Electronic Journal of Probability

Electron. J. Probab. **31** (2026), article no. 2, 1–30.
ISSN: 1083-6489 https://doi.org/10.1214/25-EJP1462

The smallest singular value of large random rectangular Toeplitz and circulant matrices

Alexei Onatski[*] Vladislav Kargin[†]

Abstract

Let $\{x_i\}_{i\in\mathbb{Z}}$ be a sequence of i.i.d. standard normal random variables. We consider two classes of rectangular random matrices: **Toeplitz** $\mathbf{X} = \left(x_{j-i}\right)_{1\le i\le p, 1\le j\le n}$ and **circulant** $\mathbf{X} = \left(x_{(j-i) \mod n}\right)_{1\le i\le p, 1\le j\le n}$. We let $p, n \to \infty$ so that $p/n \to c \in (0, 1]$. Our main result is that, for any $c \in (0, 1]$, the smallest eigenvalue of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$ converges to zero both in probability and in expectation. We further establish a lower bound on the rate of this convergence, showing it is faster than any poly-logarithmic rate yet slower than any polynomial rate. In the "rectangular circulant" setting, we additionally derive a simple, explicit upper bound on the convergence rate in terms of c. This upper bound implies that, for sufficiently small c, the decay of the smallest eigenvalue to zero can be made arbitrarily slow on a polynomial scale.

Keywords: random rectangular Toeplitz matrix; smallest singular value.
MSC2020 subject classifications: 60B20; 60F05.
Submitted to EJP on April 2, 2025, final version accepted on December 18, 2025.

1 Introduction

Let x_i, $i \in \mathbb{Z}$ be a sequence of i.i.d. standard normal random variables. Consider a $p \times n$ random Toeplitz matrix with $n \ge p$

$$\mathbf{X} = \begin{pmatrix} x_0 & x_1 & x_2 & & \cdots & x_{n-2} & x_{n-1} \\ x_{-1} & x_0 & x_1 & \ddots & & x_{n-3} & x_{n-2} \\ x_{-2} & x_{-1} & x_0 & \ddots & & x_{n-4} & x_{n-3} \\ \vdots & \ddots & \ddots & \ddots & & \vdots & \vdots \\ x_{1-p} & x_{2-p} & \cdots & & & x_{n-1-p} & x_{n-p} \end{pmatrix}. \tag{1.1}$$

The (i, j)-th entry of $\mathbf{S} := \frac{1}{n}\mathbf{X}\mathbf{X}^\top$ can be viewed as a sample autocovariance at lag $|i - j|$ based on the sample $x_{1-p}, \ldots, x_{n-1}$. If only the observations with non-negative

[*]University of Cambridge, United Kingdom. E-mail: ao319@cam.ac.uk
[†]Binghamton University, United States of America. E-mail: vkargin@binghamton.edu

indices are available, one may replace $x_{1-p}, x_{2-p}, \ldots, x_{-1}$ in the above definition of $\mathbf{X}$ by $x_{n+1-p}, x_{n+2-p}, \ldots, x_{n-1}$, respectively, to obtain a $p \times n$ "rectangular circulant" matrix

$$\mathbf{X} = \begin{pmatrix} x_0 & x_1 & x_2 & & \cdots & x_{n-2} & x_{n-1} \\ x_{n-1} & x_0 & x_1 & \ddots & & x_{n-3} & x_{n-2} \\ x_{n-2} & x_{n-1} & x_0 & \ddots & & x_{n-4} & x_{n-3} \\ \vdots & \ddots & \ddots & \ddots & & \vdots & \vdots \\ x_{n+1-p} & x_{n+2-p} & \cdots & & & x_{n-1-p} & x_{n-p} \end{pmatrix}. \tag{1.2}$$

Then the entries of $\mathbf{S}$ become the so called circular sample autocovariances [5, sec. 6.5.2].

The sample autocovariance and circular autocovariance matrices and their inverses are fundamental objects in time series analysis. The norm of $\mathbf{S}^{-1}$, and hence the smallest eigenvalue of $\mathbf{S}$, plays a key role in the study of autoregressive spectral density estimates and optimal prediction (see, for example, [8], [9, 10], [44], [26, 27]).

Random Toeplitz and circulant matrices are also critical in the randomized preprocessing of linear systems of equations ([39]) and in compressed sensing ([41]). In these applications, the smallest eigenvalue of $\mathbf{S}$ is of cental importance, as it governs the condition number, thereby directly affecting the numerical stability and accuracy of various computational methods.

This paper shows that the smallest eigenvalue of $\mathbf{S}$, for both the Toeplitz and circulant cases, converges to zero in probability and in expectation when $p, n \to \infty$ so that $p/n \to c \in (0, 1]$.

This convergence holds regardless of how small $c > 0$ is, in sharp contrast to the standard sample covariance setting—where all elements of $\mathbf{X}$ are i.i.d.—and in which, by a well-known result (see [45]), the smallest eigenvalue of $\mathbf{S}$ almost surely converges to $(1 - \sqrt{c})^2$ whenever $c \leq 1$. As $c \to 0$, that limit approaches 1, the population variance of x_i.

Note that, similar to the case of sample covariance matrices, any two rows of a random Toeplitz or circulant matrix $\mathbf{X}$ are nearly orthogonal. Therefore, the fact that the smallest eigenvalue converges to zero *for any* $c \in (0, 1]$ may come as a surprise, as it did to us.

Moreover, we establish that the convergence of the smallest eigenvalue to zero is faster than any poly-logarithmic rate in n. In the "rectangular circulant" case, we show that, conversely, the convergence is slower than any polynomial rate in n whenever c is sufficiently small. This extremely slow convergence makes it difficult to detect or even intuitively guess the limiting behavior via Monte Carlo experiments.

To illustrate this difficulty, Figure 1 shows MC results for a random circulant matrix (1.2) with $p = 100, n = 1,000$ (top panel), and for $p = 1,000, n = 10,000$ (bottom panel). Results for a random Toeplitz matrix (1.1) are very similar, and we do not report them here. The left panel plots the average over $10,000$ MC replications of the eigenvalues of $\mathbf{S}$, starting from the largest (order number 1 on the horizontal axis) and ending with the smallest (order number p). The right panel shows the histogram of the $10,000$ MC realizations of the smallest eigenvalue.

We see that the smallest eigenvalue, λ_p, remains relatively insensitive to the ten-fold increase in the values of p and n. The histogram of its MC replications concentrates and rather sluggishly shifts to the left. In fact, the left boundary of the histogram (more precisely, the smallest MC replication of λ_p) *increases* from 0.1582 to 0.1595 when the values of p and n increase in our simulation experiment. Arguably, one may easily misinterpret such MC results as suggesting a convergence of the smallest eigenvalue to some positive value.

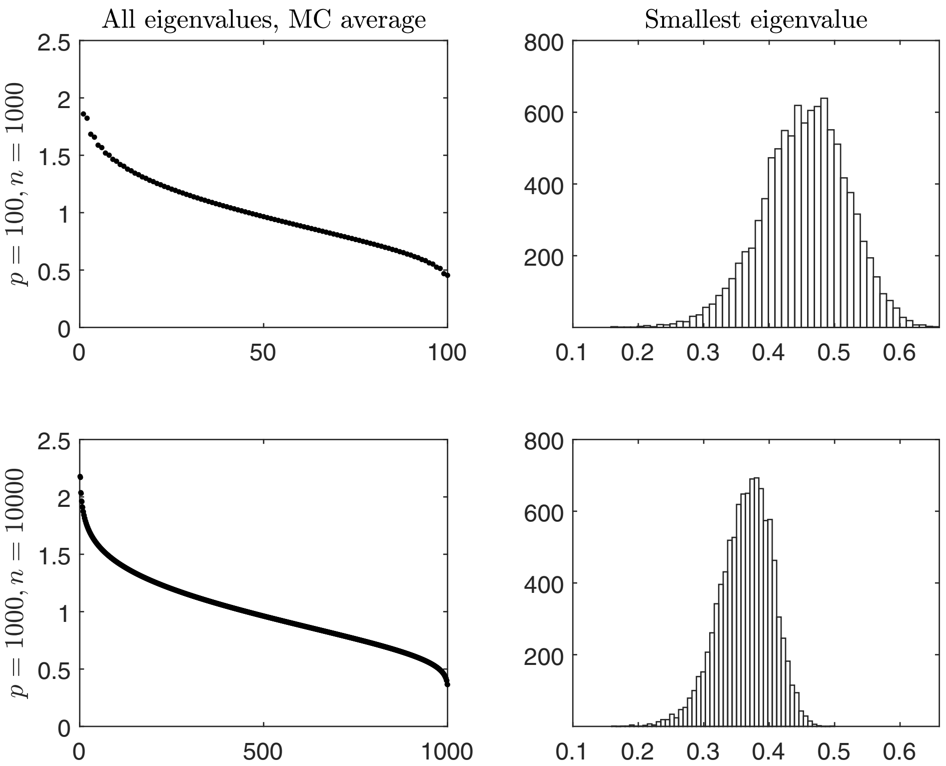

Figure 1: Monte Carlo results for random circulant matrix (1.2). The number of MC replications is 10,000. Top panel: $p = 100, n = 1,000$. Bottom panel: $p = 1,000, n = 10,000$. Left panel: scatter plot showing MC average of the ordered eigenvalues of $\mathbf{S}$. Right panel: histogram of MC realizations of the smallest eigenvalue.

Interestingly, the Monte Carlo analysis in [13] is employed to illustrate their theoretical findings on the existence of the limiting spectral distribution of $\mathbf{S}$. (See Theorems 1 and 4 in [13] for the Toeplitz and circulant cases, respectively.) Their results suggest that, for the Toeplitz case with $c = 1/3$, the support of the limiting distribution is bounded away from zero. Inspired by this evidence, [13] pose a question regarding the positivity of the infimum of the support and its precise value.

Our finding that λ_p converges to zero casts doubt on the hypothesis that the support of the limiting spectral distribution is separated from zero when $c < 1$. However, it is not inconsistent with this hypothesis. Conceivably, $o(p)$ of the relatively small eigenvalues could remain below a positive lower bound of the support of the limiting spectral distribution, so that the smallest eigenvalue does converge to zero and the hypothesis still holds. We leave proving or rejecting the hypothesis for future study.

There have been many previous studies of the asymptotic behavior of the singular values of $\mathbf{X}/\sqrt{n}$ (square roots of the eigenvalues of $\mathbf{S}$) for random Toeplitz and circulant[1] $\mathbf{X}$ under various assumptions on the generating sequence x_i. These studies can be roughly categorized into four strands.

The first, most numerous, strand consists of the research on the asymptotic behavior of the empirical distribution of the singular values. For some, but not all, important works following this direction see [14], [19], [25], [13], [30], [17], and [34]. The second

[1] For a book-long treatment of "patterned" random matrices, including Toeplitz and circulant matrices, see [12].

strand focuses on the fluctuations of the linear spectral statistics, that is, of the averages of some regular test functions of the singular values. We would like to mention, not pretending to be thorough, [20], [31], [28], and [3, 4]. The third group is represented by studies of the largest singular value. This group includes, but is not limited to [36], [16], [1, 2], and [43].

The most relevant for this paper is the fourth group, which constitutes the previous literature on the smallest eigenvalue of $\mathbf{S}$ for the random Toeplitz and circulant $\mathbf{X}$. This literature has been focusing on the square case $p = n$, hence $c = 1$. Meckes in Corollary 5 of [37] shows that the smallest eigenvalue of $\mathbf{S}$ for a (complex) Gaussian circulant matrix $\mathbf{X}$ is distributed as an exponential random variable with mean $1/n$. Pan et al. [40] derive probabilistic bounds on the smallest eigenvalue of $\mathbf{S}$ for Toeplitz matrices with i.i.d. Gaussian x_i. When x_i are i.i.d. and satisfy the Lyapunov condition, but not necessarily Gaussian, Barrera and Manrique-Mirón [7] prove that the smallest singular value of the circulant $\mathbf{X}$ (which equals the square root of the product of n and the smallest eigenvalue of $\mathbf{S}$ for the circulant case) is asymptotically distributed according to the Rayleigh distribution with c.d.f. $R(x) = 1 - \exp\{-x^2/2\}, x \geq 0$. This, of course, implies that the smallest eigenvalue of $\mathbf{S}$ itself converges to zero in probability. Manrique-Mirón [32] derives an upper bound on the probability that the smallest singular value of the circulant $\mathbf{X}$ belongs to a shrinking neighborhood of zero. Manrique-Mirón [33] studies the smallest singular value of a random tridiagonal Toeplitz matrix.

Note that when $p = n$, the smallest eigenvalue of $\mathbf{S}$ converges to zero even in the standard case of sample covariance matrices when all elements of $\mathbf{X}$ are i.i.d. For readers interested in the second-order asymptotic behavior of the smallest eigenvalue in this scenario, we refer to [48].

As explained in [11], obtaining upper bounds on the smallest singular value of a square matrix of the form $\mathbf{X} - zI$, with $z \in \mathbb{C}$, is a crucial step in establishing the limiting empirical distribution of the eigenvalues of a non-Hermitian matrix $\mathbf{X}$ (see, in particular, their Lemma 4.12). By analogy, deriving similar upper bounds on the rate at which λ_p converges to zero for square Toeplitz matrices $\mathbf{X}$ may prove useful for studying the limiting empirical distribution of their eigenvalues—a direction that we leave for future research. In the present paper, we obtain such upper bounds only for circulant matrices. The limiting empirical distribution of the eigenvalues of square non-symmetric circulant matrices is, in fact, known to be complex normal under mild moment conditions on the entries (see, e.g., Theorem 3.2.1 in [15]).

To our knowledge, our paper is the first to analyze the asymptotic behavior of λ_p for rectangular Toeplitz or circulant $\mathbf{X}$ with $c < 1$.

The remainder of the paper is organized as follows. The next section presents our main results (Theorems 2.1-2.6) and outlines their proofs. Section 3 reports results of the Monte Carlo analysis. Section 4 contains further details of the proofs of Theorems 2.1-2.6. Section 5 concludes. Proofs of technical lemmas are given in the Appendix.

2 Main results and proof outlines

In this paper, we consider the asymptotic regime where $p, n \to \infty$. We assume that $n := n(p)$ is a function of p such that as $p \to \infty$, $p/n \to c \in (0, 1]$. We will abbreviate this as $p, n \to_c \infty$. Recall that λ_p denotes the smallest eigenvalue of $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^\top$, where $\mathbf{X}$ is either a random Toeplitz matrix (1.1) or a random "rectangular circulant" matrix (1.2). In what follows, we will omit the quotation marks around the term "rectangular circulant", which will always correspond to matrices of type (1.2).

Theorem 2.1 (Convergence to zero)**.** *Suppose that* $p, n \to_c \infty$*, where* $\lim p/n = c \in (0, 1]$*. Then, for both Toeplitz and circulant cases, there exists a constant* $\beta > 0$ *that may* *depend on c, such that*

$$\lambda_p = o_{\mathrm{P}}\left(\exp\left\{-\beta \log^{1/3}(p)\right\}\right). \tag{2.1}$$

Remark 2.2. For a deterministic sequence $a_p > 0$, the relation $X_p = o_{\mathrm{P}}(a_p)$ means X_p/a_p converges to zero in probability, $X_p/a_p \xrightarrow{\mathrm{P}} 0$; equivalently, for every $\varepsilon > 0$, $\Pr\{|X_p| > \varepsilon a_p\} \to 0$ as $p \to \infty$. Here $e^{-\beta(\log p)^{1/3}} \to 0$ at a *subpolynomial* rate: it decays slower than any $p^{-\alpha}$ ($\alpha > 0$) but faster than any $(\log p)^{-k}$ ($k \in \mathbb{N}$).

For any rectangular $p \times n$ matrix $\mathbf{X}$ with $p > n$, we have $\lambda_p = 0$. Thus the convergence $\lambda_p \to 0$ is trivial for $c > 1$. Consequently, the definitions (1.1)–(1.2) and all our results focus on the nontrivial regime $c \in (0, 1]$.

Random rectangular Hankel matrices $\mathbf{X} = (x_{i+j})_{1 \le i \le p,\ 1 \le j \le n}$ and reverse circulant matrices $\mathbf{X} = (x_{(i+j) \bmod n})_{1 \le i \le p,\ 1 \le j \le n}$ can be obtained from the corresponding Toeplitz and circulant matrices, respectively, by reversing the order of columns and appropriately adjusting the indices of x_i. Such a column reordering does not affect the singular values of the matrix. Consequently, we obtain the following corollary.

Corollary 2.3. *Theorem 2.1 holds not only for Toeplitz and circulant matrices $\mathbf{X}$ but also for Hankel and reverse circulant matrices $\mathbf{X}$.*

The following theorem shows that λ_p converges to zero not only in probability, but also in expectation.

Theorem 2.4. *Under assumptions of Theorem 2.1, for every $\kappa \ge 1$, $\mathbb{E}(\lambda_p)^\kappa \to 0$.*

Function $\exp\left\{-\beta \log^{1/3}(p)\right\}$ in (2.1) converges to zero faster than $\log^{-\alpha} p$ but slower than $p^{-\alpha}$ for any $\alpha > 0$. Hence, Theorem 2.1 implies that λ_p converges to zero in probability faster than any inverse power of logarithm. Of course, the theorem does not give us any upper bound on the convergence rate because even the function which is identically equal to zero (instant convergence) can be classified as $o_{\mathrm{P}}\left(\exp\left\{-\beta \log^{1/3}(p)\right\}\right)$.

We can upper bound the convergence rate in probability (polynomially) only in the circulant (and reverse circulant) setting. Here, by an upper bound on the rate of convergence we mean an upper bound on the decay exponent $R_p = -\log \lambda_p / \log p$. Thus, statements of the form $\lambda_p > p^{-\gamma}$ provide such upper bounds, $\limsup R_p \le \gamma$, with high probability.

Theorem 2.5. *[Lower bound for λ_p, Regime I] Suppose that $\mathbf{X}$ is a random rectangular circulant matrix* (1.2)*. Let $m \ge 2$ be a fixed integer, and suppose that $p, n \to_c \infty$ with $\lim p/n = c < \frac{1}{\pi(m-1)}$. Then for any small $\epsilon > 0$, we have*

$$\Pr\left\{\lambda_p > p^{-\frac{1}{m} - \epsilon}\right\} \to 1.$$

Theorem 2.6 (Lower bound for λ_p, Regimes II and III)**.** *Under assumptions of Theorem 2.5, if $\lim p/n = c \le \frac{1-\delta}{m}$, where $\delta > 0$ is an arbitrarily small fixed real number, then for any $\epsilon > 0$, we have*

$$\Pr\left\{\lambda_p > p^{-\frac{1}{2} - \frac{1}{2m} - \epsilon}\right\} \to 1.$$

For $\lim p/n = c \in [1/2, 1]$, $\Pr\left\{\lambda_p > p^{-1-\epsilon}\right\} \to 1$.

Theorem 2.5 implies that, for sufficiently small c, the rate of convergence of λ_p to zero is arbitrarily polynomially slow. Specifically, $\lambda_p \xrightarrow{P} 0$ slower than $p^{-\alpha}$, where $\alpha = \frac{\pi c}{\pi c + 1} + \epsilon$ and $\epsilon > 0$ is an arbitrarily small constant. However, this theorem provides no upper bound on the convergence rate for $c \ge \pi^{-1}$.

Theorem 2.6 uses different proof ideas to cover the case $\pi^{-1} \le c < 1/2$ at the expense of a much worse upper bound for smaller c. The statement of the theorem for $c \in [1/2, 1]$ follows by a simple embedding argument from the result of [7] about the convergence of $\sqrt{n\lambda_p}$ to the Rayleigh distribution in the square circulant case, mentioned in the Introduction. This argument holds for all $0 < c \le 1$ but for $c < 1/2$ it gives a worse rate than in our result.

As shown in [32, Theorem 2.6], if $\mathbf{X}$ is a square circulant matrix with i.i.d. entries x_i having a finite second moment, then for any fixed $\epsilon > 0$ and $\rho \in (0, 1/4)$ we have

$$\Pr\{\lambda_p \le \epsilon^2 p^{-1-2\rho}\} \le C\left(\frac{\epsilon^2+\epsilon}{p^{2\rho}} + \frac{1}{p^{1/2-o(1)}}\right), \tag{2.2}$$

where the constant $C > 0$ depends only on the distribution of the x_i. In particular, this result implies that the conclusion of Theorem 2.6 for $c \in [1/2, 1]$ continues to hold when the x_i are not Gaussian but merely have a finite second moment.

Moreover, their result yields the same lower bound for every $0 < c \le 1$ for such non-Gaussian matrices.

A more recent work, [33], establishes the limiting distribution of the minimum of the absolute values of the eigenvalues of a random tridiagonal Toeplitz matrix. Both this analysis and that of [32], which leads to (2.2), rely on explicit representations of the eigenvalues of $\mathbf{X}$ in terms of its entries x_i. However, no such representations are currently known for the singular values of rectangular matrices $\mathbf{X}$. Consequently, the techniques developed in [32, 33] cannot be directly applied to our Theorems 2.1, 2.4–2.6.

Proof outlines. Here we introduce a setup and outline the proofs of the above theorems. Further details of the proofs can be found in Sections 4.1-4.4.

Consider a $p\times n$ rectangular circulant matrix $\mathbf{X}$ as defined in (1.2). Let $\widetilde{\mathbf{X}}$ be a Toeplitz matrix consisting of the last $\widetilde{n} := n-p+1$ columns of $\mathbf{X}$. It follows that $\mathbf{X}\mathbf{X}^\top - \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top \ge 0$, implying that the smallest eigenvalue of $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ is less than or equal to the smallest eigenvalue of $\mathbf{X}\mathbf{X}^\top$. Therefore, to establish Theorems 2.1 and 2.4 for $p \times \tilde{n}$ Toeplitz matrices, it suffices to prove them for $p \times n$ circulant matrices. Consequently, throughout our proofs, we assume that $\mathbf{X}$ is a rectangular circulant matrix.

About Theorem 2.1: The key element of the proof of Theorem 2.1 is an upper bound on λ_p, expressed in the form of a weighted average of the periodogram of $x_0, \ldots, x_{n-1}$. The variational characterization of the smallest eigenvalue yields

$$\lambda_p \le \frac{1}{n}\mathbf{u}^*\mathbf{X}\mathbf{X}^\top\mathbf{u} \tag{2.3}$$

for any complex p-dimensional vector $\mathbf{u} := (u_0, u_1, \ldots, u_{p-1})^\top$ with Euclidean norm 1. Let $\mathbf{X}_C$ be the $n \times n$ circulant matrix with the first p rows constituting $\mathbf{X}$, and let $\mathbf{u}_C$ be an n-dimensional vector $\mathbf{u}_C := (u_0, u_1, \ldots, u_{p-1}, 0, \ldots, 0)^\top$. Then obviously

$$\mathbf{u}^*\mathbf{X}\mathbf{X}^\top\mathbf{u} = \mathbf{u}_C^*\mathbf{X}_C\mathbf{X}_C^\top\mathbf{u}_C. \tag{2.4}$$

Since $\mathbf{X}_C^\top$ is an $n \times n$ real circulant matrix, it admits the spectral decomposition[2]

$$\mathbf{X}_C^\top = \sqrt{n}\mathcal{F}^* \operatorname{diag}(\mathcal{F}\mathbf{x})\,\mathcal{F}, \tag{2.5}$$

where $\mathcal{F}$ is the (scaled) unitary $n \times n$ Discrete Fourier Transform (DFT) matrix with elements[3] $\mathcal{F}_{s,t} = \exp\{\frac{2\pi \mathrm{i}}{n}st\}/\sqrt{n}$, and $\mathbf{x} := (x_0, x_1, x_2, \ldots, x_{n-1})^\top$ is the first column of $\mathbf{X}_C^\top$. Let $\mathbf{y} := \mathcal{F}\mathbf{x}$ be the (scaled) DFT of $\mathbf{x}$ and let $\mathbf{P_u} := \sqrt{n}\mathcal{F}\mathbf{u}_C$ be the DFT of $\mathbf{u}_C$. Then using (2.5) and (2.4) in (2.3), we obtain

$$\lambda_p \le \frac{1}{n}\mathbf{u}_C^*\mathbf{X}_C\mathbf{X}_C^\top\mathbf{u}_C = \frac{1}{n}\|\mathbf{y}\odot\mathbf{P_u}\|^2 = \frac{1}{n}\sum_{s=0}^{n-1}|y_s|^2|P_{\mathbf{u}s}|^2, \tag{2.6}$$

[2] See e.g. [23], p. 202.

[3] We start indexing form 0, so that the element in the first row and the first column of $\mathcal{F}$ is denoted as $\mathcal{F}_{0,0}$.

where $\odot$ denotes the element-wise Hadamard product, and $y_s, P_{\mathbf{u}s}$ are the s-th elements of $\mathbf{y}$ and $\mathbf{P_u}$. In the proof, we will seek a $\mathbf{P_u}$ that ensures the right-hand side of (2.6) is small with high probability.

In the sum in (2.6), $|y_s|^2$ can be interpreted as the periodogram of $\mathbf{x}$ at the fundamental frequencies $\omega_s := 2\pi s/n$, while $P_{\mathbf{u}s} = P_{\mathbf{u}}\left(e^{\mathrm{i}\omega_s}\right)$, where

$$P_{\mathbf{u}}(z) := \sum_{j=0}^{p-1} u_j z^j$$

is a polynomial with the coefficients given by the elements of $\mathbf{u}$. Hence, (2.6) bounds λ_p by a weighted average of the periodogram of $\mathbf{x}$ with weights determined by the magnitude of the polynomial $P_{\mathbf{u}}(z)$ of degree $p-1$ on a grid of size n on the unit circle.

Since $\mathcal{F}$ is unitary and $\mathbf{x}$ is normally distributed, $\mathbf{y} = \mathcal{F}\mathbf{x}$ has a complex normal distribution. Precisely,

$$\mathbf{y} = \begin{cases} \left(y_0,\, y_1,\, \ldots,\, y_{\frac{n-1}{2}},\, \overline{y_{\frac{n-1}{2}}},\, \ldots,\, \overline{y_1}\right)^\top, & \text{if } n \text{ is odd,} \\ \left(y_0,\, y_1,\, \ldots,\, y_{\frac{n}{2}-1},\, y_{\frac{n}{2}},\, \overline{y_{\frac{n}{2}-1}},\, \ldots,\, \overline{y_1}\right)^\top, & \text{if } n \text{ is even.} \end{cases} \tag{2.7}$$

where $y_0, y_{n/2}$ are i.i.d. $\mathcal{N}(0,1)$ and $\{y_i, 0 < i < n/2\}$ are independent from them i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ (that is, random variables of form $a_1 + \mathrm{i}a_2$, where a_1 and a_2 are independent $\mathcal{N}(0,1/2)$). In particular, for any $0 < s < t < n/2$, $|y_s|^2$ and $|y_t|^2$ are independent random variables having exponential distribution $\mathrm{Exp}(1)$. We also note that the facts that $\mathcal{F}$ is unitary and $\|\mathbf{u}\| = 1$ imply that $\frac{1}{n}\sum_{s=0}^{n-1}|P_{\mathbf{u}s}|^2 = 1$.

If we choose $\mathbf{u}$ so that $|P_{\mathbf{u}s}|^2$ for all s are roughly of the same order of magnitude, the law of large numbers would imply that the right hand side of (2.6) is $O_P(1)$, which would not be helpful as a bound for λ_p. Hence, we will aim at choosing $\mathbf{u}$ so that $|P_{\mathbf{u}s}|^2$ very quickly decays when $\min\{s, n-s\}$ grows.

As follows from the DFT interpretation of $\mathbf{P_u}$, if instead of vector $\mathbf{u}$ with coordinates u_s, $s = 0, \ldots, p-1$ we consider vector $\mathbf{u}_\tau$ with coordinates $\exp\{-\omega_\tau s\mathrm{i}\}u_s$, $s = 0, \ldots, p-1$, the corresponding $P_{\mathbf{u}_\tau,s}$ would satisfy the identity $P_{\mathbf{u}_\tau,s} = P_{\mathbf{u},(s-\tau) \bmod n}$. In what follows, we will omit notation $\bmod n$ from any coordinate indices of n-dimensional vectors. For example, we simply write

$$P_{\mathbf{u}_\tau,s} = P_{\mathbf{u},s-\tau}.$$

Pickings several integers $\tau_1, \ldots, \tau_K$, defining $\mathbf{u}_k := \mathbf{u}_{\tau_k}$, and using the above identity, we obtain the following stronger version of bound (2.6):

$$\lambda_p \le \min_{k=1,\ldots,K} \frac{1}{n}\sum_{s=0}^{n-1} |y_s|^2 |P_{\mathbf{u},s-\tau_k}|^2. \tag{2.8}$$

Since by construction the coordinates of $\mathbf{P_u}$ will be quickly decaying, picking integers $\tau_1, \ldots, \tau_K$ so that they are sufficiently far apart will ensure that the different sums under the minimum are almost independent, hence it will be relatively easy to evaluate the minimum.

Recall that $\mathbf{P_u}$ is the DFT of vector $\mathbf{u}_C$, that "adds" $n-p$ zero coordinates to $\mathbf{u}$. If we interpret this vector as a signal "localized in time" in the sense that only the first p observations (corresponding to the entries of vector $\mathbf{u}$) are non-zero, then the task of choosing $\mathbf{u}$ so that the coordinates of $\mathbf{P_u}$ quickly decay is similar to the problem of choosing a time-localized signal that has the DFT which is well concentrated in the frequency domain. There exists a large literature on this topic, see [46], [22], [6] and references therein.

In Section 4.1, we follow [6] and consider the Gaussian Fourier pair,[4] which is well localized in both time and frequency domain. This choice will imply tight bounds on $|P_{\mathbf{u},s-\tau_k}|^2$ in (2.8) yielding the convergence of λ_p to zero with the speed stated in Theorem 2.1.

About Theorem 2.4: Theorem 2.1 and the continuous mapping theorem imply that $(\lambda_p)^\kappa \xrightarrow{\mathrm{P}} 0$ as $p \to \infty$. To prove Theorem 2.4, it is sufficient to show that $\mathbb{E}(\lambda_p)^{\kappa+\varepsilon} \le C < \infty$ for some $\varepsilon > 0$, where C is a constant that does not depend on p. Indeed, this implies that $(\lambda_p)^\kappa$ is uniformly integrable and therefore the convergence in probability implies the convergence in expectation (Theorem 5.5.4 in [24]). In Section 4.2, we prove that

$$\mathbb{E}\lambda_p^\kappa \le C_{\kappa,c} < \infty \qquad \text{for every} \quad \kappa \ge 1 \tag{2.9}$$

with a constant $C_{\kappa,c}$ that may depend on κ and $c := \lim p/n$. Our detailed proof in Section 4.2 relies on the inequality (2.6) and inequalities for sub-exponential random variables (see sections 2.7 and 2.8 of [49]).

About Theorem 2.5: The starting point of our proof of Theorem 2.5 is the fact that

$$\lambda_p = \sum_{j=0}^{n-1} |y_j|^2 \left(\frac{1}{n} |P_{\mathbf{u}}(e^{\mathrm{i}\omega_j})|^2 \right)$$

for some real $\mathbf{u}$ with $\|\mathbf{u}\| = 1$. We aim to show that the right-hand side of the above equality is greater than $p^{-\frac{1}{m}-\epsilon}$ *for all* real $\mathbf{u}$ with $\|\mathbf{u}\| = 1$, for all sufficiently large p on a sequence of events A_p such that $\Pr A_p \to 1$.

Note that $\frac{1}{n}|P_{\mathbf{u}}(e^{\mathrm{i}\omega})|^2$ is a cosine trigonometric polynomial of order $p-1$. We denote it as $T_{p-1}(\omega)$, suppressing its dependence on $\mathbf{u}$. Since $\|\mathbf{u}\| = 1$, we have

$$\int_0^{2\pi} T_{p-1}(\omega)\mathrm{d}\omega = \int_0^{2\pi} \frac{1}{n}|P_{\mathbf{u}}(e^{\mathrm{i}\omega})|^2 \mathrm{d}\omega = \frac{2\pi}{n} \tag{2.10}$$

and, assuming for concreteness that n is odd (the case of even n requires a similar analysis, and we omit it) and letting $N := (n+1)/2$,

$$T_{p-1}(0) + 2\sum_{j=1}^{N-1} T_{p-1}(\omega_j) = \sum_{j=0}^{n-1} \frac{1}{n}|P_{\mathbf{u}}(e^{\mathrm{i}\omega_j})|^2 = 1. \tag{2.11}$$

Further, define

$$\xi_0 := |y_0|^2, \qquad \xi_j := |y_j|^2 + |y_{n-j}|^2 = 2|y_j|^2, \quad j = 1, \dots, N-1.$$

With this notation, we have

$$\begin{aligned} \lambda_p = \sum_{j=0}^{N-1} \xi_j T_{p-1}(\omega_j) \quad &= \quad \xi_0 T_{p-1}(0) + \sum_{j=1}^{N-1} \xi_{(j)} T_{p-1}(\omega_{\sigma(j)}) \\ &\ge \quad \frac{\xi_0}{2} T_{p-1}(0) + \xi_{(r)} \sum_{j=r+1}^{N-1} T_{p-1}(\omega_{\sigma(j)}), \end{aligned} \tag{2.12}$$

where $\xi_{(j)}$ with $j \ge 1$ is the j-th order statistic for the sequence $\xi_1, \dots, \xi_{N-1}$, σ is the (random) permutation of the indices such that $\xi_{(j)} = \xi_{\sigma(j)}$, and $r := \lfloor p^{1-1/m-\epsilon/2} \rfloor$.

We will assume that $\min\{\xi_0, \xi_{(r)}\} > r/N$ on events A_p. We can make such an assumption because

$$\Pr\left(\min\{\xi_0, \xi_{(r)}\} > r/N\right) \to 1 \tag{2.13}$$

[4][6] also considers the so-called Kaiser-Bessel Fourier pair. It turns out that in our setting, it leads to the same results, so we do not consider it here.

as $p \to \infty$. Section 4.3 contains a quick proof of (2.13). Therefore, (2.12) implies that on A_p,

$$\lambda_p > \frac{r}{N}\left(\frac{1}{2}T_{p-1}(0) + \sum_{j=r+1}^{N-1} T_{p-1}(\omega_{\sigma(j)})\right) =: \frac{r}{N}S_1. \tag{2.14}$$

Since $\frac{r}{N} \gg p^{-1/m-\epsilon}$, Theorem 2.5 would follow if we show that S_1 is bounded away from zero over all non-negative trigonometric polynomials T_{p-1}, satisfying (2.10) and (2.11).

The main idea of our proof is that a small value of S_1 yields a large total variation of $T_{p-1}(\omega)$ on $[0, 2\pi]$, with high probability. Indeed, normalization (2.11) implies that, if values of $T_{p-1}(\omega_{\sigma(j)})$ are small for $j > r$ (so that S_1 is small), they must be large for some $j \le r$. On the other hand, positions of $\omega_{\sigma(j)}$ with $j \le r$ and with $j > r$ in the interval $[0, \pi]$ are randomly mixed. This leads to a large total variation of $T_{p-1}(\omega)$ (see Section 4.3 for details).

As we further show in Section 4.3, such a large total variation of $T_{p-1}(\omega)$ contradicts the integral version of the Bernstein inequality for the derivative of trigonometric polynomials (see e.g. Theorem 2.5 in Chapter 4 of [21]). Hence, S_1 cannot be small, which concludes our proof of Theorem 2.5.

About Theorem 2.6: Unfortunately, the Bernstein inequality for trigonometric polynomials does not allow us to obtain a non-trivial upper bound on the convergence rate of λ_p for p, n such that $\lim p/n \ge 1/\pi$. To cover this case we use Lemma 4.4 (due to [47], see Section 4.4) showing that trigonometric polynomials cannot decay too fast near the point of their maximum.

Using an inequality similar to (2.14), but with $r_1 \approx \sqrt{r}$ replacing r, and assuming that the right-hand side of such an inequality is small, Section 4.4 shows that the maximum of the corresponding trigonometric polynomial on $[0, \pi]$ must be relatively large. Since, by Lemma 4.4, the trigonometric polynomial must remain large in a relatively large neighborhood of the maximum point, such a neighborhood will include points $\omega_{\sigma(j)}$ with $j > r$, with high probability. The original inequality (2.14) then would imply that λ_p cannot be too small.

Balancing the lower bounds on λ_p provided by the modified and the original inequality (2.14) gives us the upper bound on the rate of the convergence of λ_p to zero, stated by Theorem 2.6.

Discussion. Theorems 2.1 and 2.4–2.6 remain valid without modification when $x_i, i \in \mathbb{Z}$ is a sequence of i.i.d. standard complex normal random variables, rather than real normal ones. The main difference from the real normal case is that the entries of the vector $\mathbf{y} = \mathcal{F}\mathbf{x}$ become i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$ instead of having the more involved dependence structure described by (2.7). This requires only minor adjustments to the proofs, which we outline in the last section of the Appendix.

Extending our results to non-Gaussian i.i.d. sequences x_i, $i \in \mathbb{Z}$ satisfying suitable moment conditions would require substantial additional work. Much of our analysis relies on properties of weighted sums $\sum_j c_j |y_j|^2$, where the variables $|y_j|^2$ are independent standard exponential random variables representing the periodogram of the sequence $(x_0, \ldots, x_{n-1})$ at the fundamental frequencies $\omega_j := 2\pi j/n$.

As discussed in [29], although the periodogram of an i.i.d. sequence evaluated at the fundamental frequencies resembles an i.i.d. sequence of standard exponential variables, this analogy can be misleading for weighted sums. A concrete step toward extending our results to the non-Gaussian setting would be to generalize the accurate lower bound for $\Pr\left\{\sum_j c_j |y_j|^2 \le \epsilon_p\right\}$ for small ϵ_p, derived in Step 1 of the proof of Theorem 2.1 for i.i.d. exponential $|y_j|^2$, to the case where the quantities $|y_j|^2$ correspond to the periodogram of a non-Gaussian i.i.d. sequence.

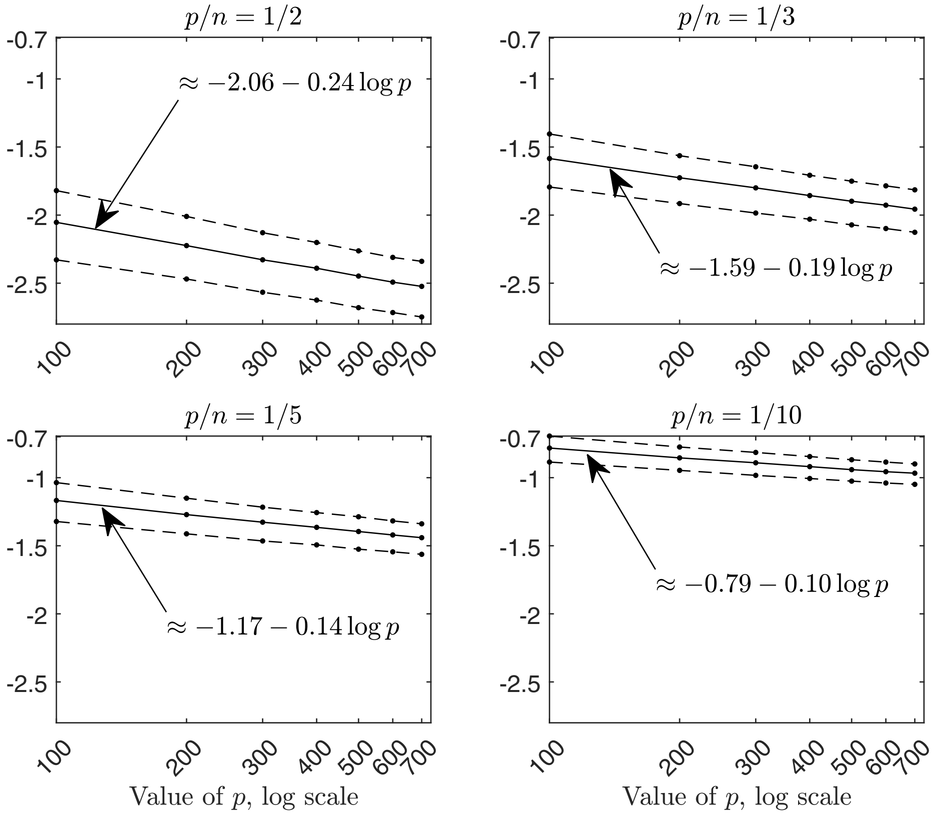

Figure 2: 25, 50, and 75 percentiles of the Monte Carlo distribution of $\log \lambda_p$ for $p \times n$ rectangular circulant $\mathbf{X}$. Based on 10,000 MC replications. Equations reported in the figure correspond to the ordinary least squares estimates of the median line based on seven observations (dot markers).

3 Monte Carlo

Theorems 2.1-2.6 leave a considerable gap between the lower and upper bounds on the rate of convergence of λ_p to zero. In this section, we perform Monte Carlo (MC) analysis to get some idea about the actual rate of the convergence. We simulate $p \times n$ rectangular circulant and Toeplitz matrices $\mathbf{X}$, and compute the smallest eigenvalue λ_p of $\mathbf{X}\mathbf{X}^\top/n$. We make 10,000 MC replications for four different ratios: $p/n = \frac{1}{2}, \frac{1}{3}, \frac{1}{5}$ and $\frac{1}{10}$, and seven different values of p: $100, \ldots, 700$. Our simulations were done in MATLAB on a basic laptop computer. Therefore, allowing for $p > 700$ would have been too time consuming.

Figure 2 reports the logarithm of $k = 25\%, 50\%$, and 75% quantiles, $Q_k(\lambda_p)$, of the Monte Carlo distribution of λ_p for rectangular circulant matrix $\mathbf{X}$. We plot $\log Q_k(\lambda_p)$ against $\log p$. The solid lines correspond to the median ($k = 50\%$), the dashed line above it – to $k = 75\%$, and the dashed line below it – to $k = 25\%$. All the graphs look linear for all values of $p/n = \frac{1}{2}, \frac{1}{3}, \frac{1}{5}$ and $\frac{1}{10}$. The lines representing the graphs for different percentiles k become flatter as p/n decreases. At the same time, the intersections of the lines with the vertical axis become closer to zero. This suggests that the model

$$Q_k(\lambda_p) = \alpha_{c,k} p^{-\theta_c}, \tag{3.1}$$

with some $\alpha_{c,k} > 0$, increasing in both k and $c = \lim p/n$, and $\theta_c > 0$, decreasing in c, might be approximating the asymptotic behavior of the quantiles of λ_p reasonably well.

If so, we would expect a polynomial rate of the convergence of λ_p to zero. Furthermore, this rate becomes slower (θ_c decreases) as c decreases.

We estimate the slope $-\theta_c$ of the solid lines (log of the MC median of λ_p) reported in Figure 2 by ordinary least squares (OLS) regression based on seven observations, corresponding to $p = 100, \ldots, 700$. The estimates are reported next to the graphs. It is interesting to compare the estimates of θ_c, $\hat{\theta}_c$, with the upper bounds on the convergence rates established in Theorem 2.5. For $p/n = \frac{1}{5}$, $\hat{\theta}_c = 0.14$, whereas the smallest upper bound provided by Theorem 2.5 for $p/n = \frac{1}{5}$ equals $(5/\pi + 1)^{-1} + \epsilon \approx 0.39$, which is approximately 2.8 times larger. For $p/n = \frac{1}{10}$, $\hat{\theta}_c = 0.10$, whereas the smallest upper bound provided by Theorem 2.5 for $p/n = \frac{1}{10}$ is $(10/\pi + 1)^{-1} + \epsilon \approx 0.24$, which is 2.4 times larger. Hence, the validity of Theorem 2.5 is supported by our MC results. At the same time, these results suggest that the bounds described by Theorem 2.5 are far from optimal.

For $p/n > 1/\pi$, Theorem 2.5 does not provide us with any bounds, so we need to use Theorem 2.6 instead. For $p/n = \frac{1}{2}$, $\hat{\theta}_c = 0.24$, whereas the smallest upper bound provided by Theorem 2.6 for $p/n = \frac{1}{2}$ is approximately 0.75, which is about 3.1 times larger. For $p/n = \frac{1}{3}$, $\hat{\theta}_c = 0.19$, whereas the smallest upper bound provided by Theorem 2.6 for $p/n = \frac{1}{3}$ is approximately 0.66, which is about 3.5 times larger. Hence, again, MC supports the theoretical findings, but the bounds described by Theorem 2.6 might be suboptimal.

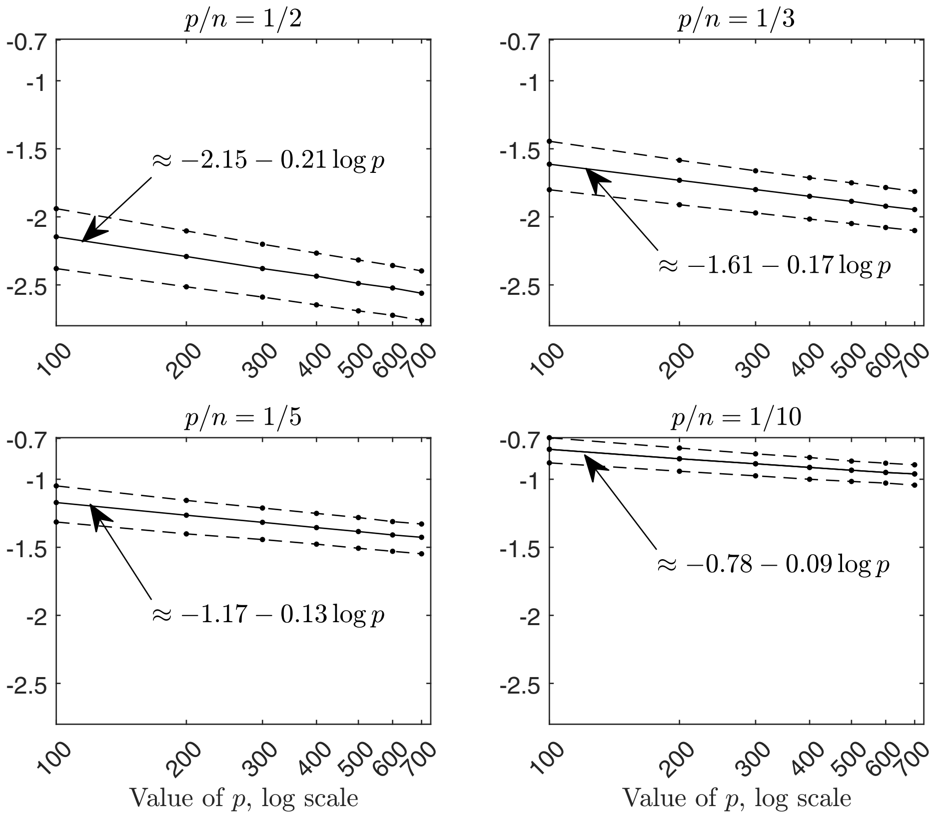

Figure 3: 25, 50, and 75 percentiles of the Monte Carlo distribution of $\log \lambda_p$ for $p \times n$ rectangular Toeplitz $\mathbf{X}$. Based on 10,000 MC replications. Equations reported in the figure correspond to the ordinary least squares estimates of the median line based on seven observations (dot markers).

Figure 3 is the equivalent of Figure 2 for Toeplitz $\mathbf{X}$. The figures are very similar, and the Ordinary Least Squares ("OLS") estimates of the slopes $-\theta_c$ differ only slightly (being smaller in absolute value) from those for rectangular circulant $\mathbf{X}$. This suggests that extrapolating the results of Theorems 2.5 and 2.6 from circulant to Toeplitz matrices would not lead to a substantial error.

4 Proof details

This section contains details of the proofs of Theorems 2.1-2.6, outlined in Section 2.

4.1 Proof details for Theorem 2.1

As explained in Section 2, the proof is based on inequality (2.8). We start by making a choice of vector $\mathbf{u}$. We define $\mathbf{u}$ as a truncated and normalized version of the so-called periodized Gaussian vector, which have good localization properties in both the position and frequency domains. We will need the following lemma. Its proof (see Appendix) is only slightly different from the proof of Proposition 6 in [6].

Lemma 4.1 (Proposition 6, [6]). *Let σ_p^2 be a positive constant that depends only on p, and let $\mathbf{w} = (w_0, ..., w_{n-1})^\top$ be the periodized Gaussian vector with*

$$w_j = (2\pi\sigma_p^2)^{-1/2} \sum_{s\in\mathbb{Z}} e^{-\frac{1}{2}(j-\lceil p/2\rceil+sn)^2/\sigma_p^2}, \quad j = 0, \dots, n-1, \tag{4.1}$$

where $\lceil x\rceil$ denotes the smallest integer that is greater than or equal to x. Then the DFT of $\mathbf{w}$, that is the vector $\hat{\mathbf{w}} = \sqrt{n}\mathcal{F}\mathbf{w}$, is also a periodized Gaussian vector with components

$$\hat{w}_k = e^{\frac{2\pi i}{n}\lceil p/2\rceil k} \sum_{m\in\mathbb{Z}} e^{-2\left(\frac{\pi\sigma_p}{n}\right)^2(k+mn)^2}, \quad k = 0, \dots, n-1. \tag{4.2}$$

We define $\mathbf{u}$ as a truncated and normalized version of $\mathbf{w}$:

$$\mathbf{u} := \frac{1}{\sqrt{\sum_{s=0}^{p-1} w_s^2}} (w_0, \dots, w_{p-1})^\top .$$

The following lemma uses Lemma 4.1 to show that the coordinates of the corresponding vector $\mathbf{P_u}$ decay quickly. A proof of the lemma is in the Appendix.

Lemma 4.2. *Suppose that $p, n \to_c \infty$ with $c \in (0, 1]$ and $p \le n$, and that $\sigma_p \to \infty$ so that $\sigma_p = o(p)$ and $\sqrt{p} = o(\sigma_p)$. Then for all sufficiently large p, we have*

$$\frac{1}{n}|P_{\mathbf{u}s}|^2 \le \frac{2^6\sigma_p}{n} e^{-4\left(\frac{\pi\sigma_p}{n}\right)^2 s^2} + \frac{2^8\sigma_p^3}{np^2} e^{-\frac{p^2}{4\sigma_p^2}}, \qquad s \in [0, n/2]. \tag{4.3}$$

Furthermore, $|P_{\mathbf{u}s}| = |P_{\mathbf{u},n-s}| = |P_{\mathbf{u},-s}|$.

Changing the summation index in (2.8) and recalling that we understand indices modulo n, we obtain

$$\lambda_p \le \min_{k=1,\dots,K} \frac{1}{n} \sum_{s=0}^{n-1} |y_s|^2 |P_{\mathbf{u},s-\tau_k}|^2 = \min_{k=1,\dots,K} \frac{1}{n} \sum_{s=0}^{n-1} |y_{s+\tau_k}|^2 |P_{\mathbf{u}s}|^2. \tag{4.4}$$

Let us choose the integers τ_k, $k = 1, \dots, K$ as follows. Consider slowly increasing integers $b_p = o(p)$ and let $K := \lfloor \frac{n}{4b_p+2} \rfloor - 1$, where $\lfloor\cdot\rfloor$ denotes the integer part of a real number. Let

$$\tau_k := k(2b_p + 1), \quad k = 1, 2, \dots, K. \tag{4.5}$$

For any $k = 1, \dots, K$, define

$$Y_k := \frac{1}{n}\sum_{s=-b_p}^{b_p} |y_{s+\tau_k}|^2 |P_{\mathbf{u}s}|^2 \quad \text{and}$$
$$R_k := \frac{1}{n}\sum_{s=0}^{n-1} |y_{s+\tau_k}|^2 |P_{\mathbf{u}s}|^2 - Y_k.$$

Note that (2.7) and our definitions of b_p and K imply that Y_k, $k = 1, \dots, K$, are independent and identically distributed. Furthermore, R_k depend only on $P_{\mathbf{u}s}$ with $b_p < s < n - b_p$, and therefore, by Lemma 4.2 it can be shown that they are “small” (after carefully choosing the sequences σ_p and b_p).

Keeping this in mind, we use (4.4) to obtain the following bound on the probability that $\lambda_p \le e^{-2\beta \log^{1/3} p}$ with $\beta > 0$ possibly depending on c. Let $\epsilon_p := \frac{1}{2} e^{-2\beta \log^{1/3} p}$, then we have

$$\begin{aligned}
\Pr\left(\lambda_p \le e^{-2\beta\log^{1/3} p}\right) &\ge \Pr\left(\min_{k=1,\dots,K}(Y_k + R_k) \le 2\epsilon_p\right) \\
&\ge \Pr\left(\min_{k=1,\dots,K} Y_k \le \epsilon_p \text{ and } \max_{k=1,\dots,K} R_k \le \epsilon_p\right) \\
&\ge \Pr\left(\min_{k=1,\dots,K} Y_k \le \epsilon_p\right) - \Pr\left(\max_{k=1,\dots,K} R_k > \epsilon_p\right) \\
&= 1 - \left(\Pr\left(Y_k > \epsilon_p\right)\right)^K - \Pr\left(\max_{k=1,\dots,K} R_k > \epsilon_p\right) \\
&\ge 1 - \left(1 - \Pr\left(Y_k \le \epsilon_p\right)\right)^K - \sum_{k=1}^{K} \Pr\left(R_k > \epsilon_p\right). \qquad (4.6)
\end{aligned}$$

The remaining two steps of the proof will establish a lower bound on $\Pr(Y_k \le \epsilon_p)$ and an upper bound on $\Pr(R_k > \epsilon_p)$ that would imply that the right hand side of (4.6) converges to unity. Such a convergence yields

$$\Pr\left(\frac{\lambda_p}{\exp\{-\beta \log^{1/3} p\}} \le \exp\{-\beta \log^{1/3} p\}\right) \to 1, \qquad (4.7)$$

which implies that $\lambda_p = o_{\mathrm{P}}\left(\exp\{-\beta \log^{1/3} p\}\right)$ as stated by Theorem 2.1.

Step 1: establishing a lower bound on $\Pr(Y_k \le \epsilon_p)$. Using (2.7) in the definition of Y_k yields[5]

$$Y_k = \frac{1}{n}\sum_{s=-b_p}^{b_p} \frac{1}{2}\chi_s^2(2)|P_{\mathbf{u}s}|^2,$$

where $\chi_s^2(a)$, $s = -b_p, \dots, b_p$ are i.i.d. random variables having chi-squared distribution with a degrees of freedom. Denoting $\frac{1}{2n}|P_{\mathbf{u}s}|^2$ as d_s and recalling that by Lemma 4.2 $|P_{\mathbf{u}s}| = |P_{\mathbf{u},-s}|$, we represent Y_k in the form

$$Y_k = d_0\chi_0^2(2) + \sum_{s=1}^{b_p} d_s \chi_s^2(4) \le \sum_{s=0}^{b_p} d_s\chi_s^2(4),$$

where $\chi_s^2(4)$ is obtained from $\chi_s^2(2)$ by adding an independent chi-squared random variable with two degrees of freedom.

[5]We remind the reader that the exponential random variable $\mathrm{Exp}(1)$ can be interpreted as $1/2$ times chi-squared random variable with two degrees of freedom.

By Lemma 4.2, for all $0 \le s \le n/2$ and all sufficiently large p, we have

$$d_s \le r_s + \Delta, \quad \text{where } r_s := \frac{2^5 \sigma_p}{n} e^{-4(\frac{\pi\sigma_p}{n})^2 s^2}, \quad \Delta := \frac{2^7 \sigma_p^3}{np^2} e^{-\frac{p^2}{4\sigma_p^2}}. \tag{4.8}$$

Therefore, we have

$$Y_k \le \sum_{s=0}^{\infty} r_s \chi_s^2(4) + \Delta \chi^2(4(b_p+1)) =: Z_1 + Z_2,$$

where $\chi^2(4(b_p+1))$ is a random variable having chi-squared distribution with $4(b_p+1)$ degrees of freedom. Clearly,

$$\Pr(Y_k \le \epsilon_p) \ge \Pr(Z_1 \le \epsilon_p/2) - \Pr(Z_2 > \epsilon_p/2). \tag{4.9}$$

We now establish a lower bound on $\Pr(Z_1 \le \epsilon_p/2)$ and an upper bound on $\Pr(Z_2 > \epsilon_p/2)$.

First, we find a lower bound on $\Pr(Z_1 \le \epsilon_p/2)$. To simplify expressions to appear, we will introduce the following new notation:

$$\alpha_p := 2\pi\sigma_p/n, \qquad \tilde{\epsilon}_p := \pi\epsilon_p/2^6, \qquad \tilde{Z}_1 := \sum_{s=0}^{\infty} \frac{1}{2}\alpha_p e^{-(\alpha_p s)^2} \chi_s^2(4),$$

so that

$$\Pr(Z_1 \le \epsilon_p/2) = \Pr(\tilde{Z}_1 \le \tilde{\epsilon}_p).$$

To derive a lower bound on $\Pr(\tilde{Z}_1 \le \tilde{\epsilon}_p)$, we are going to use the upper and lower bounds on the probability of small Gaussian ellipsoids derived in section 2 of [35].

That paper considers the random variable $z = \sum_{i=1}^{\infty} x_i^2/a_i^2$, where a_i, $i = 1, 2, \dots$ is a deterministic sequence with $\sum a_i^{-2} < \infty$ and x_i are i.i.d. standard normal random variables. It derives an upper bound UB and a lower bound LB on $\Pr(z < \epsilon)$ and establishes a simple inequality (their inequality (15)) that bounds the ratio UB/LB from above. This simple inequality can be used to obtain a simple lower bound on LB in terms of UB. We are going to use such a simple lower bound below.

Note that $\tilde{Z}_1$ can be represented in the form $\sum_{i=1}^{\infty} x_i^2/a_i^2$ with

$$a_{4s+1}^2 = \dots = a_{4s+4}^2 = 2e^{(\alpha_p s)^2}/\alpha_p, \qquad s = 0, 1, \dots.$$

[35]'s upper bound on $\Pr(\tilde{Z}_1 \le \tilde{\epsilon}_p)$ is derived as follows. Since for any $t < 1/2$,

$$\mathbb{E} \exp\{t\chi_j^2(4)\} = (1-2t)^{-2},$$

we have for any $\ell_p > 0$,

$$\Pr(\tilde{Z}_1 \le \tilde{\epsilon}_p) \le \mathbb{E}\left(\exp\left\{-\ell_p(\tilde{Z}_1 - \tilde{\epsilon}_p)\right\}\right) = \exp\{\ell_p\tilde{\epsilon}_p\} \prod_{s=0}^{\infty} \left(1 + \ell_p\alpha_p e^{-(\alpha_p s)^2}\right)^{-2} =: UB. \tag{4.10}$$

[35] denote such a UB as $UB(s_\epsilon)$, where their s_ϵ corresponds to our $\ell_p\tilde{\epsilon}_p$ and their ϵ corresponds to our $\tilde{\epsilon}_p$.

Now define $\delta(\ell_p, \tilde{\epsilon}_p)$ as

$$\delta(\ell_p, \tilde{\epsilon}_p) = 1 - \frac{2}{\ell_p\tilde{\epsilon}_p} \sum_{s=0}^{\infty} \frac{1}{(\ell_p\alpha_p)^{-1} e^{(\alpha_p s)^2} + 1}. \tag{4.11}$$

This is an equivalent of equation (11) in [35]. Suppose that ℓ_p is chosen so that $\delta(\ell_p, \tilde{\epsilon}_p) \in (0, 1/2)$. (Later on we will verify that such a choice is possible.) Then, as follows from (15) of [35], a lower bound LB on $\Pr(\tilde{Z}_1 \le \tilde{\epsilon}_p)$ satisfies

$$LB \ge UB \exp\{-2\ell_p \tilde{\epsilon}_p \delta(\ell_p, \tilde{\epsilon}_p)\} \left(1 - \frac{1}{\ell_p \tilde{\epsilon}_p \delta^2(\ell_p, \tilde{\epsilon}_p)}\right).$$

The explicit form of LB is of no interest to us here. Taking logarithm of both sides and using (4.10), we obtain

$$\begin{aligned} \log \Pr(\tilde{Z}_1 \le \tilde{\epsilon}_p) \ge (1 - 2\delta(\ell_p, \tilde{\epsilon}_p))\, \ell_p \tilde{\epsilon}_p \quad &- \quad 2\sum_{s=0}^{\infty} \log\left(1 + \ell_p \alpha_p e^{-(\alpha_p s)^2}\right) \\ &+ \quad \log\left(1 - \frac{1}{\ell_p \tilde{\epsilon}_p \delta^2(\ell_p, \tilde{\epsilon}_p)}\right). \end{aligned} \tag{4.12}$$

We will need the following lemma. Its proof is in the Appendix.

Lemma 4.3. *For any sequences* $\ell_p > 0$ *and* $\alpha_p > 0$ *such that* $\ell_p \alpha_p \to \infty$ *while* $\alpha_p \to 0$ *and* $\alpha_p \log^{1/2}(\ell_p \alpha_p) \to 0$*, we have*

$$\sum_{s=0}^{\infty} \log\left(1 + \ell_p \alpha_p e^{-(\alpha_p s)^2}\right) = \frac{2}{3} \frac{\log^{3/2}(\ell_p \alpha_p)}{\alpha_p}(1 + o(1)), \tag{4.13}$$

$$\sum_{s=0}^{\infty} \frac{1}{(\ell_p \alpha_p)^{-1} e^{(\alpha_p s)^2} + 1} = \frac{\log^{1/2}(\ell_p \alpha_p)}{\alpha_p}(1 + o(1)). \tag{4.14}$$

Now define σ_p and ℓ_p so that

$$\alpha_p = \frac{2\pi\sigma_p}{n} = \left(\frac{2^5 \beta^3}{\log p}\right)^{1/2}, \quad \text{and} \quad \ell_p \alpha_p = \frac{3 \log^{1/2}(1/\epsilon_p)}{\tilde{\epsilon}_p}, \tag{4.15}$$

and recall that

$$\tilde{\epsilon}_p = 2^{-6}\pi\epsilon_p = 2^{-7}\pi e^{-2\beta \log^{1/3} p}.$$

Note that the conditions of Lemma 4.3 are satisfied, so that from (4.11) and (4.14),

$$\delta(\ell_p, \tilde{\epsilon}_p) = 1 - \frac{2\log^{1/2}(\ell_p \alpha_p)}{\ell_p \alpha_p \tilde{\epsilon}_p}(1 + o(1)) = \frac{1}{3}(1 + o(1)), \tag{4.16}$$

which belongs to $(0, 1/2)$ for sufficiently large p, n, as required.

Further, the first two terms on the right hand side of (4.12) satisfy

$$\begin{aligned} (1 - 2\delta(\ell_p, \tilde{\epsilon}_p))\ell_p \tilde{\epsilon}_p &= \frac{\log^{1/2}(1/\epsilon_p)}{\alpha_p}(1 + o(1)), \\ -2\sum_{s=0}^{\infty} \log\left(1 + \ell_p \alpha_p e^{-(\alpha_p s)^2}\right) &= -\frac{4}{3} \frac{\log^{3/2}(1/\epsilon_p)}{\alpha_p}(1 + o(1)), \end{aligned}$$

where the the first equality follows from (4.16) and the second equation in (4.15), and the second equality follows from (4.13) and the second equation in (4.15). The last term on the right hand side of (4.12) is $o(1)$ because $\ell_p \tilde{\epsilon}_p = \frac{3\log^{1/2}(1/\epsilon_p)}{\alpha_p} \to \infty$. Therefore, overall we have the following lower bound on $\log \Pr(Z_1 \le \epsilon_p/2)$:

$$\log \Pr(Z_1 \le \epsilon_p/2) = \log \Pr(\tilde{Z}_1 \le \tilde{\epsilon}_p) \ge -\frac{4}{3} \frac{\log^{3/2}(1/\epsilon_p)}{\alpha_p}(1 + o(1)) = -\frac{2}{3} \log(p)(1 + o(1)). \tag{4.17}$$

It remains to establish an upper bound on $\Pr(Z_2 > \epsilon_p/2)$. Recall that a centered random variable X belongs to the sub-gamma family $SG(v,u)$ for $v,u>0$ if

$$\log \mathbb{E}e^{tX} \le \frac{t^2 v}{2(1-tu)}, \qquad \forall t:\ |t| < \frac{1}{u}\,.$$

If $X \in SG(v,u)$ then for every $t \ge 0$,

$$\Pr(X > \sqrt{2vt} + ut) \le e^{-t}. \tag{4.18}$$

This inequality, as well as many other results related to sub-gamma random variables, can be found in chapter 2.4 of [18].

As is well known, $C \times \chi^2(k) \in SG(2C^2k, 2|C|)$, where C is an arbitrary constant. Therefore, $Z_2 - \mathbb{E}Z_2 \in SG(\Delta^2 8(b_p+1), 2\Delta)$, and since $\mathbb{E}Z_2 = \Delta 4(b_p+1)$, we have for any $t \ge 0$,

$$\Pr\left(Z_2 > 4\Delta(b_p+1) + 2\Delta t + 4\Delta\sqrt{(b_p+1)t}\right) \le e^{-t}.$$

Using (4.15), we obtain

$$\Delta := \frac{2^7\sigma_p^3}{np^2} e^{-\frac{p^2}{4\sigma_p^2}} < \text{const} \times \frac{n^3}{\log^{3/2} p} p^{-\frac{p^2\pi^2}{n^2 2^5\beta^3}} \le p^{-100} \tag{4.19}$$

for some sufficiently small β (that depends on c) and all sufficiently large p.

Take $t = \epsilon_p/(5\Delta)$ and let $b_p = \lfloor \log p \rfloor$. Then $t > p^\gamma$ for some $\gamma > 0$ and all sufficiently large p, and hence $t \gg b_p$. Therefore, we have

$$\epsilon_p/2 = 5\Delta t/2 > 2\Delta t + 4\Delta(b_p+1) + 4\Delta\sqrt{(b_p+1)t}$$

for all sufficiently large p. This yields the following inequality

$$\Pr(Z_2 > \epsilon_p/2) < \Pr\left(Z_2 > 4\Delta(b_p+1) + 2\Delta t + 4\Delta\sqrt{(b_p+1)t}\right) \le e^{-t} < e^{-p^\gamma}. \tag{4.20}$$

Using (4.17) and (4.20) in (4.9), we obtain

$$\Pr(Y_k \le \epsilon_p) \ge p^{-\frac{2}{3}(1+o(1))}.$$

This implies that the term $(1 - \Pr(Y_k \le \epsilon_p))^K$ on the right hand side of (4.6) converges to zero, because $K = \lfloor \frac{n}{4b_p+2} \rfloor - 1 \gg p^{\frac{2}{3}(1+o(1))}$.

Step 2: establishing an upper bound on $\Pr(R_k > \epsilon_p)$. By definition of R_k, and recalling that the indices are understood modulo n, we have

$$R_k = \sum_{s=b_p+1}^{n-b_p-1} 2|y_{s+\tau_k}|^2 d_s, \tag{4.21}$$

where $d_s := \frac{1}{2n}|P_{\mathbf{u}s}|^2$. Suppose that n is odd (the analysis for even n is very similar and we omit it). Since $\mathbb{E}|y_{s+\tau_k}|^2 = 1$ for any s (see (2.7)), and $d_s = d_{n-s}$, we have

$$\mathbb{E}R_k = 4\sum_{j=b_p+1}^{(n-1)/2} d_j. \tag{4.22}$$

Further, from (2.7) and (4.21), we have the representation

$$R_k = 2\chi^2(1) d_{n-\tau_k} + \sum_{j=1}^{(n-1)/2} \gamma_j \chi_j^2(2), \tag{4.23}$$

where $\gamma_j = \sum_{s \in S_j} d_s$ with S_j being the set of all integer s between $b_p + 1$ and $n - b_p - 1$ such that $(s + \tau_k) \mod n$ equals j or $n - j$. There are at most two elements in each S_j, $j = 1, \ldots, (n-1)/2$.

Now recall that if $X \in SG(v_X, u_X)$ and $Y \in SG(v_Y, u_Y)$ are independent, then

$$X + Y \in SG(v_X + v_Y, \max\{u_X, u_Y\}).$$

Furthermore, if $X \in SG(v, u)$ then $X \in SG(v', u')$ for each $v' \geq v$, $u' \geq u$. These facts, representation (4.23), and the inequality $\gamma_j^2 \leq \sum_{s \in S_j} 2d_s^2$ yield $R_k - \mathbb{E}R_k \in SG(v, u)$ with

$$v = \sum_{s=b_p+1}^{n-b_p-1} 8d_s^2 = \sum_{s=b_p+1}^{(n-1)/2} 16d_s^2 \quad \text{and} \quad u = \max_{b_p+1 \leq s \leq n-b_p-1} 4d_s = \max_{b_p+1 \leq s \leq (n-1)/2} 4d_s.$$

On the other hand, from (4.8), (4.15) and (4.19), we have for all sufficiently large p,

$$d_s \leq (2^4/\pi)\alpha_p e^{-\alpha_p^2 s^2} + p^{-100}.$$

This inequality together with (4.15) and the fact that $b_p = \lfloor \log p \rfloor$ yield

$$u < \frac{2^9 \beta^{3/2} p^{-32\beta^3}}{\pi \log^{1/2} p} + 4p^{-100} < p^{-\gamma}$$

for some $\gamma > 0$ (in the remaining part of the proof, γ will denote a positive constant that may change its value from one appearance to another). Similarly, we obtain for all sufficiently large p,

$$\begin{aligned} v &< 16 \int_{b_p}^{(n-1)/2} \left((2^4/\pi)\alpha_p e^{-\alpha_p^2 x^2} + p^{-100} \right)^2 \mathrm{d}x \\ &< \gamma\alpha_p(1 - \Phi(2\alpha_p b_p)) + p^{-50} < p^{-\gamma}, \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cdf. Finally, using (4.22) we obtain

$$\mathbb{E}R_k < 4 \int_{b_p}^{(n-1)/2} \left((2^4/\pi)\alpha_p e^{-\alpha_p^2 x^2} + p^{-100} \right) \mathrm{d}x < p^{-\gamma}.$$

Using these inequalities in (4.18) for $X = R_k - \mathbb{E}R_k$, we obtain for any $t \geq 1$ and some $\gamma > 0$,

$$\Pr(R_k > p^{-\gamma} t) \leq e^{-t}.$$

Taking $t = p^\gamma \epsilon_p$, and noting that $t > p^\delta$ for some small $\delta > 0$ and all sufficiently large p, we conclude

$$\Pr(R_k > \epsilon_p) \leq \exp\{-p^\delta\}.$$

This implies that the term $\sum_{k=1}^K \Pr(R_k > \epsilon_p)$ on the right hand side of (4.6) converges to zero, because $K = \lfloor \frac{n}{4b_p+2} \rfloor - 1 \ll \exp\{p^\delta\}$.

As it was mentioned after (4.6), the bounds on $\Pr(Y_k \leq \epsilon_p)$ and $\sum_{k=1}^K \Pr(R_k > \epsilon_p)$ establish the convergence in (4.7) and this completes the proof of Theorem 2.1.

□

4.2 Proof details for Theorem 2.4

As discussed in Section 2, it is sufficient to prove (2.9). Let $\eta_s = |y_s|^2 - \mathbb{E}|y_s|^2$. Then by (2.6), we have

$$\lambda_p \leq \frac{1}{n} \sum_{s=0}^{n-1} |y_s|^2 |P_{\mathbf{u}s}|^2 = \frac{1}{n} \sum_{s=0}^{n-1} |P_{\mathbf{u}s}|^2 \mathbb{E}|y_s|^2 + \frac{1}{n} \sum_{s=0}^{n-1} |P_{\mathbf{u}s}|^2 \eta_s =: c_n + Z_n,$$

where $\mathbf{u}$ is an arbitrary p-dimensional vector with $\|\mathbf{u}\| = 1$. Since $\frac{1}{n}\sum_{s=0}^{n-1}|P_{\mathbf{u}s}|^2 = 1$ and $\mathbb{E}|y_s|^2 = 1$, we have $c_n \le 1$. Therefore,

$$\|\lambda_p\|_\kappa \le 1 + \|Z_n\|_\kappa, \quad \kappa \ge 1,$$

where $\|\xi\|_\kappa$ denotes the L^κ norm of a random variable ξ, $\|\xi\|_\kappa = \left(\mathbb{E}|\xi|^k\right)^{1/\kappa}$. It remains to bound $\|Z_n\|_\kappa$. Let the ψ_1-norm of a random variable ξ be defined as

$$\|\xi\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}\exp(|\xi|/t) \le 2\}.$$

(The variables with finite ψ_1-norm are called sub-exponential.) Random variables η_s are centered χ^2 and therefore sub-exponential, and $\|\eta_s\|_{\psi_1} \le C < \infty$ for some absolute constant C. By Theorem 2.8.2 in [49], we have (assuming, for concreteness, that n is odd)

$$\Pr\left\{\Big|\sum_{s=0}^{(n-1)/2} w_s\eta_s\Big| \ge t\right\} \le 2\exp\left(-\gamma\min\left(\frac{t^2}{C^2\|w\|_2^2}, \frac{t}{C\|w\|_\infty}\right)\right), \tag{4.24}$$

where $\gamma > 0$ is an absolute constant, $w_0 = \frac{1}{n}|P_{\mathbf{u}0}|^2$ and $w_s = \frac{1}{n}\left(|P_{\mathbf{u}s}|^2 + |P_{\mathbf{u},n-s}|^2\right)$ for $s > 0$.

Let us take $\mathbf{u} = p^{-1/2}(1,\dots,1)^\top$. Then it is straightforward to verify that

$$\frac{1}{n}|P_{\mathbf{u}0}|^2 = \frac{p}{n}, \quad \text{and} \quad \frac{1}{n}|P_{\mathbf{u}s}|^2 = \frac{1}{pn}\left(\frac{\sin(\pi sp/n)}{\sin(\pi s/n)}\right)^2, 1 \le s \le n-1.$$

The latter expression is the value of the Fejér kernel (divided by n) at $2\pi s/n$. Recall that for $p \in \mathbb{N}$, the 2π-periodic Fejér kernel is

$$F_{p-1}(t) = \frac{1}{p}\left(\frac{\sin\left(\frac{pt}{2}\right)}{\sin\left(\frac{t}{2}\right)}\right)^2 = \sum_{k=-(p-1)}^{p-1}\left(1 - \frac{|k|}{p}\right)e^{ikt}, \qquad t \in \mathbb{R}.$$

Since $\sin x \ge 2x/\pi$ for all $x \in [0, \pi/2]$, we have

$$\frac{1}{n}|P_{\mathbf{u}s}|^2 \le \frac{n}{4p}\frac{1}{s^2}, \quad 1 \le s \le (n-1)/2.$$

Therefore, we see that $\max\{\|w\|_2^2, \|w\|_\infty\} \le C_c$ for some constant C_c that depends on $\lim p/n = c$. By Proposition 2.7.1 in [49], this implies that $\left\|Z_n\right\|_{\psi_1} \le C_c$, where the value of C_c may change from one appearance to another, and

$$\left\|Z_n\right\|_\kappa \le C\kappa\left\|Z_n\right\|_{\psi_1} \le C_{\kappa,c},$$

which is what we wanted to prove. □

4.3 Proof details for Theorem 2.5

To complete the proof of Theorem 2.5 outlined in Section 2 we need, first, to establish (2.13), and second, to show that the term S_1 in (2.14) is bounded away from zero.

Recall that (2.13) reads as $\Pr\left(\min\{\xi_0, \xi_{(r)}\} > r/N\right) \to 1$. Since $\xi_0 \sim \chi^2(1)$ and $r/N \to 0$, we have $\Pr(\xi_0 > r/N) \to 1$. Further, as follows e.g. from equation 1.9 of [42], the order statistic $\xi_{(r)}$ has the following representation

$$\xi_{(r)} = \sum_{i=1}^{r}\frac{\zeta_i}{N-i}, \tag{4.25}$$

where $\zeta_1, \dots, \zeta_{N-1}$ are mutually independent random variables with $\mathrm{Exp}(1/2)$ distribution. From (4.25), we have

$$\mathbb{E}\xi_{(r)} = \sum_{i=1}^{r} \frac{2}{N-i}, \qquad \operatorname{Var}\xi_{(r)} = \sum_{i=1}^{r} \frac{4}{(N-i)^2}.$$

In particular,

$$\mathbb{E}\xi_{(r)} > 2r/N, \qquad \operatorname{Var}\xi_{(r)} < 5r/N^2 \tag{4.26}$$

for all sufficiently large p (recall that $r := \lfloor p^{1-1/m-\epsilon/2} \rfloor$ and $N = (n+1)/2$ with $p/n \to c \in (0,1]$). Therefore, by Chebyshev's inequality,

$$\Pr\left(\xi_{(r)} < r/N\right) < \Pr\left(\xi_{(r)} - \mathbb{E}\xi_{(r)} < -r/N\right) \le \frac{5r/N^2}{r^2/N^2} \to 0.$$

Hence $\Pr(\xi_{(r)} > r/N) \to 1$, which completes the proof of (2.13).

We now turn to the proof of the boundedness of S_1 away from zero. Recall that

$$S_1 := \frac{1}{2} T_{p-1}(0) + \sum_{j=r+1}^{N-1} T_{p-1}(\omega_{\sigma(j)}), \qquad \text{and let} \qquad S_2 := \sum_{j=1}^{r} T_{p-1}(\omega_{\sigma(j)}).$$

As follows from (2.11), $S_2 = 1/2 - S_1$. We show that S_1 is bounded away from zero by using the L_1 version of the Bernstein inequality for the derivative of a cosine trigonometric polynomial (see e.g. Theorem 2.5 in Chapter 4 of [21]):

$$\int_0^\pi \left|T'_{p-1}(\omega)\right| \mathrm{d}\omega \le (p-1) \int_0^\pi T_{p-1}(\omega) \mathrm{d}\omega = (p-1)\frac{\pi}{n}. \tag{4.27}$$

Here the last equality follows from (2.10).

Let us call ω_i type 2 if $i = \sigma(j)$ with $1 \le j \le r$ and type 1 otherwise. Then the set $\{\omega_0, \omega_1, \dots, \omega_{N-1}\}$ splits into a union of K interlacing clusters (a cluster may contain only one element) of type 1 and type 2. The first cluster is of type 1 because ω_0 is of type 1. Assuming that none of $\sigma(j) : 1 \le j \le r$ equals $N-1$ on A_p (such an assumption is without loss of generality because the probability of the latter event equals $(N-r-1)/(N-1)$, which converges to 1), the last cluster is also of type 1 because ω_{N-1} is of type 1. In particular, K is an odd integer.

Picking just one element from each of the clusters, we obtain a sequence $\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_K}$ with odd elements of type 1 and even elements of type 2. Clearly,

$$\int_0^\pi \left|T'_{p-1}(\omega)\right| \mathrm{d}\omega \quad \ge \quad |T_{p-1}(\omega_{i_1}) - T_{p-1}(\omega_{i_2})| + \dots + |T_{p-1}(\omega_{i_{K-1}}) - T_{p-1}(\omega_{i_K})|. \tag{4.28}$$

Figure 4 illustrates this inequality. On the other hand, the right hand side of (4.28) is no smaller than

$$\sum_{\text{even } 0<k<K} (T_{p-1}(\omega_{i_k}) - T_{p-1}(\omega_{i_{k+1}})) - \sum_{\text{odd } 0<k<K} (T_{p-1}(\omega_{i_k}) - T_{p-1}(\omega_{i_{k+1}}))$$
$$\ge \quad 2 \sum_{\text{even } 0<k<K} T_{p-1}(\omega_{i_k}) - \left(T_{p-1}(\omega_{i_1}) + 2 \sum_{\text{odd } 1<k\le K} T_{p-1}(\omega_{i_k}) \right).$$

Since all ω_{i_k} with odd k are of type 1 and T_{p-1} is a non-negative trigonometric polynomial, the expression in the large brackets in the latter display is no larger than $2S_1$. Hence, combining the latter inequality with (4.27) and (4.28) and recalling that all ω_{i_k} with even k are of type 2, we obtain

$$\frac{(p-1)\pi}{n} \ge 2 \sum_{\omega_{i_k} \text{ of type 2}} T_{p-1}(\omega_{i_k}) - 2S_1. \tag{4.29}$$

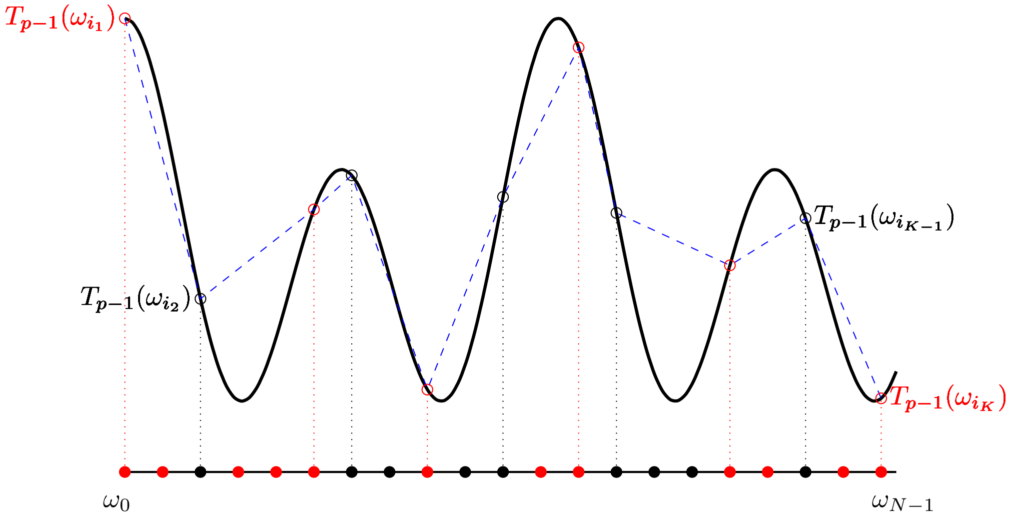

Figure 4: Illustration of the inequality (4.28). Adjacent red dots form clusters of ω_i of the first type. Adjacent black dots form clusters of ω_i of the second type. Total variation of the smooth graph, representing the left hand of (4.28), is larger than that of the piece-wise linear dashed graph, representing the right hand side of (4.28).

Let us now show that ω_i of type 2 cannot cluster in groups of m or more adjacent ω_i's with probability approaching one as $p \to \infty$. Recall that ω_i is of type 2 if $i = \sigma(j)$ with $1 \leq j \leq r$, and that σ is a random permutation uniformly distributed on the set of all permutations of the integers $1, \dots, N-1$. Consider the probability that randomly picked r integers out of $1, \dots, N-1$ contain at least m adjacent integers, that is, integers $i + j$ with $j = 1, \dots, m$ for some $0 \leq i \leq N-1-m$. We denote this probability as $P(N, r, m)$. We need to prove that $P(N, r, m) \to 0$ as $N \to \infty$.

Denote the number of possible ordered sequences of r integers out of $N-1$ that would contain at least m adjacent integers as $\gamma(N, r, m)$. Then,

$$P(N, r, m) = \frac{\gamma(N, r, m)}{(N-1)!/(N-1-r)!}.$$

On the other hand, $\gamma(N, r, m)$ is no larger than the product of the following three numbers: $r!/(r-m)!$ - the number of possible choices of m ordered places out of r; $N-m$ - the number of possible choices of m adjacent integers out of $1, \dots, N-1$; and $(N-1-m)!/(N-1-r)!$ - the number of possible choices of ordered sequences of length $r-m$ out of remaining $N-1-m$ integers. Therefore,

$$P(N, r, m) \leq \frac{\frac{r!}{(r-m)!} \times (N-m) \times \frac{(N-1-m)!}{(N-1-r)!}}{\frac{(N-1)!}{(N-1-r)!}} = \frac{r \times \cdots \times (r-m+1)}{(N-1) \times \cdots \times (N-m+1)}.$$

Clearly, as $N \to \infty$ while m is being fixed, the right hand side converges to zero as long as $r \ll N^{\frac{m-1}{m}}$, which is indeed the case for $r := \lfloor p^{1-1/m-\epsilon/2} \rfloor$.

To summarize, as we have just shown, ω_i of type 2 cannot cluster in groups of m or more adjacent ω_i's with probability approaching one as $p \to \infty$. Therefore, returning to inequality (4.29), if we pick ω_{i_k} of the second types so that the corresponding $T_{p-1}(\omega_{i_k})$ is the maximum possible in the cluster that ω_{i_k} belongs to (as illustrated in Figure 4), we have, on a sequence of high probability events,

$$\sum_{\omega_{i_k} \text{ of type 2}} T_{p-1}(\omega_{i_k}) \geq \frac{1}{m-1} S_2.$$

Combining this with (4.29) and using the fact that $S_2 = 1/2 - S_1$, we obtain

$$S_1 \geq \frac{m-1}{2m} \left(\frac{1}{m-1} - \frac{(p-1)\pi}{n} \right). \tag{4.30}$$

But the right hand side is bounded away from zero for all sufficiently large p as long as $\lim p/n < \frac{1}{\pi(m-1)}$, which is required by Theorem 2.5. This concludes the proof. □

4.4 Proof details for Theorem 2.6

We need the following result (see its proof in the Appendix), showing that trigonometric polynomials cannot decay too fast near the point of their maximum.

Lemma 4.4. *Let $T_k(\omega) = a_0 + \sum_{j=1}^{k} (a_j \cos(j\omega) + b_j \sin(j\omega))$ be a non-negative trigonometric polynomial of degree k with real coefficients, and suppose that $\max T_k(\omega) =: M > 0$. If $T_k(\bar{\omega}) = M$, then*

$$T_k(\bar{\omega} + \eta) \geq M \cos^2 \frac{k\eta}{2} \qquad \textit{for all } |\eta| \leq \pi/k.$$

The setting of our proof of Theorem 2.6 is exactly the same as that of the proof of Theorem 2.5 above. In particular, we keep the definitions of r, of type 1 and type 2 clusters, and of the events A_p. Now let $r_1 := \lfloor p^{1/2-1/(2m)-\epsilon/4} \rfloor$ so that $r_1/\sqrt{r} \to 1$, and assume that events A_p are adjusted to imply that $\xi_{(r_1)} > r_1/N$. Such an adjustment is feasible because $\Pr(\xi_{(r_1)} > r_1/N) \to 1$, as can be shown similarly to (2.13) above. Let

$$S_{11} := \frac{1}{2} T_{p-1}(0) + \sum_{j=r_1+1}^{N-1} T_{p-1}(\omega_{\sigma(j)}), \qquad S_{21} := \sum_{j=1}^{r_1} T_{p-1}(\omega_{\sigma(j)}) = 1/2 - S_{11},$$

where the latter equality follows from (2.11). Similarly to (2.14), on A_p we have

$$\lambda_p > \frac{r_1}{N} S_{11}. \tag{4.31}$$

Further, let $M := \max_\omega T_{p-1}(\omega) = T_{p-1}(\bar{\omega})$. Since $T_{p-1}(\omega)$ is an even function, we can assume without loss of generality that $\bar{\omega} \in [0, \pi]$. Clearly, $M \geq S_{21}/r_1 = (1/2 - S_{11})/r_1$. This inequality and Lemma 4.4 imply that

$$T_{p-1}(\bar{\omega} + \eta) \geq \frac{1/2 - S_{11}}{r_1} \cos^2 \frac{p\eta}{2} \qquad \text{for all } |\eta| \leq \frac{\pi}{p}. \tag{4.32}$$

(Strictly following Lemma 4.4 yields even a stronger statement with $\cos^2 \frac{(p-1)\eta}{2}$ and $|\eta| \leq \frac{\pi}{p-1}$ in the above display.)

Consider the following segment with a center at $\bar{\omega}$:

$$I := [\bar{\omega} - (1-\delta_0)\pi/p, \bar{\omega} + (1-\delta_0)\pi/p],$$

where constant $\delta_0 > 0$ is such that, for all sufficiently large p, $p/n \leq (1-\delta_0)/m$ (the existence of such a constant follows from the conditions of Theorem 2.6). Inequality (4.32) implies that, for any point $\omega_{\sigma(j)} := \frac{2\pi\sigma(j)}{n} \in I$, we must have

$$T_{p-1}(\omega_{\sigma(j)}) \geq \frac{1/2 - S_{11}}{r_1} \cos^2 \frac{(1-\delta_0)\pi}{2}.$$

Note that the number of adjacent points of form $2\pi s/n$, $s \in \mathbb{Z}$ in I cannot be smaller than

$$\left\lfloor \frac{(1-\delta_0)2\pi/p}{2\pi/n} \right\rfloor = \left\lfloor \frac{(1-\delta_0)n}{p} \right\rfloor \geq m.$$

Therefore, if $I \subseteq [\omega_1, \omega_{N-1}]$, at least one of the points $\omega_s \in I$ must be of type 1, and hence, by (2.12) we must have

$$\lambda_p \geq \xi_{(r)} T_{p-1}(\omega_s) > \frac{r}{N} \frac{1/2 - S_{11}}{r_1} \cos^2 \frac{(1-\delta_0)\pi}{2}.$$

Combining this with (4.31), we obtain

$$\lambda_p > \max \left\{ \frac{r_1}{N} S_{11}, \frac{r}{N} \frac{1/2 - S_{11}}{r_1} \cos^2 \frac{(1-\delta_0)\pi}{2} \right\}.$$

Since $S_{11} \in [0, 1/2]$, we conclude that $\lambda_p > Cr_1/N$ for some positive constant C, which implies that, on A_p, $\lambda_p > p^{-1/2-1/(2m)-\epsilon}$ for all sufficiently large p, and Theorem 2.6 holds.

To conclude the proof, we need to consider the cases where $I \nsubseteq [\omega_1, \omega_{N-1}]$. This is possible only if $\bar{\omega} \in [0, (1-\delta_0)\pi/p]$ or if $\bar{\omega} \in [\pi - (1-\delta_0)\pi/p, \pi]$. Obviously, in the former case, $\omega_1 \in I$, whereas in the latter case, $\omega_{N-1} \in I$. Let us adjust events A_p so that they imply that both ω_1 and ω_{N-1} are of type 1. Then, we can proceed as in the case $I \subseteq [\omega_1, \omega_{N-1}]$, setting ω_s to ω_1 or ω_{N-1}. □

5 Conclusion

This paper proves that, somewhat unexpectedly, the smallest non-zero eigenvalues of the rectangular $p \times n$ Toepliz and circulant matrices, (1.1) and (1.2), converge in probability and in expectation to zero as $p \to \infty$ for all $c := \lim p/n(p) \in (0, 1]$. We show that the rate of the convergence is faster than poly-log, and derive some polynomial upper bounds on the rate. Our Monte Carlo exercises suggest that the actual rate might be polynomial and that the upper bounds provided by Theorems 2.5 and 2.6 could be suboptimal.

In future research, it would be valuable to prove that the rate of convergence is indeed polynomial and characterize it precisely. Additionally, it would be interesting to extend our results to the case of non-Gaussian entries of the Toeplitz and circulant matrices.

Supplementary Material

Appendix: Proofs of technical lemmas.

Proof of Lemma 4.1

Recall the Poisson summation formula:

$$\sum_{j \in \mathbb{Z}} e^{-2\pi \mathrm{i} j\omega} f(j) = \sum_{m \in \mathbb{Z}} \hat{f}(\omega + m), \tag{5.1}$$

where $f(x)$ is any integrable function and $\hat{f}(\omega) = \int f(x) e^{-2\pi \mathrm{i}\omega x} \mathrm{d}x$ is its Fourier transform. Applying (5.1) to the Fourier pair

$$f(x) = (2\pi\sigma_p^2)^{-1/2} \exp\left\{ -\frac{(x - \lceil p/2 \rceil)^2}{2\sigma_p^2} \right\} \quad \text{and} \quad \hat{f}(\omega) = \exp\left\{ -2\pi \mathrm{i} \lceil p/2 \rceil \omega - 2(\pi\sigma_p\omega)^2 \right\},$$

we obtain

$$(2\pi\sigma_p^2)^{-1/2}\sum_{j\in\mathbb{Z}} e^{-2\pi \mathrm{i} j\omega}\exp\left\{-\frac{(j-\lceil p/2\rceil)^2}{2\sigma_p^2}\right\} = e^{-2\pi\mathrm{i}\lceil p/2\rceil\omega}\sum_{m\in\mathbb{Z}}\exp\left\{-2(\pi\sigma_p(\omega+m))^2\right\}.$$

Setting $\omega = -k/n$ and grouping terms on the left hand side yields (4.2).

Proof of Lemma 4.2

First, borrowing some ideas from the proof of Proposition 7 of [6], we bound $|\hat{w}_s|$, defined in (4.2). We have

$$|\hat{w}_s| = \sum_{m\in\mathbb{Z}} e^{-2(\frac{\pi\sigma_p}{n})^2(s+mn)^2} = \sum_{m\geq 1} e^{-2(\frac{\pi\sigma_p}{n})^2(s+mn)^2} + e^{-2(\frac{\pi\sigma_p}{n})^2 s^2} + \sum_{m\leq -1} e^{-2(\frac{\pi\sigma_p}{n})^2(s+mn)^2}.$$

For the first sum on the right hand side, we have

$$\begin{aligned}\sum_{m\geq 1} e^{-2(\frac{\pi\sigma_p}{n})^2(s+mn)^2} &< \sum_{m\geq 1} e^{-2(\frac{\pi\sigma_p}{n})^2(s^2+2sn+(mn)^2)}\\ &= e^{-2(\frac{\pi\sigma_p}{n})^2(s^2+2sn)}\sum_{m\geq 1} e^{-2(\frac{\pi\sigma_p}{n})^2(mn)^2}\\ &< e^{-2(\frac{\pi\sigma_p}{n})^2(s^2+2sn)}\int_0^\infty e^{-2\pi^2\sigma_p^2x^2}\mathrm{d}x\\ &= \frac{1}{\sqrt{8\pi}\sigma_p}e^{-2(\frac{\pi\sigma_p}{n})^2(s^2+2sn)}. \end{aligned} \tag{5.2}$$

If $0\leq s\leq n/2$, then

$$\begin{aligned}\sum_{m\leq -1} e^{-2(\frac{\pi\sigma_p}{n})^2(s+mn)^2} &= \sum_{m\geq 0} e^{-2(\frac{\pi\sigma_p}{n})^2(n-s+mn)^2}\\ &\leq e^{-2(\frac{\pi\sigma_p}{n})^2 s^2} + \sum_{m\geq 1} e^{-2(\frac{\pi\sigma_p}{n})^2(n-s+mn)^2}.\end{aligned}$$

If $0 \leq s \leq n/2$, the elements of the latter sum are no larger than those of $\sum_{m\geq 1} e^{-2(\frac{\pi\sigma_p}{n})^2(s+mn)^2}$. Hence, the argument that led us to (5.2) yields

$$\sum_{m\leq -1} e^{-2(\frac{\pi\sigma_p}{n})^2(s+mn)^2} \leq e^{-2(\frac{\pi\sigma_p}{n})^2 s^2} + \frac{1}{\sqrt{8\pi}\sigma_p}e^{-2(\frac{\pi\sigma_p}{n})^2(s^2+2sn)}.$$

Summing up, for $0\leq s\leq n/2$ we have

$$|\hat{w}_s| \leq 2e^{-2(\frac{\pi\sigma_p}{n})^2 s^2}\left(1+\frac{e^{-4s\pi^2\sigma_p^2/n}}{\sqrt{8\pi}\sigma_p}\right). \tag{5.3}$$

By definition, it is clear that, for $n/2 < s \leq n-1$, $|\hat{w}_s| = |\hat{w}_{n-s}|$.

We can similarly obtain an upper bound on $|w_j|$, defined in (4.1). For $\lceil p/2\rceil \leq j \leq \lceil p/2\rceil + n/2$, we have

$$\begin{aligned}\sum_{s\geq 0} e^{-\frac{1}{2}(j-\lceil p/2\rceil+sn)^2/\sigma_p^2} &\leq e^{-\frac{1}{2}(j-\lceil p/2\rceil)^2/\sigma_p^2}\left(1+\sum_{s\geq 1} e^{-\frac{1}{2}(sn)^2/\sigma_p^2}\right)\\ &\leq e^{-\frac{1}{2}(j-\lceil p/2\rceil)^2/\sigma_p^2}\left(1+\int_0^\infty e^{-\frac{1}{2}(xn)^2/\sigma_p^2}\mathrm{d}x\right)\\ &= e^{-\frac{1}{2}(j-\lceil p/2\rceil)^2/\sigma_p^2}\left(1+\frac{\sqrt{2\pi}\sigma_p}{2n}\right).\end{aligned}$$

Further,

$$\sum_{s<0} e^{-\frac{1}{2}(j-\lceil p/2\rceil+sn)^2/\sigma_p^2} \quad = \quad \sum_{s\geq 0} e^{-\frac{1}{2}(n+\lceil p/2\rceil-j+sn)^2/\sigma_p^2}.$$

Since by assumption, $n+\lceil p/2\rceil - j \geq j - \lceil p/2\rceil \geq 0$, the latter sum is no larger than the sum studied in the preceding display. Hence overall

$$\sum_{s\in\mathbb{Z}} e^{-\frac{1}{2}(j-\lceil p/2\rceil+sn)^2/\sigma_p^2} \leq 2e^{-\frac{1}{2}(j-\lceil p/2\rceil)^2/\sigma_p^2}\left(1+\frac{\sqrt{2\pi}\sigma_p}{2n}\right),$$

and thus, for $\lceil p/2\rceil \leq j \leq \lceil p/2\rceil + n/2$, we have

$$|w_j| \leq (2\pi\sigma_p^2)^{-1/2} e^{-\frac{1}{2}(j-\lceil p/2\rceil)^2/\sigma_p^2}\left(2+\frac{\sqrt{2\pi}\sigma_p}{n}\right). \tag{5.4}$$

Now note that for any $s=0,\dots,n-1$,

$$\left|\sum_{j=p}^{n-1} e^{\frac{2\pi \mathrm{i}}{n}sj} w_j\right| \leq \sum_{j=p}^{n-1}|w_j| \leq 2\sum_{j=p}^{\lfloor\lceil p/2\rceil+n/2\rfloor}|w_j|.$$

The latter inequality follows from the fact that, by definition (4.1), $w_j = w_{2\lceil p/2\rceil+n-j}$. In particular, for any j such that $\lfloor\lceil p/2\rceil+n/2\rfloor < j \leq n-1$, we have $w_j = w_k$ with $k := 2\lceil p/2\rceil + n - j$ satisfying $p \leq k \leq \lfloor\lceil p/2\rceil + n/2\rfloor$.

Using bound (5.4) in the above display, we obtain

$$\begin{aligned}\left|\sum_{j=p}^{n-1} e^{\frac{2\pi \mathrm{i}}{n}sj} w_j\right| \quad &< \quad 2(2\pi\sigma_p^2)^{-1/2}\left(2+\frac{\sqrt{2\pi}\sigma_p}{n}\right)\sum_{j=p}^{\infty} e^{-\frac{1}{2}(j-\lceil p/2\rceil)^2/\sigma_p^2} \\ &< \quad 2(2\pi\sigma_p^2)^{-1/2}\left(2+\frac{\sqrt{2\pi}\sigma_p}{n}\right)\int_{p/2-2}^{\infty} e^{-\frac{1}{2}x^2/\sigma_p^2}\mathrm{d}x.\end{aligned}$$

This yields

$$\left|\sum_{j=p}^{n-1} e^{\frac{2\pi \mathrm{i}}{n}sj} w_j\right| \leq \left(4+\frac{\sqrt{8\pi}\sigma_p}{n}\right)\left(1-\Phi\left(\frac{p-4}{2\sigma_p}\right)\right),$$

where Φ denotes the cumulative distribution function of the standard normal random variable. Since for any $t>0$,

$$1-\Phi(t) < \frac{1}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/2}, \tag{5.5}$$

we conclude that for $p>4$,

$$\left|\sum_{j=p}^{n-1} e^{\frac{2\pi \mathrm{i}}{n}sj} w_j\right| \leq \left(\frac{8}{\sqrt{2\pi}}+\frac{4\sigma_p}{n}\right)\frac{\sigma_p}{p-4}e^{-\frac{(p-4)^2}{8\sigma_p^2}}. \tag{5.6}$$

Next, let us establish a lower bound on $\sqrt{\sum_{j=0}^{p-1} w_j^2}$. From the definition of w_j we have

$$w_j > (2\pi\sigma_p^2)^{-1/2} e^{-\frac{1}{2}(j-\lceil p/2\rceil)^2/\sigma_p^2}.$$

Therefore,

$$\begin{aligned}\sum_{j=0}^{p-1} w_j^2 &> (2\pi\sigma_p^2)^{-1}\sum_{j=0}^{p-1} e^{-(j-\lceil p/2\rceil)^2/\sigma_p^2}\\ &> (2\pi\sigma_p^2)^{-1}\left(\int_{-p/2}^{p/2} e^{-x^2/\sigma_p^2}\mathrm{d}x - 2\right)\\ &= \frac{1}{\sqrt{\pi}\sigma_p}\Phi\left(\frac{p}{\sqrt{2}\sigma_p}\right) - \frac{1}{2\sqrt{\pi}\sigma_p} - \frac{1}{\pi\sigma_p^2}.\end{aligned}$$

Using (5.5), we arrive at the inequality

$$\sum_{j=0}^{p-1} w_j^2 > \frac{1}{2\sqrt{\pi}\sigma_p} - \frac{1}{\pi\sigma_p^2} - \frac{1}{\pi p}e^{-p^2/(4\sigma_p^2)}. \tag{5.7}$$

By definition,

$$\frac{1}{n}|P_{\mathbf{u}s}|^2 = \frac{1}{n}\left|(\sqrt{n}\mathcal{F}\mathbf{u}_C)_s\right|^2 = \frac{1}{n\sum_{j=0}^{p-1} w_j^2}\left|\hat{w}_s - \sum_{j=p}^{n-1} e^{\frac{2\pi \mathrm{i}}{n}sj}w_j\right|^2.$$

Combining (5.3), (5.6) and (5.7), we arrive at the following inequality for any integer $s \in [0, n/2]$ and $p > 4$

$$\frac{1}{n}|P_{\mathbf{u}s}|^2 < \frac{8e^{-4(\frac{\pi\sigma_p}{n})^2 s^2}\left(1+\frac{e^{-4s\pi^2\sigma_p^2/n}}{\sqrt{8\pi}\sigma_p}\right)^2 + 2\left(\frac{8}{\sqrt{2\pi}}+\frac{4\sigma_p}{n}\right)^2\frac{\sigma_p^2}{(p-4)^2}e^{-\frac{(p-4)^2}{4\sigma_p^2}}}{n\left(\frac{1}{2\sqrt{\pi}\sigma_p} - \frac{1}{\pi\sigma_p^2} - \frac{1}{\pi p}e^{-p^2/(4\sigma_p^2)}\right)}.$$

Since $p, n \to_c \infty$ and $\sigma_p \to \infty$ so that $\sigma_p = o(p)$, we have for all sufficiently large p and any $s \in [0, n/2]$

$$\begin{aligned}\left(1+\frac{e^{-4s\pi^2\sigma_p^2/n}}{\sqrt{8\pi}\sigma_p}\right)^2 &\leq \sqrt{2},\\ \left(\frac{8}{\sqrt{2\pi}}+\frac{4\sigma_p}{n}\right)^2\frac{\sigma_p^2}{(p-4)^2} &\leq \sqrt{2}\frac{64}{2\pi}\frac{\sigma_p^2}{p^2},\\ \frac{1}{2\sqrt{\pi}\sigma_p} - \frac{1}{\pi\sigma_p^2} - \frac{1}{\pi p}e^{-p^2/(4\sigma_p^2)} &\geq \frac{1}{\sqrt{2}}\frac{1}{2\sqrt{\pi}\sigma_p},\end{aligned}$$

which (together with $1 < \sqrt{\pi} < 2$ and $p = o(\sigma_p^2)$) yields

$$\frac{1}{n}|P_{\mathbf{u}s}|^2 < \frac{2^6\sigma_p}{n}e^{-4(\frac{\pi\sigma_p}{n})^2 s^2} + \frac{2^8\sigma_p^3}{np^2}e^{-\frac{p^2}{4\sigma_p^2}}.$$

Finally, the equality $|P_{\mathbf{u}s}| = |P_{\mathbf{u},n-s}| = |P_{\mathbf{u},-s}|$ follows directly from the definition of $\mathbf{P_u}$ as the DFT of $\mathbf{u}_C$.

Proof of Lemma 4.3

For any positive integer J, we have

$$\begin{aligned}\sum_{j=0}^{\infty}\log\left(1+\ell_p\alpha_p e^{-(\alpha_p j)^2}\right) &\geq \sum_{j=0}^{J}\left(\log(\ell_p\alpha_p) - (\alpha_p j)^2\right)\\ &= (J+1)\log(\ell_p\alpha_p) - \frac{\alpha_p^2}{6}J(J+1)(2J+1).\end{aligned}$$

Setting $J = \left\lfloor \frac{\log^{1/2}(\ell_p\alpha_p)}{\alpha_p} \right\rfloor$, we obtain

$$\sum_{j=0}^{\infty} \log\left(1+\ell_p\alpha_p e^{-(\alpha_p j)^2}\right) \geq \frac{2}{3}\frac{\log^{3/2}(\ell_p\alpha_p)}{\alpha_p}(1+o(1)). \tag{5.8}$$

On the other hand, for the above choice of J, we have $\ell_p\alpha_p e^{-(\alpha_p J)^2} \geq 1$ and therefore,

$$\begin{aligned} \sum_{j=0}^{\infty} \log\left(1+\ell_p\alpha_p e^{-(\alpha_p j)^2}\right) &\leq \sum_{j=0}^{J} \log\left(2\ell_p\alpha_p e^{-(\alpha_p j)^2}\right) + \int_J^{\infty} \log\left(1+\ell_p\alpha_p e^{-(\alpha_p x)^2}\right) \mathrm{d}x \\ &= \frac{2}{3}\frac{\log^{3/2}(\ell_p\alpha_p)}{\alpha_p}(1+o(1)) + \int_J^{\infty} \log\left(1+\ell_p\alpha_p e^{-(\alpha_p x)^2}\right) \mathrm{d}x. \end{aligned}$$

Let us show that the latter integral is of order lower than $\log^{3/2}(\ell_p\alpha_p)/\alpha_p$. Integrating by parts, we obtain

$$\begin{aligned} &\int_J^{\infty} \log\left(1+\ell_p\alpha_p e^{-(\alpha_p x)^2}\right) \mathrm{d}x \\ =\ & -J\log(1+\ell_p\alpha_p e^{-(\alpha_p J)^2}) + 2\alpha_p^3\ell_p \int_J^{\infty} \frac{x^2 e^{-(\alpha_p x)^2}}{1+\ell_p\alpha_p e^{-(\alpha_p x)^2}}\mathrm{d}x \\ \leq\ & 2\alpha_p^3\ell_p \int_J^{\infty} x^2 e^{-(\alpha_p x)^2}\mathrm{d}x = \ell_p\Gamma\left(3/2, (\alpha_p J)^2\right) \\ =\ & J\ell_p\alpha_p e^{-(\alpha_p J)^2}(1+o(1)). \end{aligned}$$

Here $\Gamma(a,z) := \int_z^{\infty} e^{-t}t^{a-1}dt$ is the complementary incomplete Gamma function, and the last equality follows from the well-known asymptotics for this function (see e.g. [38], p. 66). On the other hand,

$$\ell_p\alpha_p e^{-(\alpha_p J)^2} \leq \ell_p\alpha_p e^{-(\log^{1/2}\ell_p\alpha_p - \alpha_p)^2} \leq e^{2\alpha_p \log^{1/2}(\ell_p\alpha_p)} \to 1.$$

Therefore, the integral is bounded from above by a quantity of order lower than $\log^{3/2}(\ell_p\alpha_p)/\alpha_p$. To summarize,

$$\sum_{j=0}^{\infty} \log\left(1+\ell_p\alpha_p e^{-(\alpha_p j)^2}\right) \leq \frac{2}{3}\frac{\log^{3/2}(\ell_p\alpha_p)}{\alpha_p}(1+o(1)). \tag{5.9}$$

Combining (5.8) and (5.9) yields (4.13).

Next, let now

$$J = \left\lfloor \frac{\log^{1/2}(a\ell_p\alpha_p)}{\alpha_p} \right\rfloor,$$

where $a > 0$ is an arbitrarily small constant. Such a J is well defined for all sufficiently large p. We have $(\ell_p\alpha_p)^{-1}e^{(\alpha_p J)^2} \leq a$, and therefore

$$\sum_{j=0}^{\infty} \frac{1}{(\ell_p\alpha_p)^{-1}e^{(\alpha_p j)^2}+1} \geq \frac{1+J}{1+a} \geq \frac{\log^{1/2}(a\ell_p\alpha_p)}{\alpha_p(1+a)}. \tag{5.10}$$

On the other hand,

$$\begin{aligned} \sum_{j=0}^{\infty} \frac{1}{(\ell_p\alpha_p)^{-1}e^{(\alpha_p j)^2}+1} &\leq 1+J+\ell_p\alpha_p \sum_{j=J+1}^{\infty} e^{-(\alpha_p j)^2} \\ &\leq 1+J+\ell_p\alpha_p \int_J^{\infty} e^{-(\alpha_p x)^2}\mathrm{d}x \\ &= 1+J+\ell_p\sqrt{\pi}(1-\Phi(\sqrt{2}\alpha_p J)) \\ &\leq 1+J+\frac{\ell_p}{2\alpha_p J}e^{-\alpha_p^2 J^2}. \end{aligned}$$

But

$$\frac{\ell_p}{J}e^{-\alpha_p^2 J^2} \leq (1+o(1))\frac{\ell_p\alpha_p e^{-\alpha_p^2\left(\log^{1/2}(a\ell_p\alpha_p)/\alpha_p-1\right)^2}}{\log^{1/2}(a\ell_p\alpha_p)}$$
$$\leq (1+o(1))\frac{e^{2\alpha_p\log^{1/2}(a\ell_p\alpha_p)}}{a\log^{1/2}(a\ell_p\alpha_p)} = \frac{1+o(1)}{a\log^{1/2}(a\ell_p\alpha_p)} = o(1).$$

Therefore,

$$\sum_{j=0}^{\infty}\frac{1}{(\ell_p\alpha_p)^{-1}e^{(\alpha_p j)^2}+1} \leq \frac{\log^{1/2}(a\ell_p\alpha_p)}{\alpha_p}(1+o(1)). \tag{5.11}$$

Inequalities (5.10), (5.11) and the fact that a is an arbitrarily small positive number yield (4.14).

Proof of Lemma 4.4

Our proof of Lemma 4.4 closely follows the proof of a lemma on page 1511 of [47]. We make only a few minor changes. Consider trigonometric polynomial $\Psi_k(\eta) = T_k(\bar\omega+\eta) - M\cos^2(k\eta/2)$. It has a double root at $\eta = 0$, and, if $T_k(\bar\omega+\eta) < M\cos^2\frac{k\eta}{2}$ for some $0 < \eta < \pi/k$, it also has at least one additional root (the roots are counted according to their multiplicity) on each of the segments $[j\pi/k, (j+1)\pi/k]$ with $j = 0, \dots, 2k-2$. This would imply that $\Psi_k(\eta)$ has $2+2k-1 = 2k+1$ roots, which is impossible because its degree is k. The possibility that $T_k(\bar\omega+\eta) < M\cos^2\frac{k\eta}{2}$ for some $-\pi/k < \eta < 0$ is eliminated similarly.

The complex normal case

This subsection outlines the adjustments required in the proofs of Theorems 2.1 and 2.4–2.6 when $x_i, i \in \mathbb{Z}$ is a sequence of i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ random variables.

In Theorem 2.1, the analysis of the upper bound on $\Pr(R_k > \epsilon_p)$ (Step 2 of the proof) becomes slightly simpler. Since all $y_{s+\tau_k}$ are now i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ random variables, the dependence structure described for the real-valued case in (2.7) is no longer present. Consequently, the parameters v and u of the corresponding sub-Gamma family can be reduced by half.

In the proof of Theorem 2.4, the upper limit $(n-1)/2$ in the summation of equation (4.24) should be replaced with $n-1$, and the weights w_s redefined as $w_s = \frac{1}{n}|P_{\mathbf{u}s}|^2$. No further modifications are required.

In the proof of Theorem 2.5, the main change is that the trigonometric polynomial T_{p-1} is no longer a cosine polynomial but a general nonnegative trigonometric polynomial of order $p-1$. In particular, $T_{p-1}(\omega)$ is not an even function of the frequency ω. As a result, we redefine

$$S_1 := \sum_{j=r+1}^{n} T_{p-1}(\omega_{\sigma(j)}), \qquad S_2 := \sum_{j=1}^{r} T_{p-1}(\omega_{\sigma(j)}).$$

In this case, $S_2 = 1 - S_1$, instead of $S_2 = \frac{1}{2} - S_1$ in the original proof. Furthermore, instead of (4.27), we now have (by the same theorem)

$$\int_0^{2\pi}|T'_{p-1}(\omega)|\,\mathrm{d}\omega \leq (p-1)\int_0^{2\pi}|T_{p-1}(\omega)|\,\mathrm{d}\omega = (p-1)\frac{2\pi}{n}.$$

This modification strengthens the final inequality of the proof, (4.30), yielding

$$S_1 \geq \frac{m-1}{m}\left(\frac{1}{m-1} - \frac{(p-1)\pi}{n}\right).$$

Similarly, in the proof of Theorem 2.6, we set

$$S_{11} := \sum_{j=r_1+1}^{n} T_{p-1}(\omega_{\sigma(j)}), \qquad S_{21} := 1 - S_{11}.$$

In place of (4.29), we now have

$$\lambda_p > \frac{r_1}{n} S_{11},$$

and in place of (4.30),

$$T_{p-1}(\bar{\omega} + \eta) \geq \frac{1 - S_{11}}{r_1} \cos^2 \frac{p\eta}{2}, \qquad \text{for all } |\eta| \leq \frac{\pi}{p}.$$

As in the original proof, these relations lead to the final inequality

$$\lambda_p > C \, \frac{r_1}{N},$$

for some positive constant C.

References

[1] R. Adamczak, *A few remarks on the operator norm of random Toeplitz matrices*, Journal of Theoretical Probability **23** (2010), 85–108. MR2591905

[2] R. Adamczak, *On the operator norm of random rectangular Toeplitz matrices*, Progress in Probability, vol. 66, Springer, Basel, 2013. MR3443504

[3] K. Adhikari and K. Saha, *Fluctuations of eigenvalues of patterned random matrices*, Journal of Mathematical Physics **58** (2017). MR3666201

[4] K. Adhikari and K. Saha, *Universality in the fluctuation of eigenvalues of random circulant matrices*, Statistics and Probability Letters **138** (2018), 1–8. MR3788711

[5] T.W. Anderson, *The Statistical Analysis of Time Series*, Wiley, New York, 1971. MR0283939

[6] Alex H. Barnett, *How exponentially ill-conditioned are contiguous submatrices of the fourier matrix?*, SIAM Review **64** (2022), no. 1, 105–131. MR4373870

[7] G. Barrera and P. Manrique-Mirón, *The asymptotic distribution of the condition number for random circulant matrices*, Extremes **25** (2022), 567–594. MR4505180

[8] K.N. Berk, *Consistent autoregressive spectral estimates*, Annals of Statistics **2** (1974), no. 3, 489–502. MR0421010

[9] R.J. Bhansali, *Linear prediction by autoregressive model fitting in the time domain*, Annals of Statistics **6** (1978), 224–231. MR0461824

[10] R.J. Bhansali, *Asymptotically efficient autoregressive model selection for multistep prediction*, Ann. Inst. Statist. Math. **48** (1996), 577–602. MR1424784

[11] C. Bordenave and D. Chafaï, *Around the circular law*, Probability Surveys **9** (2012), 1–89. MR2908617

[12] A. Bose, *Patterned Random Matrices*, CRC Press, Boca Raton, FL, 2018. MR3823788

[13] A. Bose, S. Gangopadhyay, and A. Sen, *Limiting spectral distribution of XX' matrices*, Annales de l'Institut Henri Poincaré - Probabilités et Statistiques **46** (2010), 677–707. MR2682263

[14] A. Bose and J. Mitra, *Limiting spectral distribution of a special circulant*, Statistics and Probability Letters **60** (2002), 111–120. MR1945684

[15] A. Bose and K. Saha, *Random Circulant Matrices*, Chapman and Hall/CRC, 2018. MR3838430

[16] A. Bose and A. Sen, *Spectral norm of random large dimensional noncentral toeplitz and hankel matrices*, Electronic Communications in Probability **12** (2007), 29–35. MR2284045

[17] A. Bose and P. Sen, *$XX^\top$ matrices with independent entries*, ALEA- Latin American Journal of Probability and Mathematical Statistics **20** (2023), 75–125. MR4554226

[18] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration Inequalities: A nonasymptotic Theory of Independence*, Oxford university press, Oxford, UK, 2013. MR3185193

[19] W. Bryc, A. Dembo, and T. Jiang, *Spectral measure of large random Hankel, Markov and Toeplitz matrices*, The Annals of Probability **34** (2006), 1–38. MR2206341

[20] S. Chatterjee, *Fluctuations of eigenvalues and second order Poincaré inequalities*, Probability Theory and Related Fields **143** (2009), 1–40. MR2449121

[21] R.A. DeVore and G.G. Lorentz, *Constructive Approximations*, Springer-Verlag, Berlin, 1993. MR1261635

[22] D.L. Donoho and P.B. Stark, *Uncertainty principle and signal recovery*, SIAM Journal of Applied Mathematics **49** (1989), no. 3, 906–931. MR0997928

[23] G.H. Golub and C.F. VanLoan, *Matrix Computations*, The John Hopkins University Press, Baltimore and London, 1996. MR1417720

[24] Allan Gut, *Probability: a Graduate Course*, 2nd ed., Springer, New York, NY, 2013. MR2977961

[25] C. Hammond and S. J. Miller, *Distribution of eigenvalues for the ensemble of real symmetric toeplitz matrices*, Journal of Theoretical Probability **18** (2005), no. 3, 537–566. MR2167641

[26] Ching-Kang Ing and Ching-Zong Wei, *On same realization prediction in an infinite-order autoregressive process*, Journal of Multivariate Analysis **85** (2003), 130–155. MR1978181

[27] Ching-Kang Ing and Ching-Zong Wei, *Order selection for same-realization predictions in autoregressive processes*, The Annals of Statistics **33** (2005), 2423–2474. MR2211091

[28] A.S. Kiran Kumar, S. N. Maurya and K. Saha, *Linear eigenvalue statistics of XX' matrices*, Journal of Mathematical Physics **64** (2023), no. 12. MR4679204

[29] P. Kokoszka and T. Mikosch, *The periodogram at the Fourier frequencies*, Stochastic Processes and their Applications **86** (2000), 49–79. MR1741196

[30] D.-Z. Liu and Z.-D. Wang, *Limit distribution of eigenvalues for random Hankel and Toeplitz band matrices*, Journal of Theoretical Probability **24** (2011), no. 4, 988–1001. MR2851241

[31] D.-Z. Liu, X. Sun, and Z.-D. Wang, *Fluctuations of eigenvalues for random Toeplitz and related matrices*, Electronic Journal of Probability **17** (2012), no. 95. MR2994843

[32] P. Manrique-Mirón, *Random Toeplitz matrices: The condition number under high stochastic dependence*, Random Matrices: Theory and Applications **11** (2022), no. 3. MR4440253

[33] P. Manrique-Mirón, *Condition number of random tridiagonal Toeplitz matrix*, arXiv:2305.11971v1 (2023).

[34] Shambhu Nath Maurya, *Limiting spectral distribution of toeplitz and hankel matrices with dependent entries*, Statistics & Probability Letter **209** (2024), 110092. MR4711202

[35] E. Mayer-Wolf and O. Zeitouni, *The probability of small Gaussian ellipsoids and associated conditional moments*, The Annals of Probability **21** (1993), no. 1, 14–24. MR1207213

[36] M. W. Meckes, *On the spectral norm of a random Toeplitz matrix*, Electronic Communications in Probability **12** (2007), 315–325. MR2342710

[37] M. W. Meckes, *Some results on random circulant matrices*, in IMS Collections, High Dimensional Probability V: The Luminy Volume, **5** (2009), 213–223. MR2797949

[38] F. W. J. Olver, *Asymptotics and Special Functions*, A K Peters, Ltd., CRC Press, 1997. MR1429619

[39] V. Y. Pan, G. Qian, and A.-L. Zheng, *Randomized preprocessing versus pivoting*, Linear Algebra and Its Applications **438** (2013), 1883–1899. MR3005262

[40] V. Y. Pan, J. Svadlenka, and L. Zhao, *Estimating the norms of random circulant and Toeplitz matrices and their inverses*, Linear Algebra and Its Applications **468** (2015), 197–210. MR3293250

[41] H. Rauhut, *Circulant and Toeplitz Matrices in Compressed Sensing*, In Proc. SPARS'09 (2009).

[42] A. Rényi, *On the theory of order statistics*, Acta Mathematica Academiae Scientiarum Hungaricae **4** (1953), 191–231. MR0061792

[43] A. Sen and B. Virág, *The top eigenvalue of the random Toeplitz matrix and the sine kernel*, The Annals of Probability **41** (2013), 4050–4079. MR3161469

[44] R. Shibata, *Asymptotically efficient selection of the order of the model for estimating parameters of a linear process*, Annals of Statistics **8** (1980), 147–164. MR0557560

[45] J. W. Silverstein, *The smallest eigenvalue of a large dimensional Wishart matrix*, The Annals of Probability **13** (1985), no. 4, 1364–1368. MR0806232

[46] D. Slepian, *Prolate spheroidal wave functions. fourier analysis, and uncertainty — V: The discrete case*, The Bell System Technical Journal **57** (1978), no. 5, 1371–1430. MR0181766

[47] S.B. Stechkin, *Generalizations of some inequalities of S. N. Bernstein*, **60** Dokl. AS USSR (1948), no. 9, 1511–1514 (In Russian). MR0024993

[48] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values*, Geometric and Functional Analysis **20** (2010), 260–297. MR2647142

[49] Roman Vershynin, *High-dimensional Probability*, Cambridge University Press, https://doi.org/10.1017/9781108231596, 2018. MR3837109

Acknowledgments. We are grateful to Cristina Toninelli and two anonymous referees for their constructive comments and suggestions on earlier versions of this paper.