WHITENING LONG RANGE DEPENDENCE IN LARGE SAMPLE
COVARIANCE MATRICES OF MULTIVARIATE STATIONARY PROCESSES

PENG TIAN AND JIANFENG YAO

ABSTRACT. Let $X$ be an $N \times T$ data matrix which can be represented as $X = C_N^{1/2} Z R_T^{1/2}$ with $Z$ an $N \times T$ random matrix whose rows are spherically symmetric, $R_T$ a deterministic $T \times T$ positive definite Toeplitz matrix, and $C_N$ a deterministic $N \times N$ nonnegative definite matrix. In particular, $Z$ can have i.i.d standard Gaussian entries. We prove the weak consistency of an unbiased estimator $\hat{R}_T = (\hat{r}_{i,j})$ of $\xi_N R_T$ where $\xi_N = N^{-1} \text{tr} C_N$, $\hat{r}_k$ is the average of the entries on the $k$th diagonal of $T^{-1}X^*X$. When each row of $X$ are long range dependent, i.e. the spectral density of Toeplitz matrix $R_T$ is regularly varying at 0 with exponent $a \in (-1,0)$, we prove that although $\hat{R}_T$ may not be consistent in spectral norm, a weaker consistency of the form $\| R^{-1/2}_T \hat{R}_T R_T^{-1/2} - \xi_N I \| \overset{P}{\rightarrow} 0$ still holds when $N, T \to \infty$ with $N \gg \log^{3/2} T$. We also establish useful probability bounds for deviations of the above convergence. It is shown next that this is strong enough for the implementation of a whitening procedure. We then apply the above result to a complex Gaussian signal detection problem where $C_N$ is a finite rank perturbation of the identity.

1. INTRODUCTION

The society and technology is generating enormous amount of data every day to be analyzed. Most traditional statistic tools work well with low dimensional and i.i.d data only. However data in the real world are often high dimensional and highly correlated, which makes the traditional tools dramatically inaccurate. For this reason, studies on large dimensional random matrices and long memory processes have been very active in recent decades.

These developments have allowed, for example, to retrieve messages of signals in a set of large dimensional signal-plus-noise data. Let $x_1, \ldots, x_T$ be a set of received antenna data which are i.i.d $N$-dimensional random vector such that

\[ x_i = A s_i + \varepsilon_i, \]

where $A$ is a low rank deterministic matrix, $s_i \sim \mathcal{CN}(0, I_N)$ represents the signal, and $\varepsilon_i \sim \mathcal{CN}(0, \sigma^2 I_N)$ is a complex Gaussian white noise independent of $s_i$. By using the spiked eigenvalues of the sample covariance matrix

\[ S = \frac{1}{T} \sum_{i=1}^T x_i x_i^*, \]

we can detect the number of signals (i.e. the number of spiked eigenvalues of the population covariance matrix $\text{Cov}(x_i) = A \text{Cov}(s_i) A^* + \sigma^2 I_N$). See for example Section 11.6 of [21].

The model (2) can be represented in the form of random matrix as

\[ S = \frac{1}{T} X X^* = \frac{1}{T} C_N^{1/2} Z Z^* C_N^{1/2}, \]

where $X = C_N^{1/2} Z$, with $Z$ a $N \times T$ random matrix with i.i.d standard complex Gaussian entries and $C_N$ a $N \times N$ deterministic positive semi-definite Hermitian matrices. In the above example, $C_N = A \text{Cov}(s_i) A^* + \sigma^2 I_N$ is a finite rank perturbation of the scaled identity matrix. The results
in random matrix theory have showed that when the dimension $N$ grows in the same pace with
the sample size $T$, $S$ is no longer a good estimator of $C_N$. Let $\lambda_1(S) \geq \cdots \geq \lambda_N(S)$ denote
the eigenvalues of $S$, and define the empirical spectral distribution (ESD) of $S$ as
\[
\mu^S := \frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j(S)}.
\]
It is well known that whenever the ESD of $C_N$ converges weakly, then as $N,T \to \infty$ with $N/T \to c \in (0, \infty)$, with probability one, the ESD $\mu^S$ converges weakly to the Marchenko-Pastur (MP) law.
See [12, 17, 16]. The spiked eigenvalues of $S$ are also well studied in this high dimensional regime.
Precisely, only the spiked eigenvalues of $C_N$ larger than $\sigma^2(1 + \sqrt{c})$ generate spiked eigenvalues of $S$. Suppose that $\alpha > \sigma^2(1 + \sqrt{c})$ is a spiked eigenvalue of $C_N$ of multiplicity $m \geq 1$, it generates $m$
spiked eigenvalues of $S$ converging to $\alpha + \frac{\sigma^2 \alpha}{\alpha - \sigma^2}$, fluctuating in the scale of $\sqrt{n}$ like the eigenvalues
of a $m \times m$ GUE matrix. See [2, 9, 10, 1, 11].
Sometimes the i.i.d assumption in the time domain cannot be satisfied. One way to release this
assumption is to consider the separable model
\[
X = C^\frac{1}{2}_N Z R^\frac{1}{2}_T
\]
and the corresponding sample covariance matrix
\[
S = \frac{1}{T} C^\frac{1}{2}_N Z R_T Z^* C^\frac{1}{2}_N.
\]
In this case, Zhang [22] studied the limiting spectral distribution (LSD) of $S$ and proved that the
Stieltjes transforms of the LSD of $S$ are uniquely determined by a system of equations. The spiked
eigenvalues of this separable model for general $C_N$, $R_T$ appears very recently, in particular, in the
case where $C_N$, $R_T$ are both bounded in spectral norm, the spiked eigenvalues of $S$ are studied in [6]; and in the case where $R_T$ is a Toeplitz matrix with unbounded spectral norm, the largest
eigenvalues of $S$ are studied in [19]. For two random variables $X, Y$, $X \preceq Y$ means that they follow
the same distribution.
Note that a spiked eigenvalue of the separable model (4) can be caused by either $C_N$ or $R_T$. It
is thus difficult to determine the spikes of $C_N$ from $S$. For example, in the case where the largest
eigenvalues of $R_T$ tend to infinity as in [19], the largest $m$ eigenvalues of $S$ are asymptotically
equivalent to
\[
\frac{\text{tr} C_N}{T} \lambda_1(R_T), \ldots, \frac{\text{tr} C_N}{T} \lambda_m(R_T)
\]
as $N,T \to \infty$ and $N/T \to c \in (0, \infty)$. In other words, only the summary statistic $\text{tr} C_N$ appears in these limits, and there is thus no way to recover the spike eigenvalues of $C_N$.
In order to resolve this problem, in the case where the process $(x_t)_{t \in \mathbb{Z}}$ is stationary with short
range dependence (SRD), or equivalently, the time domain covariance matrix $R_T = (r_{i,j})_{i,j=1}^T$ is
Toeplitz with absolutely summable coefficients $\sum_{t=-\infty}^{\infty} |r_t| < \infty$, some authors [20, 5, 18] suggest to whiten the correlation of columns of data $X$ using some estimator of $R_T$. Precisely, let
\[
S := \frac{1}{N} X^* X = \frac{1}{N} R_T^{1/2} Z^* C_N Z R_T^{1/2} =: (s_{i,j})_{i,j=1}^T,
\]
and define two estimators
\[
R_T := (\hat{r}_{i,j})_{1 \leq i,j \leq T}, \quad R_T^b := (\hat{r}_{i,j}^b)_{1 \leq i,j \leq T}
\]
with
\[
\hat{r}_k := \frac{1}{T-|k|} \sum_{1 \leq i+j \leq T} s_{i+k,j} \mathbbm{1}_{1 \leq i+k \leq T, 1 \leq j \leq T}, \quad \hat{r}_k^b := \frac{1}{T} \sum_{1 \leq i+j \leq T} s_{i+k,j} \mathbbm{1}_{1 \leq i+k \leq T, 1 \leq j \leq T}.
\]
The difference between the two estimators is that $\hat{r}_k$ are unbiased estimators of their population
counter-parts, while $\hat{r}_k^b$ are biased ones. Let $\xi_N := N^{-1} \text{tr} C_N$. Note that $E S = \xi N R_T$, we can see that up to a normalization scalar $\xi_N$, $\hat{R}_T$ is an unbiased estimator of $R_T$, while $\hat{R}_T^b$ is a biased one.
In [20, 5] it is showed that if $C_N$ is bounded in spectral norm, then both estimators are consistent in spectral norm:

$$
\| \hat{R}_T \text{ (or } \hat{R}_T^b) - \xi_N R_T \| \xrightarrow{a.s.} 0 .
$$

Then we can use the following time-whitened sample covariance matrix

$$
\tilde{S} := \frac{1}{T} X \hat{R}_T^{-1} X^* , \quad \text{or} \quad \tilde{S}^b := \frac{1}{T} X (\hat{R}_T^b)^{-1} X^*
$$

to estimate the spiked eigenvalues of $C_N$. Let

$$
S^w := \frac{1}{N} C_N^{1/2} Z Z^* C_N^{-1/2} .
$$

Then thanks to (7), we have

$$
\| \tilde{S} \text{ (or } \tilde{S}^b) - \xi_N^{-1} S^w \| \xrightarrow{a.s.} 0 ,
$$

from which we conclude that the LSD and the asymptotics of spiked eigenvalues of $S$ or $S^b$ are the same as $S^w$.

In this paper we study this whitening procedure in the context of long range dependent (LRD) stationary processes, that is, the Toeplitz matrix $R_T = (r_{i-j})$ has a spectral density $f$ in the form

$$
f(x) = |x|^{-a} L(|x|^{-1}) \quad \text{for} \quad x \in [-\pi, \pi],
$$

where $a \in (0, 1)$ and $L$ is a function defined on $[\pi^{-1}, \infty)$ and is slowly varying at $\infty$. Such situation is frequent in applications. For example, the covariance of received signal of two elements of a uniform antenna array is

$$
r_{i-j} = \frac{1}{\theta_{\text{max}} - \theta_{\text{min}}} \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} e^{i 2 \pi d |i-j| \cos \theta} d\theta .
$$

where $[\theta_{\text{min}}, \theta_{\text{max}}]$ is the range of arrival angles, $\lambda$ is the length of wave, and $d$ is the distance between two successive antennas. To have an illustrative example, we assume that $\theta_{\text{min}} = 0$, $\theta_{\text{max}} = \pi$, $2d = \lambda$, then by changing variable as $\pi \cos \theta = s$, we have

$$
r_{i-j} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i |i-j| s} \frac{2\pi}{\sqrt{\pi^2 - s^2}} ds .
$$

Then the Toeplitz matrices have the spectral density $f(s) = 2\pi/\sqrt{\pi^2 - s^2}$ which satisfies $f(s) \sim \sqrt{2\pi/|\pi \mp s|}$ as $s \to \pm \pi$. Noting Remark 2.1, this is an example of LRD process in the space-domain.

Let the data matrix $X$ be defined as in (3) with the spectral density $f$ of $R_T$ satisfying (10). Suppose also that $f$ is lower bounded. Let $\hat{R}_T$ be defined as (5). We will establish a concentration inequality for the spectrum of $R_T^{-1/2} \hat{R}_T R_T^{-1/2}$, which implies that

$$
\| R_T^{-1/2} \hat{R}_T R_T^{-1/2} - \xi_N I \| = O \left( \frac{\log^{3/2} T}{\sqrt{N}} \right)
$$

almost surely as $N, T \to \infty$ with $N \gg \log^3 T$. In this regime we have obviously

$$
\| R_T^{-1/2} \hat{R}_T R_T^{-1/2} - \xi_N I \| \xrightarrow{a.s.} 0 .
$$

Note that here $N$ is required to be large but can be much smaller than $T$.

With (13), the whitening procedure (8) is still possible for large dimensional LRD processes. Let $\hat{S}$ be the sample covariance matrix whitened with $\hat{R}$ as in (8), assuming that the normalization $\xi_N = \text{tr} C_N / N$ is lower bounded, then thanks to (13) the convergence (9) still holds for $\hat{S}$. As a consequence, we can apply this result to the signal detection problem.

We leave $\hat{R}_T^b$ behind because in some LRD cases, the convergence $\| R_T^{-1/2} \hat{R}_T^b R_T^{-1/2} - \xi I \| \to 0$ does not hold for any $\xi > 0$ in any reasonable sense. See Remark 2.3.
Main contributions. The contributions of this paper are two-fold. First, we consider not only the complex Gaussian distributions, but the rows of \( Z \) can be some more general spherically symmetric distributions. This setting is more widely used in e.g. robust statistics, signal processing, etc. This idea comes from [18] where the authors considered a very similar model but with short memory covariance matrix \( R_T \). Moreover their proof is nearly the same as [20] and is only applicable to compound Gaussian distributions.

Second, we justify the whitening procedure in the context of LRD processes. Because \( R_T \) is no longer bounded in spectral norm (see Lemma 4.3), the global bounds in Lemma 9 of [20] are no longer enough. Precisely, let \( \hat{\Upsilon}_T(\theta) \) and \( \hat{\Upsilon}_T(\theta) \) be defined in (30), then the key problem is to estimate the probability bounds for deviations of the relative errors
\[
\frac{|\hat{\Upsilon}_T(\theta) - \Upsilon_T(\theta)|}{f(\theta)}
\]
for almost every \( \theta \in [-\pi, \pi] \). This leads to the estimation of \( \text{tr} Q_T^2(\theta)/f^2(\theta) \) for almost every \( \theta \) where \( Q_T(\theta) \) is defined in (34), which is one of the main difficulties since \( R_T \) is no longer bounded and the global bounds in Lemma 9 of [20] only gives the trivial bound (39). Using harmonic analysis techniques we managed to prove that \( \text{tr} Q_T^2(\theta)/f^2(\theta) \) is at most \( O(\log^2 T) \) uniformly in \( \theta \) if \( f \) is in the form (10). The discretization method used in [20] also needs adaption because of the presence of the denominator \( f(\theta) \) in (15).

Notations. Matrices are denoted by bold capital characters, row or column vectors are denoted by bold characters. For \( x \in \mathbb{R}, \delta_x \) denotes the Dirac measure at \( x \). For a Hermitian \( N \times N \) matrix \( S \), its eigenvalues are denoted as \( \lambda_1(S) \geq \cdots \geq \lambda_N(S) \), and \( \mu^S := N^{-1} \sum_{i=2}^N \delta_{\lambda_i(S)} \) denotes the ESD of \( S \). The largest and smallest eigenvalues of \( S \) can also be noted as \( \lambda_{\max}(S) \) and \( \lambda_{\min}(S) \) respectively. For a matrix \( A \), \( A_{i,j} \) stands for the element at the \( i \)th row and \( j \)th column, \( ||A|| \) stands for the operator norm, which coincides with the spectral norm when \( A \) is Hermitian. \( A^* \) stands for the conjugate transpose of \( A \). For a square matrix \( A \), the spectrum of \( A \) is denoted by \( \text{Spec}(A) \). For a function \( f \), \( ||f||_1 \) stands for its \( L^1 \) norm, and \( ||f||_{\infty} \) stands for the \( \infty \)-norm, which is defined as \( \text{ess sup}_x |f(x)| \). For \( x \) in the definition domain of \( f \) and \( \delta > 0 \), we define a local \( \infty \)-norm \( || \cdot ||_{(x, \delta)} \) as
\[
||f||_{(x, \delta)} := \text{ess sup}_{t \in (x-\delta, x+\delta)} \{ |f(t)| \}.
\]
If \( a, b \) are two elements of a Hilbert space, we denote the inner product of \( a, b \) as \( \langle a, b \rangle \). The symbol \( K \) denotes a constant which may take different value from one place to another. If there are more than one constant in one expression, we will denote them by \( K_1, K_2, \ldots \). We say that a constant is absolute if it does not depend on anything. 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 \( A_n \leq (\gtrsim) K B_n \) 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 \( n \).

Organization. This paper is organized as follows. In Section 2, we state the main results. The most general result is Theorem 2.1 and Corollary 2.2, while Proposition 2.3 gives a negative answer to the question whether the biased estimator \( \hat{R}_T^b \) can be used in the whitening procedure when the process is LRD. In Section 3 we discuss an application to the complex Gaussian signal plus noise model. In Section 4 we prove the theorems and propositions. Theorem 2.1 is proved in Section 4.1, and Proposition 2.3 is proved in Section 4.2.

2. The model and main results

The model and assumptions. For two positive integers \( N, T \), let
\[
X = C_1^{1/2} Z R_T^{1/2}
\]
be a random matrix where \( Z \) is a \( N \times T \) random matrix whose rows are are denoted by \( z_1, \ldots, z_N \), \( C_N \) is a \( N \times N \) nonnegative definite Hermitian matrix, and \( R_T = (r_{i-j})_{i,j=1}^T \) is a Toeplitz matrix.

We will assume that the rows \( z_n \) are spherically symmetric in \( \mathbb{R}^T \) or \( \mathbb{C}^T \). That is, for any orthonormal or unitary matrix \( U_T \), we have \( z_n \overset{d}{=} U_T \). It is well known that \( z_n \) has the stochastic representation \( z_n \overset{d}{=} \sqrt{\nu_n} u_n \), where \( \nu_n \) is a positive random variable, \( u_n \) is uniformly distributed
on the sphere of $\mathbb{R}^T$ or $\mathbb{C}^T$ of radius $\sqrt{T}$, and $\nu_n$, $u_n$ are independent. An important (proper) subclass of spherically symmetric distributions is the multivariate compound Gaussian, which can be statistically represented by $z_n = \sqrt{\nu_n}g_n$ where $g_n$ is the $T$-dimensional standard real or complex Gaussian vector, independent of $\nu_n$. In our model and results, one can put the textures $\nu_n$ into the matrix $C_N$ by writing

$$C_N = \text{diag}(\sqrt{\nu_1}, \ldots, \sqrt{\nu_N})\hat{C}_N \text{diag}(\sqrt{\nu_1}, \ldots, \sqrt{\nu_N})$$

where $\hat{C}_N$ is a deterministic hermitian matrix. Thus in the main theorem 2.1 we only need to assume that the rows $z_n$ are i.i.d Gaussian or spherical uniform vectors, and if $\nu_n$ are random, the result should be interpreted as a conditional large deviation bound.

Although the compound Gaussian is a subclass of spherically symmetric distributions, we will still state the theorem separately under these two conditions. Not only because the proof is a generalization of the Gaussian case, but also provide an easier way of verifying the conditions in the compound Gaussian case.

Denote $\xi_N := N^{-1} \text{tr} C_N$. Assume that the following conditions are satisfied:

**A1** The row vectors $z_i$ are i.i.d standard real or complex Gaussian vectors, or i.i.d random vectors following the uniform distribution on the centered sphere of radius $\sqrt{T}$ in $\mathbb{R}^T$ or $\mathbb{C}^T$.

**A2** There exist constants $C > 0$ and $\kappa > 0$ such that

$$\frac{1}{N} \text{tr} C_N^2 \leq C, \quad \|C_N\| \leq \kappa \log T.$$

**A3** The matrix $C_N$ is diagonal, i.e. $C_N = \text{diag}(c_1, \ldots, c_N)$ with $c_i \geq 0$.

**A4** The Toeplitz matrices $R_T = (r_{i-j})_{i,j=1}^T$ have a positive spectral density $f \in L^1(-\pi, \pi)$ which is bounded in any set of the form $[-\pi, \pi] \setminus (-\delta, \delta)$ with $\delta > 0$.

**A5** The spectral density $f$ is lower bounded:

$$\text{ess inf}_{\theta \in (0,2\pi)} f(\theta) > 0.$$

**A6** The spectral density $f$ is even and has the following asymptotical behavior near 0:

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

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

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

$$r_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx.$$

If $f$ is real, then $R_T$ is Hermitian; if $f$ is positive, then $R_T$ is positive definite; if $f$ is positive and even, then $R_T$ is real symmetric and positive definite. We will consider $f$ as a $2\pi$-periodic real function so that $f(x)$ is well defined by periodicity for all real $x$. Note that the assumption **A5** ensures that the smallest eigenvalue of $R_T$ is lower bounded, thus $R_T$ is invertible for all $T$ with $R_T^{-1}$ bounded in spectral norm.

**Remark 2.1.** We require in **A6** that the $f$ is singular at 0. But the special location of singularity is not essential. For any $\theta \in \mathbb{R}$, let

$$D_T(\theta) := \text{diag}(1, e^{i\theta}, \ldots, e^{i(T-1)\theta}).$$

It is easily seen that if $f$ is the spectral density of $R_T$, then $f(\cdot + \theta)$ is the spectral density of $D_T(\theta)R_TD_T(\theta)$. As $D_T(\theta)$ is unitary, the spectrum of Toeplitz matrices are invariant under the translation of the spectral density, and our results are not affected due to the fact that the complex Gaussian vectors are unitary invariant. So the example (12) also enters into our consideration.

If $R_T$ satisfies **A4**, **A6** and is the autocovariance matrix of a stationary process, then the process is long range dependent by Definition 2.1.5 (Condition IV) of [14]. We will prove that in this case, $\|R_T\| \asymp T^n L(T)$ in Lemma 4.3. Note that a similar estimation of $\|R_T\|$ is also available in [13] for Toeplitz matrices $R_T$ satisfying Definition 2.1.5 (Condition II) of [14].
Let
\[ S := \frac{1}{N} XX^* = \frac{1}{N} R_T^{1/2} Z^* C_N Z R_T^{1/2} =: (s_{i,j})_{i,j=1}^{T}, \]
and
\[ \tilde{r}_k := \frac{1}{T - |k|} \sum_i s_{i+k,i} I_{\{1 \leq i + k \leq T, 1 \leq i \leq T\}}. \]
We define
\[ (18) \quad \hat{R}_T := (\tilde{r}_{i-j})_{1 \leq i,j \leq T}. \]

Main results.

**Theorem 2.1.** Under A1- A6, there exist K > 0, an integer β ≥ 0, such that for any fixed \( x \in (0, C/(K\kappa)) \), and large enough \( N, T \), we have
\[ \mathbb{P} \left( \left\| R_T^{-1/2} \hat{R}_T R_T^{-1/2} - \xi_N I \right\| > x \right) \leq 2T^\beta \exp \left( -\frac{N x^2}{K C \log^2 T} \right). \]

If \( x \) depends also on \( N, T \) and \( x = o(1) \), we need that there exists \( \gamma > 0 \) such that \( x \gtrsim T^{-\gamma} \), and the \( \beta \) in (19) satisfies \( \beta > 2 + \alpha + \gamma \).

Using this theorem, we can prove the almost sure convergence theorem in the context of spherically symmetric distributions. Assume that

**A7** The rows of \( Z \) are i.i.d. and have the stochastic representation
\[ z_n = \xi_n \tilde{z}_n \]
where \( \nu_n \) are i.i.d positive sub-exponential variables with \( \mathbb{E} \nu_n = 1 \), and \( \tilde{z}_n \) satisfies A1.

**A8** The matrix \( C_N \) is bounded in spectral norm, and is real if \( \tilde{z}_n \) is real Gaussian; \( C_N \) is diagonal if \( \tilde{z}_n \) are non-Gaussian.

**Corollary 2.2.** Let \( X \) be defined as in (17) with \( R_T \) satisfying A4, A5, A6, \( Z \) satisfying A7, and \( C_N \) satisfying A8. Then almost surely, as \( N, T \to \infty \) with \( N/T \) stays in a compact subset of \((0, \infty)\),
\[ (20) \quad \| R_T^{-1/2} \hat{R}_T R_T^{-1/2} - \xi_N I \| \overset{a.s.}{\to} 0, \]
where \( \xi_N = N^{-1} \text{tr} C_N \).

**Remark 2.2.** In the case of spherically symmetric rows, it is also possible to establish a concentration inequality. Note that if \( \nu_n \) are i.i.d. sub-exponential variables with \( \mathbb{E} \nu_n = 1 \), let \( V_N = \text{diag}(\nu_1, \ldots, \nu_n) \), then for some \( C > 1, \kappa > 0 \), almost surely
\[ (21) \quad \frac{\text{tr}(C_N V_N)^2}{N} \leq \frac{\| C_N \|^2 \text{tr} V_N^2}{N} \leq C; \quad \max_n \nu_n \leq \kappa \log N. \]
Thus applying Theorem 2.1, there is \( K > 0 \) such that, for any \( x \in (0, C/(K\kappa)) \), one has
\[ \mathbb{P} \left( \left\| R_T^{-1/2} \hat{R}_T R_T^{-1/2} - \xi N \right\| > x \right) \leq 2T^\beta \exp \left( -\frac{N x^2}{K C \log^2 T} \right) + \mathbb{P}(\text{(21) is not satisfied}) \]
where \( \xi_N = N^{-1} \text{tr}(C_N V_N) \) which converges almost surely to \( \xi_N = N^{-1} \text{tr} C_N \).

Using the above result, assuming moreover that \( \xi_N \) is lower bounded, we can whiten the correlation of the columns of \( X \) by setting
\[ (22) \quad \hat{S} = \frac{1}{T} X \hat{R}_T^{-1} X^*. \]
Comparing time-domain whitened \( \hat{S} \) to the sample covariance matrix with true uncorrelated data columns
\[ (23) \quad S^w := \frac{1}{T} C_N^{1/2} Z Z^* C_N^{1/2}, \]
we have
\[ \| \hat{S} - \xi_N^{-1} S^w \| \overset{a.s.}{\to} 0. \]
Remark 2.3. In the short memory case, the biased estimator $\hat{R}_T^b$ defined in (5) and (6) sometimes performs better than the unbiased estimator $R_T$, because the elements near the top-right corner and bottom-left corner of $R_T$ are less accurate. By tapering these corner elements we can reduce the inaccuracy without modifying essentially the spectral properties. See [20] and [5]. However in the long memory case, the bias $\hat{E}R_T^{1/2} - R_T$, or the normalized bias $R_T^{-1/2}\hat{E}R_T^{1/2} - \xi_N I$ are no longer negligible.

Let

$$r_n = \frac{1}{(1 + |n|)^{1-a}}, \quad \text{and} \quad R_T = (r_{i-j})_{i,j=1}^T,$$

where $a \in (0, 1)$. Then from Proposition 2.2.14 of [14], $R_T$ satisfies $A_4, A_6$. From Theorem 1.5, Chapter V of [23], $R_T$ satisfy also $A_5$. Indeed, if the diagonal entry $r_0$ is large enough such that $(r_n)_{n \geq 0}$ is convex, the spectral density of $R_T$ is nonnegative definite. The minimal value of such $r_0$ is $2^n - 3^{a-1} < 1$ for $0 < a < 1$. Thus when we take $r_0 = 1$, the spectral density $f$ is larger than $1 - 2^n + 3^{a-1} > 0$.

Suppose that $C_N = I_N$, then

$$R_T^b := \hat{E}R_T = \left( 1 - \frac{|i-j|}{T} \right)^T \frac{1}{(1 + |i-j|)^{1-a}} i,j=1^T.$$

In Proposition 2.3 we show that $\| R_T^{-1/2} R_T^b R_T^{-1/2} - I \| \ne 0$ for any $\gamma > 0$. In the proof of this proposition, we note that a necessary condition of the convergence $\| R_T^{-1/2} R_T^b R_T^{-1/2} - I \| \ne 0$ is that $\lambda_{\max}(R_T^b) \sim \gamma \lambda_{\max}(R_T)$, $\lambda_{\min}(R_T^b) \sim \gamma \lambda_{\min}(R_T)$. However, in the LRD cases, the behavior of $\lambda_{\min}(R_T^b)$ is mainly determined by the entries near diagonal, while the asymptotic behavior of $\lambda_{\max}(R_T^b)$ is mainly determined by the $\varepsilon T$ entries at the top-right (bottom-left) corner (with any $0 < \varepsilon < 1$). Modifying these elements impacts on the asymptotics of the largest eigenvalue. From this point of view, most tapering methods can face the same inconsistency.

Proposition 2.3. Let $R_T$ and $R_T^b$ be defined as in Remark 2.3. Then for any $\xi > 0$,

$$\| R_T^{-1/2} R_T^b R_T^{-1/2} - I \| \ne 0$$

as $T \to \infty$.

So in this paper we focus on the unbiased estimator $R_T$.

3. Applications and Simulation Results

We take the signal-plus-noise model (1) with LRD on time-domain. Then we have

$$X = C_N^{1/2} Z R_T^{1/2}$$

with

$$C_N = A \, \text{Cov}(s) A^* + \sigma^2 I_N$$

and $Z$ is a $N \times T$ matrix having i.i.d standard Gaussian entries. Suppose that $A \, \text{Cov}(s) A^*$ has eigenvalues $\alpha_1^2 > \cdots > \alpha_p^2$ with multiplicities $m_1, \ldots, m_p$ and we set $m = m_1 + \cdots + m_p$. The ratios

$$\beta_i := \frac{\alpha_i^2}{\sigma^2}, \quad \text{for } i = 1, \ldots, p$$

can be considered as signal-to-noise ratios (SNR) of signals $\alpha_1^2, \ldots, \alpha_p^2$. When we need to infer $\alpha_1^2$ or $\beta_i$ from $S$, because the Gaussian vectors are unitary invariant, we can diagonalize $C_N$ as

$$C_N = \begin{pmatrix} A_m & 0 \\ 0 & \sigma^2 I_{N-m} \end{pmatrix},$$

where $A_m$ has eigenvalues $\alpha_i + \sigma^2$. By [21, Corollary 11.4, Theorem 11.11] and Weyl’s inequality $|\lambda_i(A) - \lambda_i(B)| \leq \| A - B \|$ for any Hermitian matrices $A, B$, the following corollary is immediate. Note that the spectrum of the normalized matrix $\xi_N^{-1} S^w$ depends on $\beta_i$’s, but not on $\sigma^2$. In the
sequel of this section we can suppose that $\sigma^2 = 1$ without loss of generality. Then the normalization $\xi_N$ is asymptotically

$$\xi_N = 1 + \frac{\sum_{i=1}^pm_i\alpha_i}{N} = 1 + O(1/N).$$

**Corollary 3.1.** Let $\tilde{S}$ be defined as in (22) with $\mathbf{Z}$ having i.i.d standard complex Gaussian entries, and $\mathbf{R}_T$ satisfying $A_4$, $A_5$, $A_6$. Suppose that $\mathbf{C}_N$ is in the form (24) where $\mathbf{A}\operatorname{Cov}(s)\mathbf{A}^*$ has eigenvalues $\alpha_1^2 > \cdots > \alpha_p^2$ with multiplicities $m_1, \ldots, m_p$. Suppose also that $N, T \to \infty$ with $N/T \to c \in (0, \infty)$. Then almost surely the ESD of $\tilde{S}$ converges to the MP law (26)

$$\mathbb{P}_{MP}(d\lambda) := (1 - e^{-1})\delta_0(d\lambda)\mathbb{I}_{\{c>1\}} + \frac{\sqrt{(\lambda^+ - \lambda)(\lambda - \lambda^-)}}{2\pi c\lambda} \mathbb{I}_{\lambda\in(\lambda^-,\lambda^+)\mathbb{d}\lambda}, \quad \lambda^\pm = (1 \pm \sqrt{c})^2.$$

Let $\beta_i = \alpha_i^2/\sigma^2$, then for any $i = 1, \ldots, p$ and $j = 1, \ldots, m_i$ (making convention that $m_0 = 0$),

$$\lambda_{m_0+\cdots+m_{i-1}+j}(\tilde{S}) \xrightarrow{\mathbb{P}} \psi(\beta_i) := \begin{cases} \beta_i + \frac{c}{\pi} + c + 1 & \beta_i > \sqrt{c} \\ (1 + \sqrt{c})^2 & \beta_i \leq \sqrt{c}. \end{cases}$$

Moreover, almost surely

$$|\lambda_{m_0+\cdots+m_{i-1}+j}(\tilde{S}) - \psi(\beta_i)| = O\left(\frac{\log^{3/2}N}{\sqrt{N}}\right).$$

Suppose that the largest SNR $\beta_1 > \sqrt{c}$ is large enough. Consider the following test in the model (24),

$$\begin{cases} H_0 : \quad \mathbf{C}_N = \sigma^2 \mathbf{I}_N, \\ H_1 : \quad \mathbf{C}_N \neq \sigma^2 \mathbf{I}_N. \end{cases}$$

Then according to Corollary 3.1, there exists $\varepsilon \gg \frac{\log^{3/2}N}{\sqrt{N}}$ such that

$$\lambda_1(\tilde{S}) \xrightarrow{H_0 \to H_1} (1 + \sqrt{c})^2 + \varepsilon.$$

Now suppose that $H_1$ is adopted and that the minimum SNR threshold $s > \sqrt{c}$ is known, i.e.

$$\beta_p \geq s > \sqrt{c}.$$

We can estimate $\beta_i$ with $m_i$ for $i = 1, \ldots, p$ using the following algorithm.

Let $d_N = o(1)$, $d_N \gg \log^{3/2}N/\sqrt{N}$ . In order to distinguish the spiked eigenvalues and the eigenvalues at the border, we suppose that $\psi(s) - (1 + \sqrt{c})^2 > 2d_N$. For $i = 1, 2, \ldots$, we set

1. $\hat{m}_0 = 0$;
2. If $\hat{m}_{i-1}$ has been estimated, let
   $$\hat{m}_i = \min\{j : \lambda_{m_0+\cdots+m_{i-1}+j}(\tilde{S}) - \lambda_{m_0+\cdots+m_{i-1}+j}(\tilde{S}) > d_N\}.$$
3. Repeat Step 2, until
   $$\lambda_{\hat{m}_1+\cdots+\hat{m}_i+1}(\tilde{S}) < \psi(s) - d_N.$$
   We set $p$ the smallest $i$ satisfying the above property.
4. We set
   $$\hat{m} = \hat{m}_1 + \cdots + \hat{m}_i.$$
5. Let
   $$\hat{\beta}_i = \frac{1}{\hat{m}_i} \sum_{j=1}^{\hat{m}_i} \psi^{-1}\left(\lambda_{m_0+\cdots+m_{i-1}+j}(\tilde{S})\right).$$
By Corollary 3.1, the above estimators are all consistent as \( N, T \to \infty \) with \( N/T \to c \). Note that the theoretical convergence rate \( \frac{\log^{3/2} N}{\sqrt{N}} \) is larger than the standard fluctuation scales \( (N^{-1/2} \text{ for spiked eigenvalues and } N^{-2/3} \text{ for the eigenvalues at the right border of the bulk of MP law}) \), the fluctuations of spiked and non-spiked border eigenvalues of \( \hat{S} \) remain unknown. Nevertheless, from the following numeric simulation we get very encouraging results.

We take \( N = 500, T = 833, c = 0.6 \), and \( C_N = \text{diag}(10, 10, 6, 4, 4, 1, \ldots) \).

Let

\[
R_T = \left( \frac{1}{(1 + |i-j|)^{0.3}} \right)_{i,j=1}^T.
\]

Then \( R_T \) satisfies A4, A6 A5. The argument can be found in Remark 2.3. We make 1000 independent realizations, and note the means and standard deviations (SD) of the largest eigenvalues of \( \tilde{S} \) and \( S^w \). The result is listed in Table 3, first column, compared with the means and SD’s of largest eigenvalues of \( S^w \) in the second column. We see that the performance of whitened model \( \tilde{S} \) is at least as good as the standard sample covariance model \( S^w \).

### Table 1. Means and standard deviations of largest eigenvalues

| \( i \) | \( \lambda_i(\tilde{S}) \) mean | SD | \( \lambda_i(S^w) \) mean | SD | \( \xi_N \lambda_i(\tilde{S}) \) mean | SD | \( \lambda_i(C_N) \) mean | SD |
|---|---|---|---|---|---|---|---|
| 1 | 10.0898 | 0.2597 | 11.0825 | 0.3036 | 11.0150 | 0.3125 | 10 |
| 2 | 9.4060 | 0.2433 | 10.3077 | 0.2832 | 10.2684 | 0.2873 | 10 |
| 3 | 6.2183 | 0.1729 | 6.7249 | 0.1934 | 6.7884 | 0.2065 | 6 |
| 4 | 4.6653 | 0.0993 | 5.0111 | 0.1089 | 5.0929 | 0.1406 | 4 |
| 5 | 4.4293 | 0.0857 | 4.7527 | 0.0948 | 4.8353 | 0.1306 | 4 |
| 6 | 4.2051 | 0.0877 | 4.5098 | 0.0968 | 4.5905 | 0.1307 | 4 |
| 7 | 2.8849 | 0.0213 | 3.0679 | 0.0231 | 3.1492 | 0.0614 | 1 |
| 8 | 2.8383 | 0.0173 | 3.0182 | 0.0188 | 3.0984 | 0.0831 | 1 |
| 9 | 2.8019 | 0.0150 | 2.9792 | 0.0163 | 3.0587 | 0.0828 | 1 |

From the numeric simulations, the top eigenvalues of \( \tilde{S} \) are a little smaller than the corresponding eigenvalues of \( S^w \) due to ignoring the normalization \( \xi_N \), which makes the estimators \( \hat{\beta}_i \) tend to be smaller than the real values. It is sometimes more convenient to estimate \( \xi_N \) which is also unknown, especially when \( N \) is not very large but there are too many large spiked eigenvalues of \( C_N \). From the theoretical result, we have

\[
\xi_N \sim \frac{\lambda_i(S^w)}{\lambda_i(\tilde{S})}
\]

for any fixed \( i \). We also know that the largest border eigenvalue of \( S^w \) converges to \((1 + \sqrt{c})^2 \). Thus an appropriate estimator of \( \xi_N \) might be

\[
\hat{\xi}_N = \frac{(1 + \sqrt{c})^2}{\lambda_{n+1}(\tilde{S})}.
\]

With the estimated \( \hat{\xi}_N \), the corrected estimation results are listed in the third column of Table 3. We can see that in this example the corrected spiked eigenvalues are closer to the corresponding eigenvalues of \( S^w \).

**Conjectures and open questions.** The fluctuations of spiked and border eigenvalues of \( \tilde{S} \) are still unknown. From the result of numeric simulations, we conjecture that the fluctuations of spiked and border eigenvalues of \( \tilde{S} \) are the same as those of \( S^w \).

### 4. Proof of main results

#### 4.1. Proof of Theorem 2.1
4.1.1. Some preliminary works. The idea of the proof is to estimate the range of eigenvalues of the Hermitian matrix $R_T^{-1/2} R_T R_T^{-1/2} - \xi_N I$. Note that under the assumption A2, $\xi_N$ is also bounded.

Note that the matrix $R_T^{-1/2} R_T R_T^{-1/2}$ has the same eigenvalues as $\hat{R}_T R_T^{-1}$, we only need to control the probability that some eigenvalues of $R_T R_T^{-1}$ are outside a neighborhood of $\xi_N$. The following lemma connects the spectrum of $R_T R_T^{-1}$ with the spectral densities of the two Toeplitz matrices. It was first proved in [8] and extended to integrable spectral densities in [15], see Theorem 2.1 of [15].

Lemma 4.1. Let $f_1, f_2 \in L^1((0, 2\pi])$ be two nonnegative integrable functions not identically zero. Let $R_{1,T}, R_{2,T}$ be two $T \times T$ Toeplitz matrices whose spectral densities are $f_1$ and $f_2$ respectively. Then for any $T \geq 1$,

$$\operatorname{Spec}(R_{1,T} R_{2,T}^{-1}) \subset \left[ \essinf_{\theta \in [0, 2\pi]} \frac{f_1(\theta)}{f_2(\theta)}, \esssup_{\theta \in [0, 2\pi]} \frac{f_1(\theta)}{f_2(\theta)} \right].$$

We note that the Toeplitz matrix $\hat{R}_T$ is random and depends on $N, T$, it does not have a fixed spectral density. To tackle this problem, for each $N$ and $T$, we define

$$\hat{f}_T(\theta) := \xi_N f(\theta) + \sum_{n=-T+1}^{T-1} (\hat{r}_n - \xi_N r_n) e^{i n \theta}$$

where $f$ is the spectral density of $R_T$. Note that the Fourier coefficients of $\hat{f}_T$ are $\hat{r}_t$ for $-T + 1 \leq t \leq T - 1$, thus for this particular $N$ and $T$, $\hat{f}_T$ is the spectral density of $R_T$, and by Lemma 4.1, the eigenvalues of $R_T R_T^{-1}$ are in the interval

$$\left[ \essinf_{\theta \in [0, 2\pi]} \frac{\hat{f}_T(\theta)}{f(\theta)}, \esssup_{\theta \in [0, 2\pi]} \frac{\hat{f}_T(\theta)}{f(\theta)} \right].$$

Thus in order to prove the theorem we only need to estimate

$$P \left( \esssup_{\theta \in [0, 2\pi]} \left| \frac{\hat{f}_T(\theta)}{f(\theta)} - \xi_N \right| > x \right)$$

for any appropriate $x > 0$. Let

$$\Upsilon_T(\theta) := \sum_{n=-T+1}^{T-1} r_n e^{i n \theta}, \quad \hat{\Upsilon}_T(\theta) := \sum_{n=-T+1}^{T-1} \hat{r}_n e^{i n \theta}.$$

Then $\hat{f}_T(\theta) - \xi_N f_T(\theta) = \hat{\Upsilon}_T(\theta) - \xi_N \Upsilon_T(\theta)$. Recall that we have $E \hat{\Upsilon}_T(\theta) = \xi_N \Upsilon_T(\theta)$ for any $\theta \in [0, 2\pi]$. Then the probability (29) becomes

$$P \left( \esssup_{\theta \in [0, 2\pi]} \left| \frac{\hat{\Upsilon}_T(\theta) - E \hat{\Upsilon}_T(\theta)}{f(\theta)} \right| > x \right).$$

This can be considered as the probability of large relative error of the estimation $\hat{\Upsilon}_T(\theta)$ with respect to $f(\theta)$. We will use a similar discretization strategy as [20]. Let

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

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

$$P \left( \left| \hat{\Upsilon}_T(\theta_j) - E \hat{\Upsilon}_T(\theta_j) \right| > x f(\theta_j) \right)$$

for each $\theta_j$. 

4.1.2. Relative error bound for individual $\theta$. Denote

$$D_T(\theta) := \text{diag}(1, e^{i\theta}, \ldots, e^{i(T-1)\theta}), \quad B_T := \left(\frac{1}{T - |i - j|}\right)_{i,j=0}^{T-1},$$

and

$$Q_T(\theta) := R_T^{1/2}D_T(\theta)B_T D_T(\theta)R_T^{1/2}.$$  

Then from Lemma 7 and 8 and (9) of [20], under A3, we have

$$\hat{\Upsilon}_T(\theta) = \frac{1}{N} \text{tr} C_N^{1/2} Z_T Q_T(\theta) Z_T^* C_N^{1/2} = \frac{1}{N} \sum_{n=1}^{N} c_n z_n Q_T(\theta) z_n^*.$$  

We assume now that $z_n$ are unitary invariant, that is, $z_n$ are complex Gaussian or uniformly distributed on the complex sphere. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_T$ be eigenvalues of $Q_T(\theta)$ (attention that $Q_T(\theta)$ need not be nonnegative definite). By the unitary invariance of the rows $z_n$, we have

$$\hat{\Upsilon}_T(\theta) - E\hat{\Upsilon}_T(\theta) \leq \frac{1}{N} \sum_{n=1}^{N} c_n \sum_{t=1}^{T} \sigma_t (|Z_{n,t}|^2 - 1).$$  

Now we discuss separately the complex Gaussian case and the complex spherical case. In the Gaussian case, the entries of $z_n$ are i.i.d. standard Gaussian random variables. Then (36) is a sum of $NT$ centered i.i.d. random variables. We write

$$\mathbb{P}\left(\frac{1}{N} \sum_{n,t} c_n \sigma_t (|Z_{n,t}|^2 - 1) > x f(\theta)\right) = \mathbb{P}\left(\left|\sum_{n,t} c_n \sigma_t (|Z_{n,t}|^2 - 1)\right| > \frac{Nx f(\theta)}{\sqrt{\sum_t \sigma_t^2}}\right).$$

The Proposition 4.6 below provides an upper bound of $\text{tr} Q_T^2(\theta)$ in terms of $f(\theta)$. Then there exists a constant $K > 0$ such that

$$\frac{Nx f(\theta)}{\sqrt{\sum_t \sigma_t^2}} \geq K N x \log T.\quad (37)$$

Writing $\sigma'_t = \sigma_t / \sqrt{\sum_t \sigma_t^2}$, then

$$\mathbb{P}\left(\left|\sum_{n,t} c_n \sigma'_t (|Z_{n,t}|^2 - 1)\right| > \frac{Nx f(\theta)}{\sqrt{\sum_t \sigma'_t^2}}\right) \leq \mathbb{P}\left(\left|\sum_{n,t} c_n \sigma'_t (|Z_{n,t}|^2 - 1)\right| > \frac{K N x}{\log T}\right).$$

Then we only need to estimate the RHS of (37) with $\sum_t (\sigma'_t)^2 = 1$. The remaining proof for Gaussian case is similar to the proof of Lemma 11 in [20]. Note that for a random variable $X$, one has $\mathbb{P}(|X| > x) = \mathbb{P}(X > x) + \mathbb{P}(-X > x)$, we only need to prove one direction, and the other direction is similar. By Chebyshev’s inequality, we have

$$\mathbb{P}\left(\sum_{n,t} c_n \sigma'_t (|Z_{n,t}|^2 - 1) > \frac{K N x}{\log T}\right) \leq \exp\left(-\tau \frac{K N x}{\log T} + \sum_{n,t} \phi(\tau c_n \sigma'_t)\right)$$

where $\phi$ is the cumulant generating function of $|Z_{n,t}|^2 - 1$:

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

and $\tau$ is any positive number such that $\phi(\tau c_n \sigma'_t)$ are well defined and finite. By the Taylor’s expansion formula $\log(1 - z) = -z^2/2 + z^3/3 - \cdots$, choosing an arbitrary $\varepsilon \in (0, 1)$, then there exists $A_\varepsilon > 0$ such that for any $|z| \leq \varepsilon$, we have

$$|\phi(z)| = |z^2|^{1/2} - z/3 + \cdots | \leq A_\varepsilon |z|^2.$$

Let $\tau$ be such that $|\tau c_n \sigma'_t| \leq \varepsilon$ for any $n, t$, then

$$\mathbb{P}\left(\sum_{n,t} c_n \sigma'_t (|Z_{n,t}|^2 - 1) > \frac{K N x}{\log T}\right) \leq \exp\left(-\tau \frac{K N x}{\log T} + A_\varepsilon \tau^2 \sum_n c_n^2\right).$$
Noting that $\sum_n c_n^2 \leq CN$ by $A2$, we then have
\[
P\left( \sum_{n,t} c_n \sigma_n^t |Z_{n,t}|^2 - 1 > \frac{KNx}{\log T} \right) \leq \exp \left( -\tau \frac{KNx}{\log T} + CA \gamma^2 \right).
\]

Let
\[
\tau = \frac{Kx}{2CA \gamma \log T},
\]
then we have
\[
P\left( \sum_{n,t} c_n \sigma_n^t |Z_{n,t}|^2 - 1 > \frac{KNx}{\log T} \right) \leq \exp \left( -\frac{K^2 N x^2}{4CA \gamma \log^2 T} \right)
\]
whenever
\[
\max_n |\tau c_n| \leq \frac{K \gamma x}{2CA \gamma} \leq \varepsilon.
\]

When $z_n$ follows the uniform distribution on the sphere $\{ z \in C^T : \|z\| = \sqrt{T} \}$. Then $\|z_n\|^2 = T$. We have
\[
z_n Q_T(\theta) z_n^* = \frac{\|z_n\|^2}{T} \text{tr} Q_T(\theta) \equiv \sum_{t=1}^T (\sigma_t - \text{tr} Q_T(\theta)/T)|Z_{n,t}|^2.
\]
From Proposition 4.6, we have
\[
\sum_{t=1}^T \left( \sigma_t - \frac{\text{tr} Q_T(\theta)/T}{T} \right)^2 \leq \text{tr} Q_T^2(\theta) \leq f^2(\theta) \gamma^2 T.
\]
Write $\sigma'_t = (\sigma_t - \text{tr} Q_T(\theta)/T)/\sqrt{\sum_t (\sigma_t - \text{tr} Q_T(\theta)/T)^2}$. Then there exists a constant $K > 0$ such that
\[
P\left( \left| \frac{1}{N} \sum_{n,t} c_n (z_n Q_T(\theta) z_n^* - \text{tr} Q_T(\theta)) \right| > xf(\theta) \right) \leq \frac{1}{N} \sum_{n,t} c_n \sigma'_t |Z_{n,t}|^2 \leq \frac{KNx}{\log T}.
\]
In this case we have $Ez_n = 0$, $\text{Cov} z_n = I_T$. But the entries of $z_n$ are not independent. We will have to estimate the cumulant generating function of $\sum_t \sigma'_t |Z_{n,t}|^2$:
\[
\Phi_T(z) := \log Ee^{zT} \sum_t \sigma'_t |Z_{n,t}|^2 = \frac{T}{2(T+1)} z^2 + \ldots
\]
This is a function depending on $T$. We want to prove that there exists uniform constants $\varepsilon > 0$ and $A > 0$ such that $|\Phi_T(z)| \leq A|z|^2$ for any $T$ and any $|z| \leq \varepsilon$. By the Taylor’s expansion,
\[
(38) \quad E e^{z T} \sum_t \sigma'_t |Z_{n,t}|^2 = \sum_{k=0}^{\infty} \frac{z^k}{k!} E \left( \sum_t \sigma'_t |Z_{n,t}|^2 \right)^k.
\]
Let $g \in C^T$ be a standard complex Gaussian vector independent of $z_n$. Then by the properties of spherical symmetric (see for example [7]), $g \leq \|g\| z_n/\sqrt{T}$. Then
\[
E \exp \left( z \frac{\|g\|^2}{T} \sum_t \sigma'_t |Z_{n,t}|^2 \right) = \sum_{k=0}^{\infty} \frac{z^k}{k!} E \left( \sum_t \sigma'_t |Z_{n,t}|^2 \right)^k = \prod_{t=1}^T \frac{1}{1 - z \sigma'_t} = \exp \left( \sum_{t=1}^T \log(1 - z \sigma'_t) \right).
\]
Note that $\sum_t \sigma'_t = 0$, $\sum_t (\sigma'_t)^2 = 1$ and $|\sigma'_t| \leq 1$. From the proof of Gaussian case, for an arbitrary $\varepsilon \in (0, 1)$, there exists $A_{\varepsilon}$ such that $|\log(1 - z \sigma'_t) + z \sigma'_t| \leq A_{\varepsilon} |z|^2 (\sigma'_t)^2$ for any $|z| \leq \varepsilon$. Thus for these $z$ we have
\[
\exp \left( \sum_{t=1}^T \log(1 - z \sigma'_t) \right) \leq \exp(A_{\varepsilon} |z|^2).
\]
Then from Cauchy’s integration formula, for any \( k \geq 0 \), we have
\[
\left| \frac{E\|g\|^{2k}}{k!T^k} - E\left(\sum_t \sigma'_t |Z_{n,t}|^2\right)^k \right| = \frac{1}{2\pi} \left| \int_{|z|=\varepsilon} \frac{1}{z^{k+1}} \prod_t \frac{1}{1 - z\sigma_t^2} \, dz \right| \leq \frac{e^{A_k\varepsilon^2}}{\varepsilon^k}.
\]

Note that
\[
\frac{T^k}{E\|g\|^{2k}} = \frac{T^k}{T(T+1) \cdots (T+k-1)} \leq 1,
\]
then for any \( |z| \leq \varepsilon/2 \), we have
\[
\left| \sum_{k=0}^{\infty} \frac{z^{k-2}}{k!} E\left(\sum_t \sigma'_t |Z_{n,t}|^2\right)^k \right| \leq 2e^{-2e^{A_k\varepsilon^2}}.
\]

Take this into (38), we get
\[
\left| Ee^{z\sum_t \sigma'_t |Z_{n,t}|^2} - 1 \right| \leq 2e^{-2e^{A_k\varepsilon^2}} |z|^2
\]
for any \( |z| \leq \varepsilon/2 \). Then by the Taylor’s expansion of \( \log(1 + z) \) again, as \( |z| \) is small enough, we have
\[
|\Phi_T(z)| = |\log Ee^{z\sum_t \sigma'_t |Z_{n,t}|^2}| \leq A|z|^2
\]
for some constant \( A \). The remaining proof for spherically uniform case is identical to the Gaussian case.

In the real case, the proof is similar since when \( C_N, Z, R_T \) are all real, one has
\[
\hat{Y}_T(\theta) = \Re(\hat{\Upsilon}(\theta)) = \frac{1}{N} \sum_{n=1}^{N} z_n \Re(Q_T(\theta)) z_n^*.
\]

Note also that
\[
\text{tr}(\Re(Q_T(\theta))^2) \leq \text{tr} Q_T^2(\theta)
\]
and
\[
\log Ee^{z|G|^2} = -\frac{1}{2} \log(1 - 2z)
\]
for standard real Gaussian variable \( G \) with \( |z| < 1/2 \). If \( g \in \mathbb{R}^T \) is a standard real Gaussian vector, we also have
\[
\frac{T^k}{E\|g\|^{2k}} = \frac{T^k}{T(T + 2) \cdots (T + 2k - 2)} \leq 1.
\]

We only need to replace the corresponding items with the above mentioned properties in the proof of complex case. We omit the details.

4.1.3. **Relative error bound for all \( \theta \) by discretization.** Let \( \beta \) be a positive integer to be determined afterwards. For \( k = 0, \ldots, T^\beta \), let
\[
\theta_k := \frac{2\pi k}{T^\beta}.
\]

For \( \theta \in [0, 2\pi) \), let \( \theta_j \) 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
\[
\frac{\hat{Y}_T(\theta) - Y_T(\theta)}{f(\theta)} \leq \frac{\hat{Y}_T(\theta) - \hat{Y}_T(\theta_j)}{f(\theta)} + \frac{\hat{Y}_T(\theta_j) - Y_T(\theta_j)}{f(\theta)} + \frac{Y_T(\theta) - Y_T(\theta_j)}{f(\theta)} =: \chi_1(\theta) + \chi_2(\theta) + \chi_3(\theta).
\]
From the proof of Lemma 10 in [20], and note that $f(\theta)$ is bounded away from 0, also note Lemma 4.3 for the bound of $\|R_T\|$, and $A2$ for the bound of $\|C_N\|$, we have

$$\sup_{\theta \in [0,2\pi]} \chi_1(\theta) \leq \sup_{\theta \in [0,2\pi]} \frac{1}{N} \|C_N\| \|Q_T(\theta) - Q_T(\theta_j)\| \|\theta - \theta_j\| \sum_{n,t} |Z_{n,t}|^2$$

$$\lesssim \sup_{\theta \in [0,2\pi]} \frac{1}{N} \|C_N\| \|R_T\| T \sqrt{\log T} \|\theta - \theta_j\| \sum_{n,t} |Z_{n,t}|^2$$

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

$$\leq T^{2+a-\beta} L(T)(\log T)^{3/2} \sup_{n,t} |Z_{n,t}|^2 / NT.$$ 

If $z_n$ are on the sphere of radius $\sqrt{T}$, we have $\sum_{n,t} |Z_{n,t}|^2 = 1$; if $z_n$ are standard complex normal, we have for any $y > 1$,

$$\mathbb{P}\left( \sum_{n,t} |Z_{n,t}|^2 / NT > y \right) \leq \exp(-NT(y-1 - \log y)).$$

For any $x > 0$ which is either fixed, or dependent on $N,T$ such that $x \geq T^{-\gamma}$, we have

$$\mathbb{P}\left( \sup_{\theta \in [0,2\pi]} \chi_1(\theta) > x \right) \leq \mathbb{P}\left( \sum_{n,t} |Z_{n,t}|^2 / NT > \frac{T^{\beta - 2 - a - \gamma}}{L(T)(\log T)^{3/2}} \right).$$

We take $\beta > 2 + a + \gamma$, let $\varepsilon = \frac{\beta - 2 - a - \gamma}{2}$, then as $T$ is large enough, we have $T^{\varepsilon} > L(T)(\log T)^{3/2}$ and $1 + \log(T^{\varepsilon}) < \frac{T}{2}$. Then

$$\mathbb{P}\left( \sup_{\theta \in [0,2\pi]} \chi_1(\theta) > x \right) \leq \exp(-NT^{1+\varepsilon}/2).$$

From the proof of Lemma 12 in [20], we have

$$\sup_{\theta \in [0,2\pi]} \chi_3(\theta) \lesssim T^2 \|\theta - \theta_j\| \|R_T\| \sqrt{\log T}$$

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

With the same $\beta > 2 + a + \gamma$, for any $x > 0$ which is either fixed, or dependent on $N,T$ such that $x \geq T^{-\gamma}$, as $T$ is large enough, we have

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

For $\chi_2(\theta)$, we note that

$$\chi_2(\theta) = \chi_3(\theta) \frac{f(\theta_j)}{f(\theta)}.$$ 

We prove that $\frac{f(\theta_j)}{f(\theta)}$ is (essentially) bounded for $\theta \in [0,2\pi]$ and $\theta_j$ defined as before. Because $f$ is supposed to be even and $2\pi$-periodic, we only need to consider $\theta \in (0,\pi)$. Note that by Lemma 4.2(c),

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

as $\theta_j \to 0^+$. Let $\delta > 0$ be such that

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

for $0 < \theta \leq \theta_j \leq \delta$. Then for any $\theta \in (0,2\pi)$, we have

$$\frac{f(\theta_j)}{f(\theta)} \leq \max \left( 2, \sup_{t \in [\delta,2\pi-\delta]} \frac{f(t)}{\inf_{t \in [0,2\pi]} f(t)} \right).$$
Denote the upper bound of \( \frac{f(\theta)}{\tr Q^2_T(\theta)} \) as \( F \). Using the result of 4.1.2, there exist \( \delta > 0, K > 0 \) such that for any \( x \in (0, \delta) \),

\[
P\left( \sup_{\theta \in (0,2\pi)} \chi_2(\theta) > x \right) \leq T^\delta P\left( \chi_2(\theta_j) > \frac{x}{F} \right) \leq 2T^\delta \exp\left( - \frac{Nx^2}{KF \log^2 T} \right).
\]

Finally, combining the above estimations for large deviation probabilities for \( \chi_i, i = 1, 2, 3 \), the result of the theorem follows.

4.1.4. Estimation of \( \tr Q^2_T(\theta) \). The above proof is based on the estimation of uniform upper bound of \( \tr Q^2_T(\theta) \). In order to estimate \( \tr Q^2_T(\theta) \), we first estimate the norm of the Toeplitz matrix \( R_T \).

**Lemma 4.2.** If \( f \) satisfies A6, then

(a) \( \sup\{f(t) : x \leq t \leq \pi\} \sim f(x) \) as \( x \to 0^+ \).
(b) \( \inf\{f(t) : 0 < t \leq \pi\} \sim f(x) \) as \( x \to 0^+ \).
(c) \( \int_0^\pi f(t) \, dt \sim \frac{x^{1-n}L(x^{-1})}{1-a} \) as \( x \to 0^+ \).
(d) \( \int_x^\pi \frac{f(t)}{t} \, dt \sim \frac{L(x^{-1})}{ax} \) as \( x \to 0^+ \).

**Proof.** By changing the variable \( u = x^{-1} \), (a) and (b) follow from Theorem 1.5.3 of [3], (c) from Proposition 1.5.10 of [3], and (d) from Proposition 1.5.8 of [3]. \( \square \)

**Lemma 4.3.** If \( (R_T) \) is a sequence of Toeplitz matrices satisfying A4 and A6, then

\[ \|R_T\| \asymp T^n L(T). \]

**Proof.** First we prove that \( \|R_T\| = O(T^n L(T)) \). It is well known that

\[ \|R_T\| \leq \sup_{\theta \in [-\pi, \pi]} |\Upsilon_T(\theta)| \]

where \( \Upsilon_T \) is defined in (30). The function \( \Upsilon_T \) has an integral formula

\[ \Upsilon_T(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) D_T(\theta - x) \, dx \]

where \( D_T(x) = \frac{\sin[(T+1/2)x]}{\sin(x/2)} \) is the \( T \)-th Dirichlet kernel function. Let \( 0 < \delta < \pi \), then we have

\[ \Upsilon_T(\theta) = \frac{1}{2\pi} \left( \int_{-\delta}^{\delta} + \int_{\delta < |x| < \pi} \right) f(x) D_T(\theta - x) \, dx \]

\[ =: \Upsilon_{\delta,T}(\theta) + \frac{1}{2\pi} \int_{\delta < |x| < \pi} f(x) D_T(\theta - x) \, dx. \]

Note that under A4,

\[ \left| \int_{\delta < |x| < \pi} f(x) D_T(\theta - x) \, dx \right| \lesssim \int_{-\pi}^{\pi} |D_T(x)| \, dx = O(\log T). \]

We just need to prove that \( \sup_\theta \Upsilon_{\delta,T} = O(T^n L(T)) \). We have

\[ |D_T(x)| \lesssim \min \left( \frac{1}{|x|}, T \right) =: h_T(x) \quad \text{for} \quad x \in [-\pi - \delta, \pi + \delta]. \]

Therefore

\[ |\Upsilon_{\delta,T}(\theta)| \lesssim \int_{-\delta}^{\delta} f(x) h_T(\theta - x) \, dx. \]

Now we prove that \( \sup_\theta |\Upsilon_T(\theta)| = O(T^n L(T)) \). We can suppose that \( f \) is decreasing on \( (0, \pi) \), because otherwise we can replace \( f \) by

\[ \hat{f}(x) := \sup_{|x| \leq t \leq \pi} \{f(t)\}, \]
Thus we can see from Lemma 4.2 (a) that \( \tilde{L}(x) := |x|^{-a} \tilde{f}(x^{-1}) \) is slowly varying at \( \infty \) and \( \tilde{f}(x) = |x|^{-a} \tilde{L}(x^{-1}) \). Let \( R_T \) be a Toeplitz matrix with spectral density \( \tilde{f} \), then because \( \tilde{f} \geq f \), we have \( \|R_T\| \geq \|R_T\| \). Assuming that \( f \) is decreasing on \([0, \pi] \), we have

\[
\int_{-\delta}^{\delta} f(x) h_f(\theta - x) \, dx = \left( \int_{-\delta}^{0} + \int_{0}^{\delta} \right) \frac{f(x)}{|\theta - x|} \mathbb{1}_{\{|x| \geq \frac{\theta}{\delta} \}} \, dx + T \int_{-\delta}^{\delta} f(x) \mathbb{1}_{\{|\theta - x| \leq \frac{\theta}{T} \}} \, dx =: P_1(\theta) + P_2(\theta) + P_3(\theta).
\]

From Lemma 4.2 (c), we have

\[ P_3(\theta) \leq P_3(0) = O(T^a L(T)). \]

Next we prove that \( P_1(\theta) \leq T^a L(T) \) uniformly in \( \theta \). When \( \theta = 0 \), from Lemma 4.2(d), we have

\[ P_1(0) = \int_{-\frac{\theta}{T}}^{\frac{\theta}{T}} f(x) \, dx = O(T^a L(T)). \]

When \( \theta < 0 \), because \((x - \theta)^{-1} \mathbb{1}_{x > \theta + \frac{\theta}{T}} \leq h(x) \) for \( x > 0 \), we have

\[ P_1(\theta) \leq P_1(0) + P_3(0) = O(T^a L(T)). \]

When \( \theta > 0 \), we have

\[ P_1(\theta) = \mathbb{1}_{\{\theta > \frac{\theta}{T} \}} \int_{0}^{\theta + \frac{\theta}{T}} f(x) \, dx + \int_{\theta + \frac{\theta}{T}}^{\delta} f(x) \, dx. \]

Because we have assumed that \( f \) is decreasing on \((0, \pi] \), we have

\[
\int_{\theta + \frac{\theta}{T}}^{\delta} \frac{f(x)}{x - \theta} \, dx \leq \int_{\theta + \frac{\theta}{T}}^{\delta} \frac{f(x - \theta)}{x - \theta} \, dx = \int_{\frac{\theta}{T}}^{\frac{\theta}{T}} \frac{f(x)}{x} \, dx = O(T^a L(T));
\]

and if \( \theta > \frac{\theta}{T} \), by Chebyshev’s sum inequality (see 43, Page 43 of [7]), we have

\[
\int_{0}^{\theta - \frac{\theta}{T}} \frac{f(x)}{\theta - x} \, dx \leq \left( \theta - \frac{1}{T} \right) \int_{\theta - \frac{\theta}{T}}^{\theta - \frac{\theta}{T}} f(x) \, dx \int_{\theta - \frac{\theta}{T}}^{\theta - \frac{\theta}{T}} \frac{1}{\theta - x} \, dx \\
\leq \int_{0}^{\theta - \frac{\theta}{T}} \frac{f(x)}{\theta - x} \, dx \leq P_1(- \frac{1}{T}) = O(T^a L(T)).
\]

The above estimations are uniform for \( \theta \), thus we have proved \( P_1(\theta) \leq T^a L(T) \) uniformly for \( \theta \). The term \( P_2(\theta) \) is treated similarly, then it follows that \( \|R_T\| \leq T^a L(T) \).

For the other direction \( \|R_T\| \geq T^a L(T) \), by Theorem 2.1 of [4], we have

\[
\|R_T\| \geq \sup_{\theta \in (-\pi, \pi)} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) F_T(x - \theta) \, dx \geq \frac{1}{2 \pi} \int_{-\pi}^{\pi} f(x) F_T(x) \, dx,
\]

where \( F_T \) denotes the Fejér kernel

\[ F_T(x) = \frac{\sin^2 \frac{T x}{2}}{T \sin^2 \frac{x}{2}}. \]

By the property of the function \( \text{sinc}(x) = \sin(x)/x \), if \( x = y/T \) with \( y \in (-1, 1) \),

\[ F_T(x) = \frac{\sin^2 \frac{y}{2}}{T \sin^2 \frac{y}{2T}} \geq \frac{T}{2} \sin^2 \left( \frac{y}{2} \right) \geq T. \]

Thus

\[ \|R_T\| \geq T \int_{-\frac{\theta}{T}}^{\frac{\theta}{T}} f(x) \, dx \geq T^a L(T). \]

□
From this lemma and (11) of [20], we get a rough global estimation

\[ \text{tr} Q_T^2(\theta) = O(T^{2a} L^2(T) \log T). \]  

We note that this bound tends to infinity, and is even much larger than \( T \) when \( a > 1/2 \). If this bound is sharp for some \( \theta \), then as \( a > 1/2 \) and \( N, T \to \infty \) at the same speed, the variance of \( \hat{\Upsilon}_T(\theta) \) does not tend to 0. However this only happens when \( \theta \) is near 0. For \( \theta \in [\delta, 2\pi - \delta] \) where \( \delta > 0 \) is an arbitrary small number, we can find a much better bound. Recall that the local norm \( \| \cdot \|_{(x, \delta)} \) is defined in (16).

**Proposition 4.4.** Let \( Q_T(\theta) \) be defined in (34) with \( R_T \) 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 \( \delta \in (0, \pi) \),

\[ \frac{\text{tr} Q_T^2(\theta)}{2 \log T} \leq \|f\|^2_{(\theta, \delta)} + \frac{K \|f\|_1 (\|f\| + \|f\|_{(\theta, \delta)})}{\delta^4 \log T} \]  

Moreover if \( f \) is continuous at \( \theta \), then

\[ \lim_{T \to \infty} \frac{\text{tr} Q_T^2(\theta)}{2 \log T} = f^2(\theta). \]

**Proof.** Let \( R_T = (r_{i-j}) \). Using the integral expression of \( r_{i-j} \), we write

\[ \text{tr} Q_T^2(\theta) = \sum_{i,j,k,l} r_{i-j} e^{i(j-k)\theta} T_{i-j} e^{i(l-i)\theta}.\]

\[ = \sum_{i,j,k,l} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-i(j-k)x} dx \int_{-\pi}^{\pi} f(y) e^{-i(l-i)y} dy \int_{-\pi}^{\pi} f(x+y+\theta) f(y+\theta) \sum_{i,j,k,l} e^{i(j-k)x+i(l-i)y} \frac{T_{i-j}}{|T-|i-j|} dx dy \]

\[ = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x+\theta) f(y+\theta) \sum_{i,j,k,l} e^{i(jx-iy)} \frac{1}{T-|i-j|} dx dy \]

Denote

\[ g(x,y) := \sum_{1 \leq i,j \leq T} e^{i(jx-iy)} \frac{T_{i-j}}{T-|i-j|}. \]

Note that the equality (42) also holds for \( R_T = I \) and correspondingly \( f = 1 \), then we get

\[ \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |g(x,y)|^2 dx dy = \text{tr} B_T^2 = 1 + 2 \sum_{k=1}^{T-1} k \sim 2 \log T. \]

If we consider \( |g(x,y)|^2 \) as a measure density on \([-\pi, \pi]^2\), the total mass of this measure is asymptotically \( 8\pi^2 \log T \).

For \( \delta \in (0, \pi) \), let \( E_{-\delta} := [-\pi, \pi] \setminus (-\delta, \delta) \). Then we have

\[ 4\pi^2 \text{tr} Q_T^2(\theta) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x+\theta) f(y+\theta) |g(x,y)|^2 dx dy \]

\[ = \left( \int_{E_{-\delta}} \int_{E_{-\delta}} f(x+\theta) f(y+\theta) |g(x,y)|^2 dx dy + \int_{E_{\delta}} \int_{E_{-\delta}} f(x+\theta) f(y+\theta) |g(x,y)|^2 dx dy \right) \]

\[ \leq \left( \int_{E_{\delta/2}} \int_{E_{\delta/2}} f(x+\theta) f(y+\theta) |g(x,y)|^2 dx dy + \int_{E_{1/2}} \int_{E_{1/2}} f(x+\theta) f(y+\theta) |g(x,y)|^2 dx dy \right) \]

\[ := P_1 + P_2 + P_3 + P_4, \]

where we have the inequality because on the RHS the integral on the region \( \{(x,y) \in [-\pi, \pi] : \delta/2 \leq |x|, |y| \leq \delta \} \) is repeated. We will show that \( P_1 + P_2 + P_3 \) is bounded, and the main contribution to the integral is \( P_4 \). Also note that the integrated functions are \( 2\pi \)-periodic on \( x \) and \( y \), so \( E_{\delta} \) and \( E_{\delta/2} \) can be replaced by \([\delta, 2\pi - \delta] \) and \([\delta/2, 2\pi - \delta/2] \) in the integrals.
We estimate the integral
\begin{equation}
\int g(x, y) = \frac{1}{T} \sum_{j=0}^{T-1} e^{i j y (x-y)} + \sum_{m=1}^{T-1} \sum_{j=0}^{T-m-1} e^{i (x+m x-y+j y)} + \sum_{j=0}^{T-m-1} e^{i (x-my-j y)}
\end{equation}

(46)

where

\begin{equation}
1 \leq \| f \|_{\infty}^2 \leq \frac{4 \sin^2(\delta/4)}{\delta^2}
\end{equation}

Next we prove that for any $0 < \delta < \pi$, $|g_2(x, y)|$ and $|g_3(x, y)|$ are bounded by $K/\delta^2$ for some absolute constant $K > 0$ on $[\delta/2, 2\pi - \delta/2]^2$. Let $z_1 = e^{ix}$, $z_2 = e^{iy}$ be two distinct points on the arc $\{ z \in \mathbb{C} : |z| = 1, \arg(z) \in [\delta/2, 2\pi - \delta/2] \}$. We have

\begin{equation}
|g_2(x, y)| = \frac{1}{2 \pi} \int_{|z_1|}^{T-1} \int_{|z_1|}^{T-1} \sum_{k=0}^{T-2} z^k \, dz \leq \sup_{z \in [z_1, z_2]} \left| 1 - z^{T-1} \right| \leq \frac{1}{\sin^2(\delta/4)},
\end{equation}

where $[z_1, z_2]$ denotes the segment between $z_1$ and $z_2$. Therefore we have $|g_2(x, y)| \leq \delta^{-2}$. The same estimation also applies to $g_3$ and we get $|g(x, y)| \leq \delta^{-2}$ for $(x, y) \in [\delta/2, 2\pi - \delta/2]^2$. Then we have

\begin{equation}
P_1 \lesssim \frac{\| f \|_{\infty}^2}{\delta^4},
\end{equation}

where the implicit constant is absolute.

For $P_2$, because $|1 - e^{i(x-y)}|^2 \geq 4 \sin^2(\delta/4)$ when $x \in (\delta, 2\pi - \delta)$ and $y \in (-\delta/2, 2\pi/2)$, we have

\begin{equation}
P_2 \leq \frac{\| f \|_{\infty}^2}{4 \sin^2(\delta/4)} \int_{\delta}^{2\pi-\delta} dx f(x) \int_{-\delta/2}^{\delta/2} |(1 - e^{i(x-y)}) g(x, y)|^2 dy.
\end{equation}

We estimate the integral $\int_{-\delta/2}^{\delta/2} |(1 - e^{i(x-y)}) g(x, y)|^2 dy$ for every $x \in (\delta, 2\pi - \delta)$. We have

\begin{equation}
\int_{-\delta/2}^{\delta/2} |(1 - e^{i(x-y)}) g(x, y)|^2 dy \leq \frac{4\delta}{T^2};
\end{equation}

using the inequality $|a - b|^2 \leq 2|a|^2 + 2|b|^2$, we have

\begin{equation}
\int_{-\delta/2}^{\delta/2} |(1 - e^{i(x-y)}) g_2(x, y)|^2 dy = \int_{-\delta/2}^{\delta/2} \sum_{k=1}^{T-1} \frac{e^{-iky} - e^{-ikx}}{k}^2 dy \leq 2 \sum_{k=1}^{T-1} \left| e^{-iky} - e^{-ikx} \right|^2 dy + 2\delta \sum_{k=1}^{T-1} \left| \frac{e^{-ikx}}{k} \right|^2
\end{equation}

\begin{equation}
\leq 2 \sum_{k=1}^{T-1} \left| e^{-iky} \right|^2 dy + 2\delta \sum_{k=1}^{T-1} \left| \frac{e^{-ikx}}{k} \right|^2
\end{equation}

\begin{equation}
= 2 \sum_{k=1}^{T-1} \left| \frac{e^{-ikx}}{k} \right|^2 dy + 2\delta \sum_{k=1}^{T-1} \left| \frac{e^{-ikx}}{k} \right|^2.
\end{equation}

For $x \in (\delta, 2\pi - \delta)$ we have

\begin{equation}
\sum_{k=1}^{T-1} \left| \frac{e^{-ikx}}{k} \right|^2 = \int_{[0, e^{-ix}]} \left| \frac{1 - z^{T-1}}{1 - z} \right| \, dz \leq \frac{2}{\sin \delta}
\end{equation}
where the complex integral is taken along the segment from 0 to $e^{-ix}$. Then the following inequality holds with two absolute constants $K_1, K_2$:

$$\int_{-\delta/2}^{\delta/2} |(1 - e^{i(x-y)}) g_2(x, y)|^2 dy \leq K_1 + K_2/\delta.$$ 

The integral expression corresponding to $g_3$ is similarly estimated. Taking these into (46), we get

$$P_2 \leq \|f\|_{(\theta, \delta)} \|f\|_1 \frac{K_1 + K_2 \delta + K_3 \delta^2 / T}{\delta^3} \lesssim \|f\|_{(\theta, \delta)} \|f\|_1.$$ 

The same bound also controls $P_3$.

For $P_4$, from (43) we have

$$P_4 \leq 8\pi^2 \|f\|_{(\theta, \delta)}^2 \log T.$$

Summarising the bounds for $P_1, P_2, P_3, P_4$ and dividing $8\pi^2 \log T$, the result follows.

To prove (41), if $f$ is continuous at $\theta$, then for any $\varepsilon > 0$, there is some $\delta > 0$ such that

$$|f(x + \theta)f(y + \theta) - f^2(\theta)| \leq \varepsilon, \quad \forall x, y \in (-\delta, \delta).$$

Note that the sum of the first three integrals in the second line of (44) is bounded by $P_1 + P_2 + P_3$, we have

$$\frac{\text{tr} \ Q_1^2(\theta)}{2 \log T} = \frac{1}{8\pi^2 \log T} \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} f(x + \theta)f(y + \theta)|g(x, y)|^2 dx dy + O\left(\frac{1}{\log T}\right).$$

All the above arguments apply also to $\mathbf{R}_T = f^2(\theta)\mathbf{I}$, then

$$\frac{\text{tr} \ B_1^2(\theta)}{2 \log T} f^2(\theta) = \frac{1}{8\pi^2 \log T} \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} f^2(\theta)|g(x, y)|^2 dx dy + O\left(\frac{1}{\log T}\right).$$

Take the difference of the last two equations and then the absolute value, also note (47) and (43), we get

$$\left|\frac{\text{tr} \ Q_1^2(\theta)}{2 \log T} - \frac{\text{tr} \ B_1^2(\theta)}{2 \log T} f^2(\theta)\right| \leq \varepsilon + O\left(\frac{1}{\log T}\right).$$

Let $T \to \infty$ and $\varepsilon \to 0$, we get the result. \qed

As a consequence of Proposition 4.4, if $f$ satisfies $\mathbf{A}_4$, then $\text{tr} \ Q_1^2(\theta) / \log T$ is uniformly bounded in $T$ and in $\theta \in (\delta, 2\pi - \delta)$ for any fixed $\delta \in (0, \pi)$. However, when $\mathbf{A}_6$ holds, it is also important to estimate $\text{tr} \ Q_1^2(\theta)$ for $\theta$ near zero.

**Lemma 4.5.** There exists a constant $K > 0$ such that for any $-\pi \leq x, y \leq \pi$ with $xy \neq 0$, we have

$$\left|\frac{1}{x - y} \sum_{k=1}^{T} e^{ikx} - e^{iky} \right| \frac{k}{\sqrt{|xy|}} \leq \frac{K}{\sqrt{|xy|}}.$$ 

**Proof.** We first assume that $xy > 0$, i.e. $x, y$ have the same sign. Assume further without loss of generality that $0 < y < x \leq \pi$. Then

$$\left|\sum_{k=1}^{T} \frac{e^{ikx} - e^{iky}}{k}\right| = \left|\sum_{k=0}^{T-1} e^{iks} \right| \leq \int_{y}^{x} \frac{1}{\sin(s/2)} ds \leq \int_{y}^{\pi} \frac{\pi}{s} ds = \pi (\log x - \log y).$$

We prove that $\frac{\log x - \log y}{x - y} \leq \frac{1}{\sqrt{|xy|}}$. Let $x = e^u, y = e^v$, it suffices to prove that

$$e^{\frac{u-v}{2}} - e^{\frac{u-v}{2}} \geq u - v.$$
For $u = v$, the equality holds. By differentiating we can see that the function

$$u \mapsto e^{\frac{u-v}{2}} - e^{\frac{v-u}{2}} - u$$

is increasing with $u$. Then for $u > v$, (48) holds and we have

$$\left| \frac{1}{x-y} \sum_{k=1}^{T} \frac{e^{ikx} - e^{iky}}{k} \right| \leq \frac{\pi}{\sqrt{|xy|}}.$$

If $-\pi \leq y < 0 < x < \pi$, we can assume without loss of generality that $|y| \leq x$. Then

$$\sum_{k=1}^{T} \frac{e^{ikx} - e^{iky}}{k} = \sum_{k=1}^{T} \frac{e^{ikx} - e^{-iky}}{k} + 2i \sum_{k=1}^{T} \frac{\sin ky}{k}$$

and by Theorem 1.3 and Remark(b) on Page 183 of [23], the series $\sum_{k=1}^{T} \frac{\sin ky}{k}$ is uniformly bounded (say, by $K$). Then from the first part of the proof, and using the inequality of arithmetic and geometric means, we have

$$\left| \frac{1}{x-y} \sum_{k=1}^{T} \frac{e^{ikx} - e^{iky}}{k} \right| \leq \frac{1}{x-(-y)} \left| \sum_{k=1}^{T} \frac{e^{ikx} - e^{-iky}}{k} \right| + \frac{2K}{x-y} \leq \frac{\pi + K}{\sqrt{|xy|}}.$$

Proposition 4.6. Let $Q_{T}(\theta)$ be defined as (34) with $R_{T}$ having spectral density $f$ satisfying A4, A5, A6. Then

$$\text{tr} \frac{Q_{T}^{2}(\theta)}{f^{2}(\theta) \log^{2} T}$$

is uniformly bounded in $\theta \in [-\pi, \pi]$ and $T \geq 1$.

Proof. If $\theta = O(1/T)$, by (39), we have $\text{tr} Q_{T}^{2}(\theta) = O(f^{2}(\theta) \log T)$; if $\theta \in [-\pi, \pi] \setminus [-\delta, \delta]$ with any $\delta > 0$, by Proposition 4.4, we have $\text{tr} Q_{T}^{2}(\theta) = O(\log T)$. So we can find two sequences of positive numbers $1/T \ll \tau_{T} < \delta_{T} = o(1)$ such that

$$\text{tr} \frac{Q_{T}^{2}(\theta)}{f^{2}(\theta) \log^{2} T}$$

is uniformly bounded in $[-\pi, -\delta_{T}] \cup [-\tau_{T}, \tau_{T}] \cup [\delta_{T}, \pi]$. In the following we prove the uniform boundedness of (49) for $|\theta| \in (\tau_{T}, \delta_{T})$.

From (42), we have

$$\text{tr} Q_{T}^{2}(\theta) \leq \frac{1}{4\pi^{2}} \left( \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x + \theta) f(y + \theta) |g(x, y)|^{2} \, dx \, dy \right).$$

From the proof of Proposition 4.4, the function $g(x, y)$ is uniformly bounded on $([-\pi, -\pi/2] \cup [\pi/2, \pi])^{2}$, thus the third integral is uniformly bounded. We only need to estimate the first integration, because the second one is similar to the first.

When $|x| \leq \pi/2, |y| \leq \pi$, we have $\frac{|x-y|}{2} < \frac{3\pi}{2}$, then

$$\frac{|x-y|}{2 |\sin \frac{\pi}{2}|} \leq \frac{2\sqrt{3}}{3\pi}.$$ 

We write $g = g_{1} + g_{2} + g_{3}$ as in the proof of Proposition 4.6, and we have $|g_{1}(x,y)| \leq 1$. By Lemma 4.5, for $|x| \leq \pi/2, |y| \leq \pi$, we have

$$|g_{2}(x,y)| \lesssim \frac{1}{\sqrt{|xy|}}.$$
By (50), we have
\[ |g_2(x, y)| \lesssim \frac{|x - y|}{2 \sin \frac{x - y}{2}} |x - y| |k|^{-1} e^{ik \theta} 
\lesssim \frac{1}{|x - y|} \log T \]

The above two bounds also controls $|g(x, y)|$ because $|g| \leq |g_1| + |g_2| + |g_3|$, where $g_2$ and $g_3$ are similarly bounded, and $|g_1| \leq 1$ can be absorbed into the other two bounds. On the other hand $|g(x, y)|$ is obviously bounded by $g(0, 0) = 2T + 1$. In the sequel we will use different bounds in different subsets of \{(x, y) \in \mathbb{R}^2 : |x| \leq \pi/2, |y| \leq \pi\}:

\[
|g(x, y)| \lesssim \begin{cases} 
T & \text{if } |x|, |y| < \frac{\pi}{2T}; \\
\frac{1}{\log T} & \text{if } |x|, |y| > \frac{\pi}{T}; \\
|x| < \frac{\pi}{T}, |y| > \frac{\pi}{T} \text{ or } |y| < \frac{\pi}{T}, |x| > \frac{\pi}{T}.
\end{cases}
\] (51)

Similar to (44), we have
\[
\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x + \theta) f(y + \theta) g(x, y)^2 \, dx \, dy \lesssim \begin{aligned}
& \int_{|x| < \frac{\pi}{T}} \int_{|y| < \frac{\pi}{T}} f(x + \theta) f(y + \theta) \, dx \, dy \\
&+ \int_{|x| < \frac{\pi}{T}} \int_{|y| > \frac{\pi}{T}} f(x + \theta) f(y + \theta) \, dx \, dy \\
&+ \int_{|x| > \frac{\pi}{T}} \int_{|y| < \frac{\pi}{T}} f(x + \theta) f(y + \theta) \, dx \, dy \\
&+ \int_{|x| > \frac{\pi}{T}} \int_{|y| > \frac{\pi}{T}} f(x + \theta) f(y + \theta) \, dx \, dy
\end{aligned}
\]

\[ =: P_1 + P_2 + P_3 + P_4, \]

where the integrals on the RHS are all taken within the region $(x, y) \in (-\pi/2, \pi/2) \times (-\pi, \pi)$, which are omitted in order to abbreviate the notations.

When $x \in [-2/T, 2/T]$ and $|\theta| \in [\tau_T, \delta_T]$, as $T$ is large enough, we have $\theta^{-1}(\theta + x) \in (1 - \varepsilon, 1 + \varepsilon)$ for some $\varepsilon \in (0, 1)$. Then from the Uniform Convergence Theorem (UCT, Theorem 1.2.1 in [3]) of slowly varying function, we have

\[
f(x + \theta) \sim \frac{|\theta|^a L(|\theta + x|^{-1})}{|\theta + x|^a L(|\theta|^{-1})} \xrightarrow{T \to \infty} 1
\]

uniformly in $x$ and $\theta$. Combining with the first clause of (51), we have

\[ P_1 \lesssim T^2 \int_{|x| < \frac{\pi}{T}} \int_{|y| < \frac{\pi}{T}} f(x + \theta) f(y + \theta) \, dx \, dy = O(f^2(\theta)). \]

For $P_2$ (and similarly $P_3$), by the third clause of (51) and (53), we have

\[
P_2 \lesssim \log^2 T \int_{|x| < \frac{\pi}{T}} \int_{|y| > \frac{\pi}{T}} |x - y|^2 f(x + \theta) f(y + \theta) \, dx \, dy \lesssim f(\theta) T^{-1} \log^2 T \int_{|y| > \frac{\pi}{T}} (|y| - \frac{\pi}{T})^2 f(y + \theta) \, dy
\]

with

\[
\int_{|y| > \frac{\pi}{T}} (|y| - \frac{\pi}{T})^2 f(y + \theta) \, dy = \int_{-\pi}^{\frac{\pi}{T}} f(y + \theta) \, dy + \int_{\frac{\pi}{T}}^{\frac{\pi}{T}} f(y + \theta) \, dy.
\] (55)

The function $f$ being even, the LHS of the above equality is even for $\theta$. Thus we can assume that $\theta > 0$. Then as $T \to \infty$, the following holds uniformly for $\theta \in (0, \pi)$ and $\theta \in (\tau_T, \delta_T)$:

\[ f(y + \theta) \leq \sup_{t \geq \omega} f(t) \sim f(\theta). \]

Thus

\[
\int_{\frac{\pi}{T}}^{\frac{\pi}{T}} f(y + \theta) \, dy = O(f(\theta)) \int_{\frac{\pi}{T}}^{\frac{\pi}{T}} \frac{1}{y^2} \, dy = O(f(\theta) T).
\]
For the other integral on the RHS of (55), we have
\[
\int_{-\pi}^{\pi} \frac{f(y + \theta)}{(y + \frac{1}{T})^2} \, dy = \int_{-\pi}^{\frac{\pi}{T} + \theta} \frac{f(y)}{y^2} \, dy + \int_{\frac{\pi}{T} + \theta}^{\pi} \frac{f(y)}{(y - \theta + \frac{1}{T})^2} \, dy
\]
\[
= \sup_{|y| > \frac{\pi}{T} + \theta} \int_{-\pi + 1/T}^{y} \frac{1}{y^2} \, dy + O(\theta^{-2}) \int_{\frac{\pi}{T} + \theta}^{\pi} f(y) \, dy
\]
\[
= O(f(\theta/2)T) + O(\theta^{-2} \theta^{-1} \nu(L(\theta^{-1}))),
\]
and by the definition of slowly varying function,
\[
f(\theta/2) = 2^a \theta^{-a} \nu(2\theta^{-1}) \sim 2^a \theta^{-a} \nu(\theta^{-1}) = 2^a f(\theta)
\]
uniformly for \( \theta \in (\pi_T, \delta_T) \), thus
\[
\int_{-\pi}^{\pi} \frac{f(y + \theta)}{(y + \frac{1}{T})^2} \, dy = O(f(\theta)T) + O(\theta^{-1} f(\theta)).
\]
Combining the above estimations, we have
\[
P_2 \leq f(\theta)T^{-1} \log^2 T[f(\theta)T + \theta^{-1} f(\theta)] = (1 + \frac{1}{T \theta})f^2(\theta) \log^2 T \sim f^2(\theta) \log^2 T.
\]
For \( P_4 \), using the second clause of (51), we have
\[
P_4 = \int_{|x| > \frac{\pi}{T}} \int_{|y| > \frac{\pi}{T}} f(x + \theta)f(y + \theta)|g(x, y)| \, dx \, dy
\]
\[
\leq \int_{|x| > \frac{\pi}{T}} \int_{|y| > \frac{\pi}{T}} f(x + \theta)f(y + \theta) \frac{1}{|xy|} \, dx \, dy
\]
\[
= \left( \int_{\frac{\pi}{T}}^{\frac{\pi}{T}} f(\theta + x) + f(\theta - x) \, dx \right) \left( \int_{\frac{\pi}{T}}^{\frac{\pi}{T}} f(\theta + y) + f(\theta - y) \, dy \right).
\]
Using similar methods as in the estimation of \( P_2 \), we have
\[
P_4 \leq f^2(\theta) \log^2 T.
\]
Summarizing the above estimations, the result follows.

4.2. Proof of Proposition 2.3. We first prove that a necessary condition of the convergence

\[
\| \mathbf{R}_T^{-1/2} \mathbf{R}_T^b \mathbf{R}_T^{-1/2} - \mathbf{\xi} \mathbf{I} \| \xrightarrow{T \to \infty} 0,
\]
is that

\[
\lim_{T \to \infty} \frac{\lambda_{\max}(\mathbf{R}_T^b)}{\lambda_{\max}(\mathbf{R}_T)} = \lim_{T \to \infty} \frac{\lambda_{\min}(\mathbf{R}_T^b)}{\lambda_{\min}(\mathbf{R}_T)} = \xi.
\]
Dividing \( \xi \) on both sides of (56), the convergence (56) is equivalent to

\[
\| \left( \mathbf{R}_T \right)^{-1/2} \mathbf{R}_T^b \left( \mathbf{R}_T \right)^{-1/2} - \mathbf{I} \| \xrightarrow{T \to \infty} 0.
\]
Take an arbitrary \( \varepsilon > 0 \). Let \( u \) be an eigenvector of \( \mathbf{R}_T \) associated with \( \lambda_{\max}(\mathbf{R}_T) \), then for large enough \( T \),

\[
1 - \varepsilon < u^*(\mathbf{R}_T)^{-1/2} \mathbf{R}_T^b (\mathbf{R}_T)^{-1/2} u = \frac{u^* \mathbf{R}_T^b u}{\xi \lambda_{\max}(\mathbf{R}_T)} \leq \frac{\lambda_{\max}(\mathbf{R}_T^b)}{\xi \lambda_{\max}(\mathbf{R}_T)}.
\]
Let \( v \) be an eigenvector of \( \mathbf{R}_T \) associated with \( \lambda_{\min}(\mathbf{R}_T) \), then for large enough \( T \),

\[
1 + \varepsilon > v^*(\mathbf{R}_T)^{-1/2} \mathbf{R}_T^b (\mathbf{R}_T)^{-1/2} v = \frac{v^* \mathbf{R}_T^b v}{\xi \lambda_{\min}(\mathbf{R}_T)} \geq \frac{\lambda_{\min}(\mathbf{R}_T^b)}{\xi \lambda_{\min}(\mathbf{R}_T)}.
\]
For an Hermitian matrix $A_T$, the convergence $\|A_T - I\| \to 0$ is equivalent to $\lambda_{\max}(A_T) \to 1, \lambda_{\min}(A_T) \to 1$. We note that

$$\lambda_i((R^b_T)^{-1/2}(\xi R_T)(R^b_T)^{-1/2}) = \frac{1}{\lambda_{T-i}(\xi R_T)^{-1/2}R^b_T(\xi R_T)^{-1/2}}.$$

then \(58\) implies

$$\|(R^b_T)^{-1/2}(\xi R_T)(R^b_T)^{-1/2} - I\| \to 0.$$

Using the same arguments of \(59\) and \(60\) we get, for large enough $T$,

\[
1 - \varepsilon \leq \frac{\xi \lambda_{\max}(R_T)}{\lambda_{\max}(R^b_T)} \quad \text{and} \quad \frac{\xi \lambda_{\min}(R_T)}{\lambda_{\min}(R^b_T)} \leq 1 + \varepsilon.
\]

Combining \(59\), \(60\) and \(61\), we have

\[
\lim_{T \to \infty} \frac{\lambda_{\max}(R^b_T)}{\xi \lambda_{\max}(R_T)} = \lim_{T \to \infty} \frac{\lambda_{\min}(R^b_T)}{\xi \lambda_{\min}(R_T)} = 1,
\]

and \(57\) follows.

However, we will prove that \(57\) cannot be satisfied by $R^b_T$ and $R_T$ defined in Remark 2.3. Indeed, let $\mathcal{K}$ and $\mathcal{K}^b$ be two integral operators on $L^2(0,1)$ defined by

$$\mathcal{K}(\varphi)(x) = \int_0^1 \frac{1}{|x-y|^{1-a}} \varphi(y) dy, \quad \mathcal{K}^b(\varphi)(x) = \int_0^1 \frac{1}{|x-y|^{1-a}} \varphi(y) dy, \quad \text{for} \ \varphi \in L^2(0,1).$$

Using the technique of the proof of Theorem 2.3 in [13], we can prove that

$$\frac{\lambda_{\max}(R_T)}{T^a} \to \lambda_1(\mathcal{K}), \quad \frac{\lambda_{\max}(R^b_T)}{T^a} \to \lambda_1(\mathcal{K}^b).$$

We can 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 $\varphi^b$ 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 where we conclude that

\[
\lim_{T \to \infty} \frac{\lambda_{\max}(R^b_T)}{\lambda_{\max}(R_T)} = \frac{\lambda_1(\mathcal{K}^b)}{\lambda_1(\mathcal{K})} < 1.
\]

For the smallest eigenvalues, let

$$f(\theta) = 1 + 2 \sum_{k=1}^{\infty} \frac{1}{(1 + k)^a} \cos(k \theta), \quad \Upsilon_T(\theta) := 1 + 2 \sum_{k=1}^{T-1} \frac{1}{(1 + k)^a} \cos(k \theta)$$

be the spectral density of $R_T$ and its partial Fourier series, and let

$$\Upsilon_T^b(\theta) := r_0^b + 2 \sum_{k=1}^{T-1} r_k^b \cos(k \theta) = 1 + 2 \sum_{k=1}^{T-1} \left(1 - \frac{k}{T}\right) \frac{1}{(1 + k)^a} \cos(k \theta).$$

Note that $\Upsilon_T^b$ is just the Cesáro mean of $\Upsilon_T$, therefore

$$\Upsilon_T^b(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) F_T(\theta - x) dx$$

where $F_T(x) = \frac{\sin^2(Tx/2)}{T \sin^2(x/2)}$ is the Fejér kernel. Then we have

$$\lambda_{\min}(R^b_T) \geq \min_{\theta} \Upsilon_T^b(\theta) \geq \text{ess inf}_{\theta} \{f(\theta)\}.$$ 

From Section 5.2(b) of [8],

$$\lim_{T \to \infty} \lambda_{\min}(R_T) = \text{ess inf}_{\theta} \{f(\theta)\}.$$
Therefore,

$$\lim_{T \to \infty} \frac{\lambda_{\min}(R_T^b)}{\lambda_{\min}(R_T)} \geq 1.$$  \hspace{1cm} (64)

The inequalities (63) and (64) contradict (57).

Acknowledgement. The authors are financially supported by Department of Statistics and Actuarial Science of the University of Hong Kong. We also thank Professor Romain Couillet in Université Grenoble-Alpes, who posed this question, for fruitful discussions.

References

[1] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.

[2] J. Baik and J. W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

[3] N. H Bingham, C. M Goldie, and J.L Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.

[4] A. Bottcher and S. Grudsky. Fejér means norms of large toeplitz matrices. *Acta Scientiarum Mathematicarum*, 69(3-4):889–900, 2003.

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

[6] X. Ding and F. Yang. Spiked separable covariance matrices and principal components. *arXiv preprint arXiv:1905.13060*, 2019.

[7] K. Fang, S. Kotz, and K. Ng. *Symmetric multivariate and related distributions*. CRC Press, 2018.

[8] U. Grenander and G. Szegö. *Toeplitz forms and their applications*, volume 321. University of California Press, 2001.

[9] K. Johansson. Shape fluctuations and random matrices. *Communications in mathematical physics*, 209(2):437–476, 2000.

[10] I.M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.

[11] N. E. Karoui. Tracy-widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, pages 663–714, 2007.

[12] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[13] F. Merlevède, J. Najim, and P. Tian. Unbounded largest eigenvalue of large sample covariance matrices: Asymptotics, fluctuations and applications. *Linear Algebra and its Applications*, 2019.

[14] V. Pipiras and M. S Taqqu. *Long-range dependence and self-similarity*, volume 45. Cambridge university press, 2017.

[15] S. Serra. On the extreme eigenvalues of hermitian (block) toeplitz matrices. *Linear algebra and its applications*, 270(1):109–129, 1998.

[16] J.W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2):331–339, 1995.

[17] J.W. Silverstein and Z. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate analysis*, 54(2):175–192, 1995.

[18] E. Terreaux, J.P. Ovarlez, and F. Pascal. Robust model order selection in large dimensional elliptically symmetric noise. *arXiv preprint arXiv:1710.06735*, 2017.

[19] P. Tian. Joint CLT for top eigenvalues of sample covariance matrices of separable high dimensional long memory processes. *arXiv:1906.00909*, 2019.

[20] J. Vinogradova, R. Couillet, and W. Hachem. Estimation of toeplitz covariance matrices in large dimensional regime with application to source detection. *IEEE Trans. Signal Processing*, 63(18):4903–4913, 2015.

[21] J. Yao, S. Zheng, and Z. Bai. *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2015.

[22] L. Zhang. *Spectral analysis of large dimensional random matrices*. PhD thesis, National University of Singapore, 2006.

[23] A. Zygmund. *Trigonometric series*, volume 2. Cambridge university press, 2002.

PENG TIAN, DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, THE UNIVERSITY OF HONG KONG, POKFULAM ROAD, HONG KONG.

E-mail address: tianpeng83@gmail.com

JIANFENG YAO, DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, THE UNIVERSITY OF HONG KONG, POKFULAM ROAD, HONG KONG.

E-mail address: jeffyao@hku.hk