\Upsilon_M(\theta) := \sum_{n=-M+1}^{M-1} r_n e^{\mathbf{i}n\theta}, \quad \hat{\Upsilon}_M(\theta) := \sum_{n=-M+1}^{M-1} \hat{r}_n e^{\mathbf{i}n\theta}. \quad (5.3)$$

Then $\hat{f}_M(\theta) - \xi_N f_M(\theta) = \hat{\Upsilon}_M(\theta) - \xi_N\Upsilon_M(\theta)$. Recall that $\mathbb{E}\hat{\Upsilon}_M(\theta) = \xi_N\Upsilon_M(\theta)$ for any $\theta \in [0, 2\pi]$. Then the RHS of (5.2) becomes

$$\mathbb{P}\left(\operatorname*{ess\,sup}_{\theta \in [0,2\pi]}\frac{|\hat{\Upsilon}_M(\theta) - \mathbb{E}\hat{\Upsilon}_M(\theta)|}{f(\theta)} > x\right). \quad (5.4)$$

This can be considered as the probability of large relative error of the estimation $\hat{\Upsilon}_M(\theta)$ with respect to $f(\theta)$. We will use a similar discretization strategy as [27]. Let

$$0 < \theta_1 < \theta_2 < \cdots < \theta_m < 2\pi$$

be an appropriate mesh of $(0, 2\pi)$, which will be precised later, then a key step is to estimate the probability

$$\mathbb{P}\left(|\hat{\Upsilon}_M(\theta_j) - \mathbb{E}\hat{\Upsilon}_M(\theta_j)| > x f(\theta_j)\right) \quad (5.5)$$

for each θ_j. This will be done in §5.2.

5.2. Relative error bound of $\hat{\Upsilon}_M(\theta)$ for individual θ

The goal of this subsection is to prove the following Proposition 5.2.

Proposition 5.2. *Let $K_1 > 0$ be an arbitrary positive constant. In both Gaussian case and spherical case, there exists $K_2 > 0$ depending on K_1, such that for any $x \in (0, CK_1/\kappa)$, for large enough M, any $N \geq 1$, and any $\theta \in (0, 2\pi)$, we have*

$$\mathbb{P}\left(\left|\hat{\Upsilon}_M(\theta) - \mathbb{E}\hat{\Upsilon}_M(\theta)\right| > x f(\theta)\right) \leq 2\exp\left(-\frac{Nx^2}{CK_2\log^2 M}\right). \quad (5.6)$$

We will prove Proposition 5.2 separately for complex Gaussian, complex spherical, and real cases. Before that, we still need some preliminary works.

Denote

$$\mathbf{D}_M(\theta) := \operatorname{diag}(1, e^{-\mathbf{i}\theta}, \ldots, e^{-\mathbf{i}(M-1)\theta}), \quad \mathbf{B}_M := \left(\frac{1}{M - |i-j|}\right)_{i,j=0}^{M-1}, \quad (5.7)$$

and

$$\mathbf{Q}_M(\theta) := \mathbf{R}_M^{1/2}\mathbf{D}_M(\theta)\mathbf{B}_M\mathbf{D}_M^*(\theta)\mathbf{R}_M^{1/2}. \quad (5.8)$$

Then from Lemma 7 and 8 and (9) of [27], under **A2**, we have

$$\hat{\Upsilon}_M(\theta) = \frac{1}{N} \operatorname{tr} \mathbf{C}_N^{1/2} \mathbf{Z} \mathbf{Q}_M(\theta) \mathbf{Z}^* \mathbf{C}_N^{1/2} = \frac{1}{N} \sum_{n=1}^{N} c_n \mathbf{z}_n \mathbf{Q}_M(\theta) \mathbf{z}_n^*, \qquad (5.9)$$

From **A1**, $\mathbf{z}_n^\top$ can be a real or complex vector. We only give the complete proof for the complex case, and list the differences between real and complex cases to ease the adaption for the real case. Thus, let us first assume that $\mathbf{z}_n^\top$ are complex Gaussian or uniformly distributed on the complex sphere. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M$ be the eigenvalues of $\mathbf{Q}_M(\theta)$ (Warning: $\mathbf{Q}_M(\theta)$ may be indefinite). By the unitary invariance of $\mathbf{z}_n^\top$, we have

$$\hat{\Upsilon}_M(\theta) - \mathbb{E}\hat{\Upsilon}_M(\theta) \overset{\mathcal{L}}{=} \frac{1}{N} \sum_{n=1}^{N} c_n \sum_{m=1}^{M} \sigma_m(|Z_{n,m}|^2 - 1). \qquad (5.10)$$

In light of (5.10), in order to prove (5.6), it is equivalent to prove

$$\mathcal{P} := \mathbb{P}\left(\left| \frac{1}{N} \sum_{n=1}^{N} c_n \sum_{m=1}^{M} \sigma_m(|Z_{n,m}|^2 - 1) \right| > x f(\theta) \right) \leq 2 \exp\left(-\frac{N x^2}{C K_2 \log^2 M} \right). \qquad (5.11)$$

For this, we find it crucial to estimate $\operatorname{tr} \mathbf{Q}_M^2(\theta) = \sum_{m=1}^{M} \sigma_m^2$. We now state the following proposition, whose proof is provided in §5.4.

Proposition 5.3. *Let $\mathbf{Q}_M(\theta)$ be defined as (5.8) with Toeplitz matrix $\mathbf{R}_M$ whose spectral density f satisfies **A3**, **A4**, **A5**. Then*

$$\frac{\operatorname{tr} \mathbf{Q}_M^2(\theta)}{f^2(\theta) \log^2 M} \qquad (5.12)$$

is uniformly bounded in $\theta \in [-\pi, \pi] \backslash \{0\}$ and $M > 1$.

5.2.1. Proof of Proposition 5.2, complex Gaussian case

Let

$$\sigma_i' = \frac{\sigma_i}{\sqrt{\sum_{m=1}^{M} \sigma_m^2}}, \quad i = 1, \ldots, M.$$

Then

$$\mathcal{P} = \mathbb{P}\left(\left| \sum_{n=1}^{N} c_n \sum_{m=1}^{M} \sigma_m'(|Z_{n,m}|^2 - 1) \right| > \frac{N x f(\theta)}{\sqrt{\sum_{m=1}^{M} \sigma_m^2}} \right).$$

By Proposition 5.3, there exists a constant $K > 0$ such that

$$\frac{N x f(\theta)}{\sqrt{\sum_m \sigma_m^2}} \geq \frac{N x}{\sqrt{K} \log M}$$

for any θ and $M > 1$. Then

$$\mathcal{P} \leq \mathbb{P}\left(\left|\sum_n c_n \sum_m \sigma'_m(|Z_{n,m}|^2 - 1)\right| > \frac{Nx}{\sqrt{K}\log M}\right). \tag{5.13}$$

Then we only need to estimate the RHS of (5.13) with $\sum_m (\sigma'_m)^2 = 1$. Let

$$\mathcal{P}_i := \mathbb{P}\left((-1)^{i-1}\sum_n c_n \sum_m \sigma'_m(|Z_{n,m}|^2 - 1) > \frac{Nx}{\sqrt{K}\log M}\right), \quad i = 1, 2.$$

Then $\mathcal{P} \leq \mathcal{P}_1 + \mathcal{P}_2$, and the estimation of $\mathcal{P}_1$ and $\mathcal{P}_2$ is similar, we only need to estimate $\mathcal{P}_1$.

Using Chernoff bound, for any $\tau > 0$, we have

$$\mathcal{P}_1 \leq \exp\left(-\frac{Nx\tau}{\sqrt{K}\log M} + \log \mathbb{E}e^{\sum_n \tau c_n \sum_m \sigma'_m(|Z_{n,m}|^2 - 1)}\right). \tag{5.14}$$

Note that the rows $\mathbf{z}_n$ of $\mathbf{Z}$ are i.i.d across n, we then have

$$\mathcal{P}_1 \leq \exp\left(-\frac{Nx\tau}{\sqrt{K}\log M} + \sum_n \Phi_M(\tau c_n)\right), \tag{5.15}$$

where Φ_M is the cumulant generating function of $\sum_m \sigma'_m(|Z_{n,m}|^2 - 1)$:

$$\Phi_M(z) := \log \mathbb{E}e^{z \sum_m \sigma'_m(|Z_{n,m}|^2 - 1)}.$$

Lemma 5.4. *When $Z_{n,m}$ are i.i.d. standard complex Gaussian variables, there exists $A > 0$ and $\varepsilon > 0$ such that when $|z| < \varepsilon$, we have*

$$|\Phi_M(z)| \leq A|z|^2$$

for any $M > 1$.

Proof. Let ϕ be the cumulant generating function of $|Z_{n,m}|^2 - 1$:

$$\phi(z) := \log \mathbb{E}e^{z(|Z_{n,m}|^2 - 1)}.$$

Then as $Z_{n,m}$ are i.i.d. standard complex Gaussian, we have

$$\Phi_M(z) = \sum_m \phi(z\sigma'_m),$$

and

$$\phi(z) = -z - \log(1 - z).$$

By the Taylor's expansion $\log(1 - z) = z - z^2/2 + z^3/3 - \cdots$, choosing an arbitrary $\varepsilon \in (0, 1)$, there exists $A > 0$ such that for any $|z| \leq \varepsilon$, we have

$$|\phi(z)| = |z|^2|1/2 - z/3 + \cdots| \leq A|z|^2.$$

Thus

$$|\Phi_M(z)| \leq \sum_m |\phi(z\sigma'_m)| \leq \sum_m A|z|^2(\sigma'_m)^2 = A|z|^2.$$

$\square$

From (5.15) and Lemma 5.4, for any $\tau > 0$ such that $|\tau c_n| \leq \varepsilon$, we have

$$\mathcal{P}_1 \leq \exp\left(-\frac{\tau N x}{\sqrt{K}\log M} + A\tau^2 \sum_n c_n^2\right). \tag{5.16}$$

Noting that $\sum_n c_n^2 \leq CN$ by **A2**, we then have

$$\mathcal{P}_1 \leq \exp\left(-\frac{\tau N x}{\sqrt{K}\log M} + CAN\tau^2\right). \tag{5.17}$$

If we can take

$$\tau = \frac{x}{2CA\sqrt{K}\log M},$$

we will minimize the RHS of (5.17) and get

$$\mathcal{P}_1 \leq \exp\left(-\frac{N x^2}{4KCA\log^2 M}\right). \tag{5.18}$$

In order to validate (5.18), we have to keep $|\tau c_n| \leq \varepsilon$ for all n. Note that $|c_n| \leq \kappa \log M$ by **A2**, we only need to keep $\tau \leq \varepsilon/(\kappa \log M)$. That is, we only need to keep

$$x \leq 2\varepsilon CA\sqrt{K}/\kappa.$$

Taking $K_1 = 2\varepsilon A\sqrt{K}$ and $K_2 = 4KA$, we conclude that, for any $x \in (0, CK_1/\kappa)$,

$$\mathcal{P}_1 \leq \exp\left(-\frac{N x^2}{CK_2\log^2 M}\right). \tag{5.19}$$

Note that K or A can be adjusted to a larger constant, which means that K_1 can be arbitrarily large, and K_2 should be adjusted correspondingly. This is exactly the statement of Proposition 5.2. We have thus proved the proposition for complex Gaussian case.

5.2.2. *Proof of Proposition 5.2, complex spherical case*

When $\mathbf{z}_n$ follows the uniform distribution on the sphere $\{\mathbf{z} \in \mathbb{C}^M : \|\mathbf{z}\| = \sqrt{M}\}$. Then $\|\mathbf{z}_n\|^2 = M$. We have

$$\begin{aligned}
\hat{\Upsilon}_M(\theta) - \mathbb{E}\hat{\Upsilon}_M &= \mathbf{z}_n \mathbf{Q}_M(\theta)\mathbf{z}_n^* - \frac{\|\mathbf{z}_n\|^2}{M}\operatorname{tr}\mathbf{Q}_M(\theta) \\
&= \mathbf{z}_n\left(\mathbf{Q}_M(\theta) - \frac{\operatorname{tr}\mathbf{Q}_M(\theta)}{M}\mathbf{I}\right)\mathbf{z}_n^* \\
&\stackrel{\mathcal{L}}{=} \sum_{m=1}^{M}\left(\sigma_m - \frac{\operatorname{tr}\mathbf{Q}_M(\theta)}{M}\right)|Z_{n,m}|^2.
\end{aligned}$$

Write

$$\sigma_m' = \frac{\sigma_m - \operatorname{tr}\mathbf{Q}_M(\theta)/M}{\sqrt{\sum_m(\sigma_m - \operatorname{tr}\mathbf{Q}_M(\theta)/M)^2}},$$

then $\sum_m \sigma'_m = 0$, $\sum_m \sigma'^2_m = 1$, and $\mathcal{P}$ defined in (5.11) becomes

$$\mathcal{P} = \mathbb{P}\left(\left|\sum_{n=1}^{N} c_n \sum_{m=1}^{M} \sigma'_m |Z_{n,m}|^2\right| > \frac{Nxf(\theta)}{\sqrt{\sum_m (\sigma_m - \operatorname{tr}\mathbf{Q}_M(\theta)/M)^2}}\right).$$

Using Proposition 5.3 again, there exists some constant $K > 0$ such that

$$\sum_{m=1}^{M}\left(\sigma_m - \frac{\operatorname{tr}\mathbf{Q}_M(\theta)}{M}\right)^2 \leq \operatorname{tr}\mathbf{Q}_M^2(\theta) \leq Kf^2(\theta)\log^2 M.$$

Then

$$\mathcal{P} \leq \mathbb{P}\left(\left|\sum_{n=1}^{N} c_n \sum_{m=1}^{M} \sigma'_m |Z_{n,m}|^2\right| > \frac{Nx}{\sqrt{K}\log M}\right).$$

Similar to the proof in the Gaussian case, we define

$$\mathcal{P}_i = \mathbb{P}\left((-1)^{i-1}\sum_{n=1}^{N} c_n \sum_{m=1}^{M} \sigma'_m |Z_{n,m}|^2 > \frac{Nx}{\sqrt{K}\log M}\right), \quad i = 1, 2$$

and we just need to estimate $\mathcal{P}_1$. Using Chernoff bound, for any $\tau > 0$, we have

$$\mathcal{P}_1 \leq \exp\left(-\frac{Nx\tau}{\sqrt{K}\log M} + \sum_n \Phi_M(\tau c_n)\right), \tag{5.20}$$

where

$$\Phi_M(z) := \log \mathbb{E}e^{z\sum_m \sigma'_m |Z_{n,m}|^2} = \frac{M}{2(M+1)}z^2 + \cdots.$$

Lemma 5.5. *When $\mathbf{z}_n^\top$ are i.i.d uniformly distributed on the complex sphere $\{\mathbf{z} \in \mathbb{C}^M : \|\mathbf{z}\| = \sqrt{M}\}$, there exists $A > 0$ and $\varepsilon > 0$ such that when $|z| < \varepsilon$, we have*

$$|\Phi_M(z)| \leq A|z|^2$$

for any $M > 1$.

Proof. On the one hand, by the Taylor's expansion of $e^{\Phi_M(z)}$, and the fact that $\mathbb{E}\sum_m \sigma'_m |Z_{n,m}|^2 = \sum_m \sigma'_m = 0$, we have

$$\mathbb{E}\exp\left(z\sum_m \sigma'_m |Z_{n,m}|^2\right) = 1 + \sum_{k=2}^{\infty} \frac{z^k}{k!}\mathbb{E}\left(\sum_m \sigma'_m |Z_{n,m}|^2\right)^k, \tag{5.21}$$

On the other hand, let $\mathbf{g} = (g_1, \ldots, g_M)^\top \in \mathbb{C}^M$ be a standard complex Gaussian vector. Then since $\mathbf{g}$ is spherically symmetric, we have $\mathbf{g} \overset{\mathcal{L}}{=} \|\mathbf{g}\|\mathbf{z}_n/\sqrt{M}$ where $\|\mathbf{g}\|$ and $\mathbf{z}_n$ are independent (see e.g. [8]). Then

$$\mathbb{E}\exp\left(z\sum_m \sigma'_m |g_m|^2\right) = \mathbb{E}\exp\left(z\frac{\|\mathbf{g}\|^2}{M}\sum_m \sigma'_m |Z_{n,m}|^2\right)$$

$$= \sum_{k=0}^{\infty} \frac{z^k \mathbb{E}\|\mathbf{g}\|^{2k}}{k!M^k}\mathbb{E}\left(\sum_m \sigma'_m |Z_{n,m}|^2\right)^k. \tag{5.22}$$

Since we know that for Gaussian variables g_m,

$$\mathbb{E} \exp\left(z \sum_m \sigma'_m |g_m|^2 \right) = \prod_{m=1}^{M} \frac{1}{1 - z\sigma'_m} = \exp\left(\sum_{m=1}^{M} \log(1 - z\sigma'_m) \right),$$

comparing to (5.22) we get

$$\sum_{k=0}^{\infty} \frac{z^k \mathbb{E}\|\mathbf{g}\|^{2k}}{k! M^k} \mathbb{E}\left(\sum_m \sigma'_m |Z_{n,m}|^2 \right)^k = \exp\left(\sum_{m=1}^{M} \log(1 - z\sigma'_m) \right). \qquad (5.23)$$

Note that $\sum_m \sigma'_m = 0$, $\sum_m (\sigma'_m)^2 = 1$ and $|\sigma'_m| \leq 1$. From the proof of Lemma 5.4, for an arbitrary $\epsilon \in (0,1)$, there exists A_ϵ such that $|\log(1 - z\sigma'_m) + z\sigma'_m| \leq A_\epsilon |z|^2 (\sigma'_m)^2$ for any $|z| \leq \epsilon$. Thus for these z we have

$$\left| \exp\left(\sum_{m=1}^{M} \log(1 - z\sigma'_m) \right) \right| \leq \exp(A_\epsilon |z|^2). \qquad (5.24)$$

Applying Cauchy's integration formula to (5.23) and using (5.24), for each $k \geq 0$, we have

$$\left| \frac{\mathbb{E}\|\mathbf{g}\|^{2k}}{k! M^k} \mathbb{E}(\sum_m \sigma'_m |Z_{n,m}|^2)^k \right| = \frac{1}{2\pi} \left| \int_{|z|=\epsilon} \frac{1}{z^{k+1}} \prod_{m=1}^{M} \frac{1}{1 - z\sigma'_m} \, \mathrm{d}z \right| \leq \frac{e^{A_\epsilon \epsilon^2}}{\epsilon^k}.$$

Note that

$$\frac{\mathbb{E}\|\mathbf{g}\|^{2k}}{M^k} = \frac{M(M+1)\cdots(M+k-1)}{M^k} \geq 1,$$

we get

$$\left| \frac{1}{k!} \mathbb{E}(\sum_m \sigma'_m |Z_{n,m}|^2)^k \right| \leq \frac{e^{A_\epsilon \epsilon^2}}{\epsilon^k}.$$

Then for any $|z| \leq \epsilon/2$, we have

$$\left| \sum_{k=2}^{\infty} \frac{z^{k-2}}{k!} \mathbb{E}(\sum_m \sigma'_m |Z_{n,m}|^2)^k \right| \leq 2\epsilon^{-2} e^{A_\epsilon \epsilon^2}.$$

Taking this into (5.21), we get

$$\left| \mathbb{E} e^{z \sum_m \sigma'_m |Z_{n,m}|^2} - 1 \right| \leq 2\epsilon^{-2} e^{A_\epsilon \epsilon^2} |z|^2$$

for any $|z| \leq \varepsilon/2$. Using the inequality $|\log(1+z)| \leq K|z|$ for $|z| \leq \epsilon < 1$, we conclude that there exists $\varepsilon > 0$, $A > 0$ such that when $|z| < \varepsilon$, we have

$$|\Phi_M(z)| = |\log \mathbb{E} e^{z \sum_m \sigma'_m |Z_{n,m}|^2}| \leq A|z|^2. \qquad \square$$

The remaining proof for spherical case is identical to the proof for Gaussian case from (5.16) onward.

5.2.3. Real case

In the real case, the proof is similar, so we omit the detail. To complete the proof, we only need to replace the corresponding items with the following mentioned properties in the proof of complex case.

The first, when $\mathbf{C}_N$, $\mathbf{Z}$, $\mathbf{R}_M$ are all real, one has

$$\hat{\Upsilon}_M(\theta) = \Re(\hat{\Upsilon}(\theta)) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{z}_n \Re(\mathbf{Q}_M(\theta)) \mathbf{z}_n^*.$$

The second, note also that

$$\mathrm{tr}(\Re\mathbf{Q}_M(\theta))^2 \leq \mathrm{tr}\,\mathbf{Q}_M^2(\theta)$$

and

$$\log \mathbb{E}e^{z|G|^2} = -\frac{1}{2}\log(1 - 2z)$$

for standard real Gaussian variable G with $|z| < 1/2$.

The third, if $\mathbf{g} \in \mathbb{R}^M$ is a standard real Gaussian vector, we have

$$\frac{M^k}{\mathbb{E}\|\mathbf{g}\|^{2k}} = \frac{M^k}{M(M+2)\cdots(M+2k-2)} \leq 1.$$

5.3. Relative error bound for all θ by discretization

Let β be a positive integer to be determined afterwards. For $k = 0, \dots, M^\beta$, let

$$\theta_k := \frac{2\pi k}{M^\beta}.$$

For $\theta \in (0, 2\pi)$, let θ_j be such that $\theta_{j-1} < \theta \leq \theta_j$ if $\theta \in (0, \pi]$, and $\theta_j \leq \theta < \theta_{j+1}$ if $\theta \in (\pi, 2\pi)$. We write

$$\frac{|\hat{\Upsilon}_M(\theta) - \Upsilon_M(\theta)|}{f(\theta)} \leq \frac{|\hat{\Upsilon}_M(\theta) - \hat{\Upsilon}_M(\theta_j)|}{f(\theta)} + \frac{|\hat{\Upsilon}_M(\theta_j) - \Upsilon_M(\theta_j)|}{f(\theta)}$$
$$+ \frac{|\Upsilon_M(\theta) - \Upsilon_M(\theta_j)|}{f(\theta)} =: \chi_1(\theta) + \chi_2(\theta) + \chi_3(\theta).$$

We first estimate the probability of large deviation of χ_2, which we will see is the main part of (2.3). Different from the SRD case, we have to treat the singularity of the spectral density f at 0. We note that

$$\chi_2(\theta) = \chi_2(\theta_j)\frac{f(\theta_j)}{f(\theta)}.$$

We now prove that $\frac{f(\theta_j)}{f(\theta)}$ is bounded for $\theta \in [0, 2\pi]$ and θ_j. Because f is supposed to be even and 2π-periodic, we only need to consider $\theta \in (0, \pi)$. Note that by Lemma 5.7(c) below, as $\theta_j \to 0^+$,

$$\frac{f(\theta_j)}{f(\theta)} \leq \frac{f(\theta_j)}{\inf_{0 < t \leq \theta_j} f(t)} \sim 1.$$

Let $\delta > 0$ be such that

$$\frac{f(\theta_j)}{f(\theta)} \leq \frac{f(\theta_j)}{\inf_{0 < t \leq \theta_j} f(t)} \leq 2$$

for $0 < \theta \leq \theta_j \leq \delta$. Then for any $\theta \in (0, \pi)$, we have

$$\frac{f(\theta_j)}{f(\theta)} \leq \max\left(2, \frac{\sup_{t \in [\delta, \pi]} f(t)}{\inf_{t \in [\delta, \pi]} f(t)}\right),$$

and by **A3**, **A4**, the RHS of the above inequality is bounded. Denote this bound as F. Using Proposition 5.2, for any $x \in (0, CFK_1/\kappa)$, as M, N are large enough, we have

$$\mathbb{P}\left(\sup_{\theta \in (\theta_{j-1}, \theta_j]} \chi_2(\theta) > x\right) \leq \mathbb{P}\left(\chi_2(\theta_j) > \frac{x}{F}\right) \leq 2\exp\left(-\frac{Nx^2}{CF^2 K_2 \log^2 M}\right).$$

Then

$$\mathbb{P}\left(\sup_{\theta \in (0, 2\pi)} \chi_2(\theta) > x\right) \leq 2M^\beta \exp\left(-\frac{Nx^2}{CF^2 K_2 \log^2 M}\right). \tag{5.25}$$

We then estimate χ_1. From the proof of Lemma 10 in [27], and note that $f(\theta)$ is bounded away from 0, also note Lemma 5.6 for the bound of $\|\mathbf{R}_M\|$, and **A2** for the bound of $\|\mathbf{C}_N\|$, we have

$$\sup_{\theta \in [0, 2\pi]} \chi_1(\theta) \leq \sup_{\theta \in [0, 2\pi]} \frac{1}{Nf(\theta)} \|\mathbf{C}_N\| \|\mathbf{Q}_M(\theta) - \mathbf{Q}_M(\theta_j)\| |\theta - \theta_j| \sum_{m,n} |Z_{n,m}|^2$$

$$\lesssim \sup_{\theta \in [0, 2\pi]} \frac{1}{N} \|\mathbf{C}_N\| \|\mathbf{R}_M\| M\sqrt{\log M} |\theta - \theta_j| \sum_{m,n} |Z_{n,m}|^2$$

$$\lesssim \frac{\kappa}{N} M^{1+a-\beta} L(M)(\log M)^{3/2} \sum_{m,n} |Z_{n,m}|^2$$

$$\leq \kappa M^{2+a-\beta} L(M)(\log M)^{3/2} \frac{\sum_{m,n} |Z_{n,m}|^2}{MN}.$$

Then for x_M satisfying $x_M \gtrsim M^{-\gamma}$ with some $\gamma > 0$, we have

$$\mathbb{P}\left(\sup_{\theta \in [0, 2\pi]} \chi_1(\theta) > x_M\right) \leq \mathbb{P}\left(\frac{\sum_{m,n} |Z_{n,m}|^2}{MN} > \frac{M^{\beta-2-a-\gamma}}{\kappa L(M)(\log(M))^{3/2}}\right). \tag{5.26}$$

We take $\beta > 2 + a + \gamma$ and let $\varepsilon = \frac{\beta - 2 - a - \gamma}{2}$, then as M is large enough, we have

$$\frac{M^{\beta-2-a-\gamma}}{\kappa L(M)(\log(M))^{3/2}} \gg M^\varepsilon > 1. \tag{5.27}$$

If $\mathbf{z}_n$ are standard complex normal, by [27, Lemma 2], we have for any $y > 1$,

$$\mathbb{P}\left(\frac{\sum_{m,n} |Z_{n,m}|^2}{MN} > y\right) \leq \exp(-MN(y - 1 - \log y)). \tag{5.28}$$

For y large enough, we have $y - 1 - \log y > y/2$. Thus if M is large enough, from (5.26), (5.27) and (5.28),

$$\mathbb{P}\left(\sup_{\theta \in [0, 2\pi]} \chi_1(\theta) > x_M\right) \le \exp(-NM^{1+\varepsilon}/2). \qquad (5.29)$$

The real Gaussian case is similar.

If $\mathbf{z}_n$ are spherical of radius $\sqrt{M}$, then $\frac{\sum_{m,n} |Z_{n,m}|^2}{MN} = 1$ and the RHS of (5.26) is eventually zero. Therefore, for both spherical and Gaussian cases, when M is large enough, (5.29) holds.

We now estimate the bound of χ_3. From the proof of Lemma 12 in [27], and note that $|\xi_N| \le \sqrt{\mathrm{tr}\,\mathbf{C}_N^2/N} \le \sqrt{C}$ we have

$$\sup_{\theta \in [0, 2\pi]} \chi_3(\theta) \lesssim M^2 |\theta - \theta_j| \|\mathbf{R}_M\| \sqrt{C \log M}$$

$$\lesssim M^{2+a-\beta} L(M) \sqrt{C \log M}.$$

For any x_M satisfying $x_M \gtrsim M^{-\gamma}$, let $\beta > 2 + a + \gamma$, then as M is large enough, we have

$$\sup_{\theta \in [0, 2\pi]} \chi_3(\theta) < x_M.$$

and thus

$$\mathbb{P}\left(\sup_{\theta \in [0, 2\pi]} \chi_3(\theta) > x_M\right) = 0. \qquad (5.30)$$

The final result follows from combining the above estimations (5.25), (5.29), (5.30), and letting the dominant item (5.25) absorb the others by appropriately changing the corresponding constants.

5.4. Proof of Proposition 5.3

In order to estimate $\mathrm{tr}\,\mathbf{Q}_M^2(\theta)$, we first estimate the norm of the Toeplitz matrix $\mathbf{R}_M$. The following lemma is a direct corollary of Theorem 2.3 in [25], so the proof is omitted.

Lemma 5.6. *If $(\mathbf{R}_M)$ is a sequence of Toeplitz matrices satisfying **A3** and **A5**, then*

$$\|\mathbf{R}_M\| \asymp M^a L(M).$$

We also need the following properties of functions regularly varying at 0.

Lemma 5.7. *If f satisfies **A5**, then*

(a) $\sup\{f(t) : x \le t \le \pi\} \sim f(x)$ *as $x \to 0^+$.*
(b) $\inf\{f(t) : 0 < t \le x\} \sim f(x)$ *as $x \to 0^+$.*
(c) $\int_0^x f(t)\,dt \sim \frac{x^{1-a}L(x^{-1})}{1-a} = \frac{xf(x)}{1-a}$ *as $x \to 0^+$.*

Proof. By changing the variable $u = x^{-1}$, (a) and (b) follow from Theorem 1.5.3 of [2], and (c) from Proposition 1.5.10 of [2]. □

The bound of $\operatorname{tr} \mathbf{Q}_M^2(\theta)$ will be estimated in different ways according to the location of θ. From Lemma 5.6 above, and (11) in [27], we get a global estimation

$$\operatorname{tr} \mathbf{Q}_M^2(\theta) = O(M^{2a} L^2(M) \log M) \tag{5.31}$$

for any $\theta \in [-\pi, \pi]$. This bound may be sharp for θ very close to the singular point 0, but not for θ farther away from 0.

In order to establish a sharper bound of $\operatorname{tr} \mathbf{Q}_M^2(\theta)$ for a regular point θ, we define, for a certain $\delta > 0$, a local ∞-norm $\|f\|_{(\theta,\delta)}$ as

$$\|f\|_{(\theta,\delta)} := \operatorname*{ess\,sup}_{t \in (\theta-\delta, \theta+\delta)} \{|f(t)|\}. \tag{5.32}$$

Proposition 5.8. *Let $\mathbf{Q}_M(\theta)$ be defined in (5.8) with $\mathbf{R}_M$ having positive spectral density $f \in L^1(-\pi, \pi)$. Then there exists an absolute constant $K > 0$ such that for any $\theta \in \mathbb{R}$ and $\delta \in (0, \pi/2)$,*

$$\frac{\operatorname{tr} \mathbf{Q}_M^2(\theta)}{2 \log M} \leq \|f\|_{(\theta,\delta)}^2 + \frac{K\|f\|_1(\|f\|_1 + \|f\|_{(\theta,\delta)})}{\delta^2 \log M}. \tag{5.33}$$

Before proving this proposition, it is convenient to express $\operatorname{tr} \mathbf{Q}_M^2(\theta)$ in terms of f. Using the integral expression of entries r_{i-j}, we write

$$\begin{aligned}
\operatorname{tr} \mathbf{Q}_M^2(\theta) &= \sum_{i,j,k,l} r_{i-j} \frac{e^{-\mathbf{i}(j-k)\theta}}{M - |j-k|} r_{k-l} \frac{e^{-\mathbf{i}(l-i)\theta}}{M - |l-i|} \\
&= \frac{1}{4\pi^2} \sum_{i,j,k,l} \int_{-\pi}^{\pi} f(x) e^{-\mathbf{i}(i-j)x} \, dx \frac{e^{-\mathbf{i}(j-k)\theta}}{M - |j-k|} \int_{-\pi}^{\pi} f(y) e^{-\mathbf{i}(k-l)y} \, dy \frac{e^{-\mathbf{i}(l-i)\theta}}{M - |l-i|} \\
&= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x+\theta) f(y+\theta) \sum_{i,j,k,l} \frac{e^{\mathbf{i}(jx-ky)-\mathbf{i}(ix-ly)}}{(M-|j-k|)(M-|l-i|)} \, dx \, dy \\
&= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x+\theta) f(y+\theta) \left| \sum_{i,j} \frac{e^{\mathbf{i}(jx-iy)}}{M - |i-j|} \right|^2 \, dx \, dy.
\end{aligned} \tag{5.34}$$

We denote

$$g(x,y) := \sum_{i,j} \frac{e^{\mathbf{i}(jx-iy)}}{M - |i-j|}.$$

For later use, it is necessary to study several bounds of this integral kernel $g(x,y)$.

Lemma 5.9. *The kernel $g(x,y) = \sum_{0 \leq i,j \leq M-1} \frac{e^{\mathbf{i}(jx-iy)}}{M-|i-j|}$ satisfies*

(1) $\frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |g(x,y)|^2 \, dx \, dy = 1 + 2 \sum_{k=1}^{M-1} \frac{1}{k} \sim 2 \log M.$

(2) $\frac{1}{2\pi}\int_{-\pi}^{\pi}|g(x,y)|^2\,\mathrm{d}x \lesssim \min(\frac{\log^2 M}{|\sin(y/2)|}, \frac{1}{\sin^2(y/2)})$.

(3) *For any* $\delta \in (0,\pi)$ *and* $x,y \in [-\pi,\pi]\backslash(-\delta,\delta)$, *we have*

$$|g(x,y)| \lesssim \frac{1}{|\sin\delta|}.$$

Proof. For (1), we replace f with 1 and $\mathbf{R}_M$ with $\mathbf{I}$ correspondingly in eq. (5.34), and we get

$$\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|g(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y = \operatorname{tr}\mathbf{B}_M^2 = 1 + 2\sum_{k=1}^{M-1}\frac{1}{k}.$$

For (2), note that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|g(x,y)|^2\,\mathrm{d}x = \sum_j\left|\sum_i \frac{e^{-\mathbf{i}iy}}{M-|i-j|}\right|^2.$$

Multiplying $|1-e^{-\mathbf{i}y}| = 2|\sin(y/2)|$ or $|1-e^{-\mathbf{i}y}|^2 = 4\sin^2(y/2)$ on both sides, and note that

$$\left|\sum_{i=0}^{M-1}\frac{e^{-\mathbf{i}iy}(1-e^{-\mathbf{i}y})}{M-|i-j|}\right|$$

$$= \left|\frac{1}{M-j} - \frac{e^{-\mathbf{i}My}}{M-|M-1-j|} + \sum_{i=1}^{M-1}\frac{(|i-j|-|i-j-1|)e^{-\mathbf{i}iy}}{(M-|i-j|)(M-|i-j-1|)}\right|$$

$$\leq \frac{1}{M-j} + \frac{1}{M-|M-1-j|} + \sum_{i=1}^{M-1}\frac{1}{(M-|i-j|)(M-|i-j-1|)}$$

$$\lesssim \frac{1}{M-j} + \frac{1}{M-|M-1-j|},$$

we get

$$|\sin(y/2)|\int_{-\pi}^{\pi}|g(x,y)|^2\,\mathrm{d}x \lesssim \log(M)\sum_{j=0}^{M-1}\left(\frac{1}{M-j} + \frac{1}{M-|M-1-j|}\right) \lesssim \log^2 M,$$

and

$$\sin^2(y/2)\int_{-\pi}^{\pi}|g(x,y)|^2\,\mathrm{d}x \lesssim \sum_{j=0}^{M-1}\left(\frac{1}{M-j} + \frac{1}{M-|M-1-j|}\right)^2 \leq K,$$

and (2) follows.

For (3), we rewrite $g(x, y)$ as

$$
g(x,y) = \frac{1}{M} \sum_{j=0}^{M-1} e^{\mathbf{i}j(x-y)} + \sum_{m=1}^{M-1} \frac{\sum_{i=0}^{M-m-1} e^{\mathbf{i}(ix+mx-iy)} + \sum_{j=0}^{M-m-1} e^{\mathbf{i}(jx-my-jy)}}{M-m}
$$

$$
= \frac{1}{M} \frac{1 - e^{\mathbf{i}M(x-y)}}{1 - e^{\mathbf{i}(x-y)}} + \frac{e^{\mathbf{i}Mx}}{1 - e^{\mathbf{i}(x-y)}} \sum_{m=1}^{M-1} \frac{e^{-\mathbf{i}(M-m)x} - e^{-\mathbf{i}(M-m)y}}{M-m}
$$

$$
+ \frac{e^{-\mathbf{i}My}}{1 - e^{\mathbf{i}(x-y)}} \sum_{m=1}^{M-1} \frac{e^{\mathbf{i}(M-m)y} - e^{\mathbf{i}(M-m)x}}{M-m}
$$

$$
= \frac{1}{M} \frac{1 - e^{\mathbf{i}M(x-y)}}{1 - e^{\mathbf{i}(x-y)}} + \frac{e^{\mathbf{i}Mx}}{1 - e^{\mathbf{i}(x-y)}} \sum_{k=1}^{M-1} \frac{e^{-\mathbf{i}kx} - e^{-\mathbf{i}ky}}{k}
$$

$$
+ \frac{e^{-\mathbf{i}My}}{1 - e^{\mathbf{i}(x-y)}} \sum_{k=1}^{M-1} \frac{e^{\mathbf{i}ky} - e^{\mathbf{i}kx}}{k}
$$

$$
=: g_1(x,y) + g_2(x,y) + g_3(x,y).
$$

$$(5.35)$$

We note that for any $x, y \in \mathbb{R}$,

$$
|g_1(x,y)| = \frac{1}{M} \left| \frac{\sin(M(x-y)/2)}{\sin((x-y)/2)} \right| \le 1.
$$

Let $z_1 = e^{-\mathbf{i}x}, z_2 = e^{-\mathbf{i}y}$, then we have

$$
|g_2(x,y)| = \left| \frac{1}{z_1 - z_2} \int_{[z_1,z_2]} \sum_{k=0}^{M-2} z^k \, \mathrm{d}z \right| \le \sup_{z \in [z_1,z_2]} \left| \frac{1 - z^{M-1}}{1 - z} \right| \le \frac{1}{|\sin \delta|},
$$

where $[z_1, z_2]$ denotes the segment between z_1 and z_2. The same bound also controls g_3, so (3) holds. $\qquad\square$

Now we are ready to prove Proposition 5.8.

Proof of Proposition 5.8. For $\delta \in (0, \pi/2)$, let $E_\delta := [-\pi, \pi] \setminus (-\delta, \delta)$. Then we have

$$
\operatorname{tr} \mathbf{Q}_M^2(\theta) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x+\theta) f(y+\theta) |g(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y
$$

$$
= \frac{1}{4\pi^2} \left(\int_{-\delta}^{\delta} \int_{-\delta}^{\delta} + \int_{E_\delta} \int_{E_\delta} + \int_{-\delta}^{\delta} \int_{E_\delta} + \int_{E_\delta} \int_{-\delta}^{\delta} \right) f(x+\theta) f(y+\theta) |g(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y
$$

$$
=: P_1 + P_2 + P_3 + P_4.
$$

$$(5.36)$$

For P_1, using Lemma 5.9 (1), we have

$$
|P_1| \le \|f\|_{\theta,\delta} \frac{1}{4\pi^2} \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} |g(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y \le 2\|f\|_{\theta,\delta} \log M.
$$

For P_2, using Lemma 5.9 (3), we have

$$|P_2| \lesssim \frac{1}{\sin^2 \delta} \int_{E_\delta} \int_{E_\delta} f(x+\theta)f(y+\theta) \, \mathrm{d}x \, \mathrm{d}y \lesssim \frac{\|f\|_1^2}{\delta^2}.$$

For P_3 and similarly for P_4, using Lemma 5.9 (2), we have

$$|P_3| \leq \|f\|_{(\theta,\delta)} \frac{1}{4\pi^2} \int_{E_\delta} f(y+\theta) \int_{-\delta}^\delta |g(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y \lesssim \frac{\|f\|_{(\theta,\delta)} \|f\|_1}{\delta^2},$$

and $|P_4|$ is controlled by the same bound.

Summing up the bounds for P_1, P_2, P_3, P_4 and dividing $8\pi^2 \log M$, the result follows. □

As a consequence of Proposition 5.8, if f is bounded in a neighborhood of a point or a set, then $\operatorname{tr} \mathbf{Q}_M^2(\theta)/\log M$ is (uniformly) bounded at this point or in this set.

To summarize what we have obtained, if f satisfies **A3**, **A4** and **A5**, then

$$\frac{\operatorname{tr} \mathbf{Q}_M^2(\theta)}{f^2(\theta)\log M} \tag{5.37}$$

is bounded uniformly in $M \geq 1$ and

1. in $\theta \in [-\frac{A}{M}, \frac{A}{M}]$ for any $A > 0$, using the global bound (5.31) and Lemma 5.7(b).
2. in $\theta \in [-\pi, -\delta] \cup [\delta, \pi]$ for any $\delta \in (0, \pi/2)$, using Proposition 5.8.

Therefore, using a classic argument, we can find two sequences of positive numbers $1/M \ll \tau_M < \delta_M \ll 1$ such that (5.12) is uniformly bounded in $[-\pi, -\delta_M] \cup [-\tau_M, \tau_M] \cup [\delta_M, \pi]$. In order to complete the proof of Proposition 5.3, it remains to prove the uniform boundedness of (5.12) for $|\theta| \in (\tau_M, \delta_M)$.

For any such θ, suppose that $\theta > 0$ without loss of generality, and denote $E_\theta := [-\pi, \pi] \setminus (-\frac{\theta}{2}, \frac{\theta}{2})$. We write

$$\operatorname{tr} \mathbf{Q}_M^2(\theta) = \frac{1}{4\pi^2} \int_{-\pi}^\pi \int_{-\pi}^\pi f(x)f(y)|g(x-\theta, y-\theta)|^2 \, \mathrm{d}x \, \mathrm{d}y$$
$$=: P_1 + P_2 + P_3 + P_4, \tag{5.38}$$

where the P_i's are the four sub-integrals defined via the following partition of the double integral

$$\int_{-\pi}^\pi \int_{-\pi}^\pi = \int_{E_\theta} \int_{E_\theta} + \int_{-\frac{\theta}{2}}^{\frac{\theta}{2}} \int_{-\frac{\theta}{2}}^{\frac{\theta}{2}} + \int_{-\frac{\theta}{2}}^{\frac{\theta}{2}} \int_{E_\theta} + \int_{E_\theta} \int_{-\frac{\theta}{2}}^{\frac{\theta}{2}}.$$

For P_1, by Lemma 5.7 (a) and the definition of regularly varying functions, with some $\varepsilon > 0$, whenever θ is small enough,

$$\sup_{x \in E_\theta} f(x) \leq (1+\varepsilon)f(\theta/2) \leq 2^a(1+\varepsilon)^2 f(\theta). \tag{5.39}$$

Thus by Lemma 5.9(1),

$$|P_1| \lesssim \frac{f^2(\theta)}{4\pi^2} \int_{E_\theta} \int_{E_\theta} |g(x-\theta, y-\theta)|^2 \, dx \, dy \lesssim f^2(\theta) \log M.$$

For P_2, by Lemma 5.9(3) and Lemma 5.7(c), we have

$$|P_2| \lesssim \sup_{x,y \in [-\frac{3\theta}{2}, -\frac{\theta}{2}]} |g(x,y)|^2 \left(\int_{-\frac{\theta}{2}}^{\frac{\theta}{2}} f(x) \, dx \right)^2 \lesssim f^2(\theta).$$

For P_3, and similarly for P_4, using (5.39) again, and using Lemma 5.9(2) and Lemma 5.7(c), we have

$$|P_3| \lesssim f(\theta) \int_{-\frac{\theta}{2}}^{\frac{\theta}{2}} f(y) \int_{-\pi}^{\pi} |g(x,y)|^2 \, dx \, dy \lesssim f^2(\theta) \log^2 M.$$

The same bound also controls P_4.

The proof of Proposition 5.3 is complete by summing up the bounds for P_1, P_2, P_3, P_4.

6. Proof of Proposition 2.4

Define

$$r_k^b := \left(1 - \frac{|k|}{M} \right) r_k,$$

and

$$\hat{\Upsilon}_M^b(\theta) := \sum_{k=-M+1}^{M-1} \hat{r}_k^b e^{\mathbf{i}k\theta}, \quad \Upsilon_M^b(\theta) := \sum_{k=-M+1}^{M-1} r_k^b e^{\mathbf{i}k\theta}.$$

Note that Υ_M^b is the Cesàro mean of $\Upsilon_M(\theta) := \sum_{k=-M+1}^{M-1} r_k e^{\mathbf{i}k\theta}$, therefore

$$\Upsilon_M^b(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) F_M(\theta - x) \, dx, \qquad (6.1)$$

where $F_M(x) = \frac{\sin^2(Mx/2)}{M \sin^2(x/2)}$ is the Fejér kernel. Thus for any $\theta \in \mathbb{R}$, we have

$$\operatorname*{ess\,inf}_t f(t) \leq \Upsilon_M^b(\theta) \leq \operatorname*{ess\,sup}_t f(t). \qquad (6.2)$$

By **A4**, Υ_M^b is positive and uniformly lower bounded from 0.

Following the same idea as §5.1, we only need to estimate

$$\mathbb{P} \left(\sup_{\theta \in [0, 2\pi]} \left| \frac{\hat{\Upsilon}_M^b(\theta)}{\Upsilon_M^b(\theta)} - 1 \right| > x \right) \qquad (6.3)$$

for any $x > 0$.

We use the same discretization strategy as §5. For any fixed $\theta \in [0, 2\pi]$, define

$$d_M(\theta) = \frac{1}{\sqrt{M}}(1, e^{-\mathbf{i}\theta}, \ldots, e^{-\mathbf{i}(M-1)\theta})^\top, \quad \mathbf{Q}_M^b(\theta) := \mathbf{R}_M^{1/2} d_M(\theta) d_M^*(\theta) \mathbf{R}_M^{1/2}$$

then by Lemma 3 of [27], we have

$$\hat{\Upsilon}_M^b(\theta) = \frac{1}{N} d_M^*(\theta) \mathbf{R}_M^{1/2} \mathbf{Z}^* \mathbf{C}_N \mathbf{Z} \mathbf{R}_M^{1/2} d_M(\theta) = \frac{1}{N} \sum_{n=1}^{N} c_n \mathbf{z}_n \mathbf{Q}_M^b(\theta) \mathbf{z}_n^*,$$

and by the unitary invariance of $\mathbf{z}_n$, note also that $\mathbf{Q}_M^b(\theta)$ is of rank one with a positive eigenvalue $d_M^*(\theta) \mathbf{R}_M d_M(\theta) = \Upsilon_M^b(\theta)$, we then have

$$\frac{\hat{\Upsilon}_M^b(\theta)}{\Upsilon_M^b(\theta)} \overset{\mathcal{L}}{=} \frac{1}{N} \sum_{n=1}^{N} c_n |Z_{n,1}|^2.$$

Then

$$\mathbb{P}\left(\left| \frac{\hat{\Upsilon}_M^b(\theta)}{\Upsilon_M^b(\theta)} - 1 \right| > x \right) = \mathbb{P}\left(\left| \frac{1}{N} \sum_{n=1}^{N} c_n(|Z_{n,1}|^2 - 1) \right| > x \right).$$

Using the same method as §5.2, we get the concentration inequality

$$\mathbb{P}\left(\left| \frac{1}{N} \sum_{n=1}^{N} c_n(|Z_{n,1}|^2 - 1) \right| > x \right) \leq 2 \exp\left(-\frac{KNx^2}{\kappa^2 \log^2 M} \right)$$

for some constant $K > 0$, for any $x > 0$ and M, N large enough.

For the discretization step, we use the same method as §5.3, and the proof of Lemma 4, Lemma 6 in [27], along with the norm bound $\|\mathbf{R}_M\| \lesssim M^a L(M)$. Note also that $\Upsilon_M^b(\theta)$ are positive and uniformly lower bounded from 0. We finally get

$$\mathbb{P}\left(\operatorname*{ess\,sup}_\theta \left| \frac{\hat{\Upsilon}_M^b(\theta)}{\Upsilon_M^b(\theta)} - 1 \right| > x \right) \leq 2M^\beta \exp\left(-\frac{KNx^2}{\kappa^2 \log^2 M} \right)$$

for some $\beta > 0$, $K > 0$ and any $x > 0$, large enough M, N. This implies that

$$\|(\mathbf{R}_M^b)^{-1/2} \hat{\mathbf{R}}_M^b (\mathbf{R}_M^b)^{-1/2} - \mathbf{I}\| \to 0$$

as $M, N \to \infty$ with $N \gg \log^3 M$.

Next we prove the inconsistency (2.9). We first prove that for two sequences of invertible matrices $\mathbf{R}_{1,M}, \mathbf{R}_{2,M}$, a necessary condition for the convergence

$$\|\mathbf{R}_{1,M}^{-1/2} \mathbf{R}_{2,M} \mathbf{R}_{1,M}^{-1/2} - \mathbf{I}\| \xrightarrow[M \to \infty]{} 0, \tag{6.4}$$

is

$$\lim_{M \to \infty} \frac{\lambda_{\max}(\mathbf{R}_{2,M})}{\lambda_{\max}(\mathbf{R}_{1,M})} = 1. \tag{6.5}$$

Take an arbitrary $\varepsilon > 0$. Let u be un eigenvector of $\mathbf{R}_{1,M}$ associated with $\lambda_{\max}(\mathbf{R}_{1,M})$, then from (6.4), for large enough M,

$$1 - \varepsilon < u^* (\mathbf{R}_{1,M})^{-1/2} \mathbf{R}_{2,M} (\mathbf{R}_{1,M})^{-1/2} u = \frac{u^* \mathbf{R}_{2,M} u}{\lambda_{\max}(\mathbf{R}_{1,M})} \leq \frac{\lambda_{\max}(\mathbf{R}_{2,M})}{\lambda_{\max}(\mathbf{R}_{1,M})}. \quad (6.6)$$

Note that $(\mathbf{R}_{1,M}^{-1/2} \mathbf{R}_{2,M} \mathbf{R}_{1,M}^{-1/2})^{-1}$ has the same eigenvalues as $\mathbf{R}_{2,M}^{-1/2} \mathbf{R}_{1,M} \mathbf{R}_{2,M}^{-1/2}$. Recall that for a sequence of Hermitian matrices $\mathbf{A}_M$, the convergence $\|\mathbf{A}_M - \mathbf{I}\| \to 0$ is equivalent to the convergence of its eigenvalues, i.e. $\lambda_{\max}(\mathbf{A}_M) \to 1$, $\lambda_{\min}(\mathbf{A}_M) \to 1$. Therefore (6.4) also implies that

$$\|(\mathbf{R}_{2,M})^{-1/2}(\mathbf{R}_{1,M})(\mathbf{R}_{2,M})^{-1/2} - \mathbf{I}\| \xrightarrow[M \to \infty]{} 0.$$

Using the same arguments as (6.6), we get, for large enough M,

$$1 - \varepsilon \leq \frac{\lambda_{\max}(\mathbf{R}_{1,M})}{\lambda_{\max}(\mathbf{R}_{2,M})}. \quad (6.7)$$

Combining (6.6) and (6.7), we have

$$\lim_{M \to \infty} \frac{\lambda_{\max}(\mathbf{R}_{2,M})}{\lambda_{\max}(\mathbf{R}_{1,M})} = 1, \quad (6.8)$$

and (6.5) follows. However, we will prove that almost surely (6.5) cannot be satisfied by $\hat{\mathbf{R}}_M^b$ and $\mathbf{R}_M$. Indeed, from (2.8) we conclude that almost surely

$$\frac{\lambda_{\max}(\hat{\mathbf{R}}_M^b)}{\lambda_{\max}(\mathbf{R}_M^b)} \to 1.$$

Thus we only need to prove that

$$\frac{\lambda_{\max}(\mathbf{R}_M^b)}{\lambda_{\max}(\mathbf{R}_M)} \nrightarrow 1. \quad (6.9)$$

Let $\mathcal{K}$ and $\mathcal{K}^b$ be two integral operators acting on $L^2(0,1)$ defined by

$$\mathcal{K}(\varphi)(x) = \int_0^1 \frac{1}{|x-y|^{1-a}} \varphi(y) \, \mathrm{d}y, \quad \mathcal{K}^b(\varphi)(x) = \int_0^1 \frac{1 - |x-y|}{|x-y|^{1-a}} \varphi(y) \, \mathrm{d}y.$$

Lemma 6.1. *Under the same assumptions as Proposition 2.4, as $M \to \infty$,*

$$\frac{\lambda_{\max}(\mathbf{R}_M)}{KM^a L(M)} \to \lambda_1(\mathcal{K}), \quad \frac{\lambda_{\max}(\mathbf{R}_M^b)}{KM^a L(M)} \to \lambda_1(\mathcal{K}^b)$$

with some absolute constant $K > 0$.

Proof. We first assume that the slowly varying function L in **A5** equals to 1. Then from [18, Proposition 2.2.16],

$$r_k \sim \frac{K}{(1 + |k|)^{1-a}}$$

as $k \to \infty$ with some absolute constant $K > 0$. Using Widom-Shampine's Lemma ([16, Lemma 5.1]) and the same method as the proof of [16, Theorem 2.3], one can prove that

$$\frac{\lambda_{\max}(\mathbf{R}_M)}{KM^a} \to \lambda_1(\mathcal{K}), \quad \frac{\lambda_{\max}(\mathbf{R}_M^b)}{KM^a} \to \lambda_1(\mathcal{K}^b). \tag{6.10}$$

If the slowly varying function L in **A5** is not constant, let

$$\tilde{f}(\theta) = \frac{1}{|\theta|^a}, \quad \theta \in [-\pi, \pi]$$

and $\tilde{\Upsilon}_M^b$, $\tilde{\mathbf{R}}_M^b$ be defined with $\tilde{f}$ in the same way as Υ_M^b, $\mathbf{R}_M^b$ with f. Note that the Féjer kernel F_M has the same upper bound as the Dirichlet kernel $D_M(\theta) = \sin((M+1/2)\theta)/\sin(\theta/2)$ used in the proof of [25, Theorem 2.3], that is, for $\theta \in [-3\pi/2, 3\pi/2]$,

$$|F_M(\theta)| = \left| \frac{D_0(\theta) + \cdots + D_{M-1}(\theta)}{M} \right| \lesssim \min\{M, |\theta|^{-1}\}.$$

Then using the same technique there, one can prove that

$$\sup_\theta \left\| \frac{\Upsilon_M^b(\theta)}{M^a L(M)} - \frac{\tilde{\Upsilon}_M^b(\theta)}{M^a} \right\| \to 0$$

as $M \to \infty$, which implies that

$$\left\| \frac{\mathbf{R}_M^b}{M^a L(M)} - \frac{\tilde{\mathbf{R}}_M^b}{M^a} \right\| \to 0.$$

Also note that by Theorem 2.3 of [25],

$$\left\| \frac{\mathbf{R}_M}{M^a L(M)} - \frac{\tilde{\mathbf{R}}_M}{M^a} \right\| \to 0,$$

together with (6.10), we have

$$\frac{\lambda_{\max}(\mathbf{R}_M)}{KM^a L(M)} \to \lambda_1(\mathcal{K}), \quad \frac{\lambda_{\max}(\mathbf{R}_M^b)}{KM^a L(M)} \to \lambda_1(\mathcal{K}^b). \qquad \square$$

From this lemma, we have

$$\frac{\lambda_{\max}(\mathbf{R}_M^b)}{\lambda_{\max}(\mathbf{R}_M)} \to \frac{\lambda_1(\mathcal{K}^b)}{\lambda_1(\mathcal{K})}. \tag{6.11}$$

We then prove that $\lambda_1(\mathcal{K}) > \lambda_1(\mathcal{K}^b)$. Indeed because the two integral kernels are positive, from the mini-max formula for the largest eigenvalue, their eigenfunctions associated with the largest eigenvalue are positive in $[0, 1]$. Let φ^b be the eigenfunction of $\mathcal{K}^b$ associated with $\lambda_1(\mathcal{K}^b)$, then

$$\lambda_1(\mathcal{K}^b) = \langle \varphi^b, \mathcal{K}^b \varphi^b \rangle = \langle \varphi^b, \mathcal{K} \varphi^b \rangle - \int_0^1 \int_0^1 |x - y|^a \varphi^b(x) \varphi^b(y) \, dx \, dy < \lambda_1(\mathcal{K}),$$

from which we conclude that

$$\lim_{M \to \infty} \frac{\lambda_{\max}(\mathbf{R}_M^b)}{\lambda_{\max}(\mathbf{R}_M)} = \frac{\lambda_1(\mathcal{K}^b)}{\lambda_1(\mathcal{K})} < 1. \tag{6.12}$$

This is the end of the proof of proposition since (6.9) is proved.

Appendix A: Additional proofs

The proof of Corollary 2.3 is in §A.1; the proof of Proposition 2.5 is in §A.2; and the proof of Proposition 3.3 is in §A.3.

A.1. Proof of Corollary 2.3

We write

$$\left\| \hat{\mathbf{R}}_M^{1/2} \mathbf{R}_M^{-1/2} - \sqrt{\xi_N}\mathbf{I} \right\| \le \|\mathbf{R}_M^{1/4}\| \left\| \mathbf{R}_M^{-1/4}\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/4} - \sqrt{\xi_N}\mathbf{I} \right\| \|\mathbf{R}_M^{-1/4}\|, \tag{A.1}$$

where $\|\mathbf{R}_M^{-1/4}\|$ is bounded, and from Lemma 5.6, $\|\mathbf{R}_M^{1/4}\| = O(M^{1/4+\varepsilon})$ with any $\varepsilon \in (0, 1/8)$.

The spectral norm $\left\| \mathbf{R}_M^{-1/4}\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/4} - \sqrt{\xi_N}\mathbf{I} \right\|$ equals to

$$\max \left\{ \left| \lambda_{\max}(\mathbf{R}_M^{-1/4}\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/4}) - \sqrt{\xi_N} \right|, \left| \lambda_{\min}(\mathbf{R}_M^{-1/4}\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/4}) - \sqrt{\xi_N} \right| \right\}.$$

The positive definite Hermitian matrix $\mathbf{R}_M^{-1/4}\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/4}$ has the same eigenvalues as $\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/2}$, so the latter matrix has M positive eigenvalues. On the other hand, all the eigenvalues of $\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/2}$ are between its smallest and largest singular values, that is,

$$\sqrt{\lambda_{\min}(\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2})} \le \lambda_{\min}(\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/2})$$
$$\le \lambda_{\max}(\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/2}) \le \sqrt{\lambda_{\max}(\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2})}.$$

From Theorem 2.2, if ξ_N is bounded away from 0, as $N, M \to \infty$ with $N/M \to c \in (0, \infty)$, we have almost surely

$$\left| \sqrt{\lambda_{\max}(\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2})} - \sqrt{\xi_N} \right| = O(M^{-1/2+\varepsilon}).$$

The same result also holds for $\sqrt{\lambda_{\min}(\mathbf{R}_M^{-1/2}\hat{\mathbf{R}}_M\mathbf{R}_M^{-1/2})}$. Therefore, we have almost surely

$$\left\| \mathbf{R}_M^{-1/4}\hat{\mathbf{R}}_M^{1/2}\mathbf{R}_M^{-1/4} - \sqrt{\xi_N}\mathbf{I} \right\| = O(M^{-1/2+\varepsilon}).$$

Taking the above estimations into (A.1), the result follows.

A.2. Proof of Proposition 2.5

We first recall the LSD of Toeplitz matrices. If a sequence of Toeplitz matrices $(\mathbf{R}_M = (r_{i-j})_{i,j=1}^M)_{M \geq 1}$ have a real spectral density $f \in L^1(-\pi, \pi)$, then by a generalized version of Szegő's Theorem [3, Theorem 2], for any continuous function φ defined on $\mathbb{R}$ such that $\varphi(x)/(1 + |x|)$ is bounded, we have

$$\lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^M \varphi(\lambda_k(\mathbf{R}_M)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(f(\theta)) \, d\theta. \qquad (A.2)$$

In particular, the LSD $\mu^{\mathbf{R}}$ of $\mathbf{R}_M$ will be defined by the identity

$$\int \varphi \, d\mu^{\mathbf{R}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(f(\theta)) \, d\theta, \quad \forall \varphi \in C_b(\mathbb{R}), \qquad (A.3)$$

where $C_b(\mathbb{R})$ denotes the set of bounded continuous functions on $\mathbb{R}$.

In order to prove Proposition 2.5, we are led to a general result relating the ratio ESD of two Toeplitz matrices with their spectral densities, which may be of independent interest.

Lemma A.1. *Let $(\mathbf{R}_M = (r_{i-j})_{i,j=1}^M)_{M \geq 1}$ be a sequence of Toeplitz matrices with real spectral density $f \in L^1(-\pi, \pi)$. Let $(f_M)_{M \geq 1}$ be a sequence of real functions in $L^1(-\pi, \pi)$. Let $\mathbf{R}_M^{(M)} = (r_{i-j}^{(M)})_{i,j=1}^M$ with*

$$r_k^{(M)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_M(\theta) e^{\mathbf{i} k\theta} \, d\theta$$

the Fourier coefficients of f_M. Let $\mu^{\mathbf{R}}$ denote the LSD of $\mathbf{R}_M$ defined in (A.3).

1. If $\|f_M - f\|_1 \to 0$, then

$$\mu^{\mathbf{R}_M^{(M)}} \xrightarrow{\mathcal{D}} \mu^{\mathbf{R}}.$$

2. In addition to (1), if moreover f is positive and bounded away from 0, and $f_M/f > a$ for some $a \in \mathbb{R}$, then

$$\mu^{\mathbf{R}_M^{(M)} \mathbf{R}_M^{-1}} \xrightarrow{\mathcal{D}} \delta_1.$$

First note that by normalization, we can assume $\xi_N = 1$ for all $N \in \mathbb{N}$ without loss of generality. Then from the proof of Proposition 2.4, almost surely, as $M, N \to \infty$ with $N \gg \log^3 M$,

$$\sup_{\theta} \left| \frac{\hat{\Upsilon}_M^b(\theta)}{\Upsilon_M^b(\theta)} - 1 \right| \to 0.$$

Then almost surely

$$\int_0^{2\pi} |\hat{\Upsilon}_M^b(\theta) - \Upsilon_M^b(\theta)| \, d\theta \leq \sup_{\theta} \left| \frac{\hat{\Upsilon}_M^b(\theta)}{\Upsilon_M^b(\theta)} - 1 \right| \int_0^{2\pi} |\Upsilon_M^b(\theta)| \, d\theta \to 0,$$

where $\int_0^{2\pi} |\Upsilon_M^b(\theta)|\, d\theta$ is bounded because $\Upsilon_M^b(\theta)$ is the Cesàro mean of the Fourier series of f, and it is well known that Υ_M^b converges to f in $L^1(0, 2\pi)$. Then we deduce that almost surely

$$\int_0^{2\pi} |\hat{\Upsilon}_M^b(\theta) - f(\theta)|\, d\theta \to 0.$$

Also note that $\hat{\Upsilon}_M^b/f \geq 0$, then $\hat{\Upsilon}_M^b$ and f satisfy the conditions of Lemma A.1. Therefore the result of Proposition 2.5 is a corollary of Lemma A.1.

It remains to prove Lemma A.1. If f_M converges in $L^1(0, 2\pi)$ to f, we denote

$$(f_M - f)_+ = \max\{f_M - f, 0\}, \quad (f_M - f)_- = \max\{f - f_M, 0\},$$

and

$$\mathbf{A}_+ = \left(\frac{1}{2\pi} \int_0^{2\pi} (f_M - f)_+(\theta) e^{\mathbf{i}(i-j)\theta}\, d\theta \right)_{i,j=1}^M,$$

$$\mathbf{A}_- = \left(\frac{1}{2\pi} \int_0^{2\pi} (f_M - f)_-(\theta) e^{\mathbf{i}(i-j)\theta}\, d\theta \right)_{i,j=1}^M.$$

Then $\mathbf{A}_+, \mathbf{A}_-$ are two positive semi-definite Toeplitz matrices satisfying

$$\frac{1}{M} \operatorname{tr} \mathbf{A}_\pm = \frac{1}{2\pi} \int_0^{2\pi} (f_M - f)_\pm\, d\theta \to 0.$$

It is easy to prove that there exists a sequence of positive numbers $(\varepsilon_M)_{M \geq 1}$ converging to 0, such that

$$\frac{\#\{k\,:\,\lambda_k(\mathbf{A}_\pm) > \varepsilon_M\}}{M} \leq \varepsilon_M$$

where "$\#S$" denotes the cardinal of the set S. From (A.2) we already have $\mu^{\mathbf{R}_M} \xrightarrow{\mathcal{D}} \mu^{\mathbf{R}}$. We next prove successively that

$$\mu^{(\mathbf{R}_M + \mathbf{A}_+)} \xrightarrow{\mathcal{D}} \mu^{\mathbf{R}} \quad \text{and} \quad \mu^{(\mathbf{R}_M + \mathbf{A}_+ - \mathbf{A}_-)} \xrightarrow{\mathcal{D}} \mu^{\mathbf{R}},$$

which is the first result of the lemma.

Let $\mathbf{A}_+ = U \operatorname{diag}(\lambda_1, \ldots, \lambda_M) U^*$ be a diagonalization of $\mathbf{A}_+$ with $\lambda_1, \ldots, \lambda_M$ its eigenvalues. Let $\mathbf{A}_+^{(1)} = U \operatorname{diag}(\lambda_1 \mathbb{1}_{\lambda_1 > \varepsilon_M}, \ldots, \lambda_M \mathbb{1}_{\lambda_M > \varepsilon_M}) U^*$, and $\mathbf{A}_+^{(2)} = U \operatorname{diag}(\lambda_1 \mathbb{1}_{\lambda_1 \leq \varepsilon_M}, \ldots, \lambda_M \mathbb{1}_{\lambda_M \leq \varepsilon_M}) U^*$. Then the rank of $\mathbf{A}_+^{(1)}$ is at most $M\varepsilon_M$, and $\|\mathbf{A}_+^{(2)}\| \leq \varepsilon_M$. Using Theorem A.43 and A.45 in [1] successively, we can prove that

$$\mu^{\mathbf{R}_M + \mathbf{A}_+} \xrightarrow{\mathcal{D}} \mu^{\mathbf{R}}.$$

Repeating the same procedure, we have also

$$\mu^{\mathbf{R}_M + \mathbf{A}_+ - \mathbf{A}_-} \xrightarrow{\mathcal{D}} \mu^{\mathbf{R}}.$$

Thus the first part of the lemma is proved.

Next we prove the second part. We will first prove the following lemma.

Lemma A.2. *Suppose that the probability measures $\mu_n(n = 1, 2, \dots)$ and μ are supported on $[a, +\infty)$. If for any $x < a$,*

$$\lim_{n \to \infty} \int \log |x - t| \, \mathrm{d}\mu_n(t) = \int \log |x - t| \, \mathrm{d}\mu(t) < \infty, \tag{A.4}$$

then μ_n converges weakly to μ.

Proof. From every subsequence of (μ_n) we can extract a subsequence converging vaguely to a positive measure ν with total mass less than or equal to 1. Take an arbitrary $x_0 < a$. Then for any $z \in \mathbb{C}\backslash(x_0, +\infty)$, because the function $t \mapsto (z - t)^{-1}$ is continuous on the support of μ_n and ν, and tends to 0 as $t \to \infty$, we have the convergence of Stieltjes transform

$$s_n(z) := \int \frac{1}{z - t} \, \mathrm{d}\mu_n(t) \xrightarrow[n \to \infty]{} \int \frac{1}{z - t} \, \mathrm{d}\nu(t) =: s(z).$$

By dominated convergence theorem, we have

$$\int_{x_0}^{z} s_n(w) \, \mathrm{d}w \xrightarrow[n \to \infty]{} \int_{x_0}^{z} s(w) \, \mathrm{d}w,$$

where the integral is taken along the segment from x_0 to z. Changing the order of integrals, we get

$$\int (\log(z - t) - \log(x_0 - t)) \, \mathrm{d}\mu_n(t) \xrightarrow[n \to \infty]{} \int_{x_0}^{z} s(w) \, \mathrm{d}w,$$

where $\log z = \log |z| + i \arg z$ with $\arg z \in [0, 2\pi)$. When $z = x \in (-\infty, x_0]$, the above convergence and the condition (A.4) imply that

$$\int_{x_0}^{x} s(w) \, \mathrm{d}w = \int \log(x - t) \, \mathrm{d}\mu(t) - \int \log(x_0 - t) \, \mathrm{d}\mu(t).$$

Extending this equality by analyticity, we have, for $z \in \mathbb{C}\backslash(x_0, +\infty)$,

$$\int_{x_0}^{z} s(w) \, \mathrm{d}w = \int \log(z - t) \, \mathrm{d}\mu(t) - \int \log(x_0 - t) \, \mathrm{d}\mu(t).$$

Differentiating both sides, we get

$$s(z) = \int \frac{1}{z - t} \, \mathrm{d}\mu(t).$$

This implies that $\mu = \nu$. Then because the vague limit μ is a probability measure, we actually have the weak convergence $\mu_n \xrightarrow{\mathcal{D}} \mu$ and Lemma A.2 is proved.

We continue the proof of Lemma A.1. By Lemma A.2, we only need to prove

$$\lim_{M \to \infty} \frac{1}{M} \log \left| \det(\mathbf{R}_M^{(M)} - x\mathbf{R}_M) \right| = \frac{1}{2\pi} \int_0^{2\pi} \log |f(\theta) - xf(\theta)| \, \mathrm{d}\theta. \tag{A.5}$$

for $x \in (-\infty, a)$. From the condition $f_M/f > a, f > 0$, we get $f_M - xf > 0$ for $x < a$, thus the Toeplitz matrix $\mathbf{R}_M^{(M)} - x\mathbf{R}_M$ is positive definite. For any $A > 1$ and $t > 0$, let

$$\ell_A(t) := \min(\log(t), A), h_A(t) := \max(\log(t), A) - A.$$

Then it is easily seen that $\ell_A(t) + h_A(t) = \log(t)$, and $\ell_A(t) \to \log(t), h_A(t) \to 0$ for every $t > 0$ when $A \to \infty$. Note that $f_M - xf \to (1-x)f$ in L^1, then from the first part of the lemma, we have

$$\frac{1}{M} \sum_{k=1}^{M} \ell_A(\lambda_k(\mathbf{R}_M^{(M)} - x\mathbf{R}_M)) \xrightarrow[M \to \infty]{} \frac{1}{2\pi} \int_0^{2\pi} \ell((1-x)f(\theta)) \, \mathrm{d}\theta. \qquad (A.6)$$

Note that if A is large, $h_A(t) < \sqrt{t}$, thus

$$\frac{1}{M} \sum_{k=1}^{M} h_A(\lambda_k(\mathbf{R}_M^{(M)} - x\mathbf{R}_M)) \leq \frac{1}{M} \sum_{k=1}^{M} (\lambda_k(\mathbf{R}_M^{(M)} - x\mathbf{R}_M))^{\frac{1}{2}} \mathbb{1}_{\lambda_k(\mathbf{R}_M^{(M)} - x\mathbf{R}_M) > A}$$

$$\leq \frac{1}{M\sqrt{A}} \operatorname{tr}(\mathbf{R}_M^{(M)} - x\mathbf{R}_M)$$

$$= \frac{1}{2\pi\sqrt{A}} \int_0^{2\pi} (f_M - xf)(\theta) \, \mathrm{d}\theta$$

$$(A.7)$$

Summing (A.6) and (A.7), and let $A \to +\infty$, we get

$$\lim_{M \to \infty} \frac{1}{M} \log \left| \det(\mathbf{R}_M^{(M)} - x\mathbf{R}_M) \right| = \frac{1}{2\pi} \int_0^{2\pi} \log |f(\theta) - xf(\theta)| \, \mathrm{d}\theta. \qquad (A.8)$$

On the other hand, by [3, Theorem 2] and the hypothesis that f is positive and bounded away from 0, we have

$$\lim_{M \to \infty} \frac{1}{M} \log |\det(\mathbf{R}_M)| = \frac{1}{2\pi} \int_0^{2\pi} \log |f(\theta)| \, \mathrm{d}\theta. \qquad (A.9)$$

Take the difference of (A.8) and (A.9), we get

$$\frac{1}{M} \log \left| \det(\mathbf{R}_M^{-1} \mathbf{R}_M^{(M)} - x\mathbf{I}) \right| \xrightarrow[M \to \infty]{} \log |1 - x|$$

for $x < a$. From Lemma A.2, we have $\mu^{\mathbf{R}_M^{-1}\mathbf{R}_M^{(M)}} \xrightarrow{\mathcal{D}} \delta_1$. The proof of Lemma A.1 is complete.

A.3. Proof of Proposition 3.3

By Proposition 3.1, we have

$$\|\mathbf{S}_w - \xi_N^{-1}\mathbf{S}_Y\| \xrightarrow{\text{a.s}} 0.$$

Let $\mathbf{v}_1, \ldots, \mathbf{v}_p$ and $\mathbf{v}_{w,1}, \ldots, \mathbf{v}_{w,p}$ be the eigenvectors associated with the p largest eigenvalues of $\mathbf{S}_Y$ and $\mathbf{S}_w$, respectively. Let $\mathbf{P}$ and $\mathbf{P}_w$ be the orthogonal projections into the subspace generated by $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ and $\{\mathbf{v}_{w,1}, \ldots, \mathbf{v}_{w,p}\}$, respectively. Using the integral formula for eigenprojections, we have

$$\mathbf{P}_w - \mathbf{P} = \frac{1}{2\pi\mathbf{i}} \int_\Gamma [(z - \mathbf{S}_w)^{-1} - (z - \xi_N^{-1}\mathbf{S}_Y)^{-1}] \, \mathrm{d}z,$$

where Γ is a contour on the complex plane surrounding the p largest eigenvalues of $\mathbf{S}_w$ and $\xi_N^{-1}\mathbf{S}_Y$, and keeping the other eigenvalues outside. Because there is a gap between bulk and spike eigenvalues, the distance between the contour and the spectrum of both matrices can be lower bounded. Therefore the resolvents $(z - \mathbf{S}_w)^{-1}$ and $(z - \xi_N^{-1}\mathbf{S}_Y)^{-1}$ are both uniformly bounded in M, N and $z \in \Gamma$. Using the formula $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, we deduce that

$$\|\mathbf{P}_w - \mathbf{P}\| \le \frac{1}{2\pi} \int_\Gamma \|(z - \mathbf{S}_w)^{-1}\| \|\mathbf{S}_w - \xi_N^{-1}\mathbf{S}_Y\| \|(z - \xi_N^{-1}\mathbf{S}_Y)^{-1}\| \, |\mathrm{d}z| \xrightarrow{\mathrm{a.s.}} 0.$$

On the other hand, by [30, Theorem 11.3], we know that the matrix $M^{-1/2}\mathbf{Y}$ is almost surely bounded in spectral norm. Together with Corollary 2.3, we have

$$\frac{1}{\sqrt{M}} \|\mathbf{Y}_w - \mathbf{Y}/\sqrt{\xi_N}\| \le \left\| \frac{1}{\sqrt{M}} \mathbf{Y} \right\| \|\mathbf{R}_M^{1/2}\hat{\mathbf{R}}_M^{-1/2} - \xi_N^{-1/2}\| \xrightarrow{\mathrm{a.s.}} 0.$$

Therefore,

$$\frac{1}{\sqrt{M}} \|\mathbf{Y}_{w,pc} - \mathbf{Y}_{pc}/\sqrt{\xi_N}\| \le \frac{1}{\sqrt{M}} \|\mathbf{P}_w - \mathbf{P}\| \|\mathbf{Y}_w\| + \frac{1}{\sqrt{M}} \|\mathbf{P}\| \|\mathbf{Y}_w - \mathbf{Y}/\sqrt{\xi_N}\| \xrightarrow{\mathrm{a.s.}} 0.$$

Then we automatically have

$$\frac{1}{\sqrt{M}} \|\mathbf{Y}_{w,pc} - \mathbf{Y}_{pc}/\sqrt{\xi_N}\|_F \xrightarrow{\mathrm{a.s.}} 0,$$

because $\mathrm{rank}(\mathbf{Y}_{w,pc} - \mathbf{Y}_{pc}/\sqrt{\xi_N}) \le \mathrm{rank}(\mathbf{Y}_{w,pc}) + \mathrm{rank}(\mathbf{Y}_{pc}) = 2p$.

Acknowledgments

References

[1] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral analysis of large dimensional random matrices* **20**. Springer. MR2567175
[2] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular variation* **27**. Cambridge University Press. MR0898871

[3] CAPIZZANO, S. (2002). Test functions, growth conditions and Toeplitz matrices. In *Proceedings of the Fourth International Conference on Functional Analysis and Approximation Theory* **2** 791–795. MR1975486

[4] COUILLET, R., GRECO, M. S., OVARLEZ, J. and PASCAL, F. (2015). 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)* 149–152. IEEE.

[5] DAI, W. (2004). Asymptotics of the sample mean and sample covariance of long-range-dependent series. *Journal of Applied Probability* 383–392. MR2057588

[6] DING, X. and YANG, F. (2021). Spiked separable covariance matrices and principal components. *The Annals of Statistics* **49** 1113–1138. MR4255121

[7] DOUKHAN, P., OPPENHEIM, G. and TAQQU, M. (2002). *Theory and applications of long-range dependence.* Springer Science & Business Media. MR1956041

[8] FANG, K., KOTZ, S. and NG, K. (2018). *Symmetric multivariate and related distributions.* CRC Press. MR1071174

[9] GIRAITIS, L., KOUL, H. L. and SURGAILIS, D. (2012). *Large sample inference for long memory processes.* World Scientific Publishing Company. MR2977317

[10] GRENANDER, U. and SZEGŐ, G. (2001). *Toeplitz forms and their applications* **321**. University of California Press. MR0094840

[11] ING, C.-K., CHIOU, H.-T. and GUO, M. (2016). Estimation of inverse autocovariance matrices for long memory processes. *Bernoulli* **22** 1301–1330. MR3474817

[12] JAY, E., SOLER, T., TERREAUX, E., OVARLEZ, J. P., PASCAL, F., DE PERETTI, P. and CHORRO, C. (2020). Improving portfolios global performance using a cleaned and robust covariance matrix estimate. *Soft Computing* 1–12.

[13] JOHNSTONE, I. M. and PAUL, D. (2018). PCA in high dimensions: An orientation. *Proceedings of the IEEE* **106** 1277–1292.

[14] LIU, W. and WU, W. (2010). Asymptotics of spectral density estimates. *Econometric Theory* 1218–1245. MR2660298

[15] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* **1** 457.

[16] MERLEVÈDE, F., NAJIM, J. and TIAN, P. (2019). Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications. *Linear Algebra and its Applications.* MR3948744

[17] MERLEVÈDE, F. and PELIGRAD, M. (2016). On the empirical spectral distribution for matrices with long memory and independent rows. *Stochastic Processes and their Applications* **126** 2734–2760. MR3522299

[18] PIPIRAS, V. and TAQQU, M. S. (2017). *Long-range dependence and self-similarity* **45**. Cambridge University Press. MR3729426

[19] RANGARAJAN, G. and DING, M. (2003). *Processes with long-range correlations: Theory and applications* **621**. Springer Science & Business Media.

[20] SERRA, S. (1998). On the extreme eigenvalues of Hermitian (block)

Toeplitz matrices. *Linear Algebra and Its Applications* **270** 109–129. MR1484077

[21] SHENG, H., CHEN, Y. and QIU, T. (2011). *Fractional processes and fractional-order signal processing: techniques and applications.* Springer Science & Business Media. MR2883096

[22] SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis* **55** 331–339. MR1370408

[23] TERREAUX, E., OVARLEZ, J. P. and PASCAL, F. (2017). Robust model order selection in large dimensional elliptically symmetric noise. *arXiv preprint arXiv:1710.06735.*

[24] TIAN, B., ZHANG, Y. and ZHOU, W. (2020). Tracy–Widom law for the largest eigenvalue of sample covariance matrix generated by VARMA. *Random Matrices: Theory and Applications* 2150022. MR4260217

[25] TIAN, P. (2022). Joint CLT for top eigenvalues of sample covariance matrices of separable high dimensional long memory processes. *Random Matrices: Theory and Applications* 2250032. MR4440258

[26] VALLET, P. and LOUBATON, P. (2017). On the performance of MUSIC with Toeplitz rectification in the context of large arrays. *IEEE Transactions on Signal Processing* **65** 5848–5859. MR3722965

[27] VINOGRADOVA, J., COUILLET, R. and HACHEM, W. (2015). Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection. *IEEE Trans. Signal Processing* **63** 4903–4913. MR3385845

[28] WOODROOFE, M. B. and VAN NESS, J. W. (1967). The maximum deviation of sample spectral densities. *The Annals of Mathematical Statistics* 1558–1569. MR0216717

[29] WU, W., HUANG, Y. and ZHENG, W. (2010). Covariances estimation for long-memory processes. *Advances in Applied Probability* **42** 137–157. MR2666922

[30] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large sample covariance matrices and high-dimensional data analysis. Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press. MR3468554

[31] ZHANG, L. (2006). Spectral analysis of large dimentional random matrices, PhD thesis, National University of Singapore.

[32] ZYGMUND, A. (2002). *Trigonometric series* **2**. Cambridge University Press. MR0236587