The Annals of Probability
2013, Vol. 41, No. 6, 4050–4079
DOI: 10.1214/13-AOP863

THE TOP EIGENVALUE OF THE RANDOM TOEPLITZ MATRIX AND THE SINE KERNEL

BY ARNAB SEN[1] AND BÁLINT VIRÁG[2]

University of Minnesota and University of Toronto

We show that the top eigenvalue of an $n \times n$ random symmetric Toeplitz matrix, scaled by $\sqrt{2n \log n}$, converges to the square of the $2 \to 4$ operator norm of the sine kernel.

1. Introduction.

An $n \times n$ symmetric random Toeplitz matrix is given by

$$\mathbf{T}_n = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \\ a_1 & a_0 & a_1 & \cdots & a_{n-2} \\ \vdots & a_1 & a_0 & \ddots & \vdots \\ a_{n-2} & \vdots & \ddots & \ddots & a_1 \\ a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 \end{bmatrix} = \bigl((a_{|i-j|})\bigr)_{0 \le i, j \le n},$$

where $(a_i)_{0 \le i \le n-1}$ is a sequence of independent random variables. This article establishes the law of large numbers for the maximum eigenvalue of this matrix as $n \to \infty$. The study of deterministic Toeplitz operators has a rich theory. See the classical book by Grenander and Szegő (1984) or more recent works by Böttcher and Grudsky (2000) and Böttcher and Silbermann (1999, 2006). In contrast, the study of random Toeplitz matrices is a relatively new field of research. The question of establishing the limiting spectral distribution of random Toeplitz matrices with independent entries was first posed in the review paper by Bai (1999). The answer was given by Bryc, Dembo and Jiang (2006) using method of moments. Since then the study of asymptotic distribution of eigenvalues of Toeplitz matrices has attracted considerable attention; for example, see Bose and Sen (2008), Chatterjee (2009), Hammond and Miller (2005), Kargin (2009) and the references therein.

The problem of studying the maximum eigenvalue of random Toeplitz matrices is raised in Bryc, Dembo and Jiang (2006), Remark 1.3. Bose and Sen (2007) established the law of the large numbers for the spectral norm of Toeplitz matrix when the entries are i.i.d. with some *positive* mean and finite variance. But

Received September 2011; revised March 2013.
[1]Supported by the EPSRC Grant EP/G055068/1.
[2]Supported by the Canada Research Chair program and the NSERC Discovery Accelerator Grant.
MSC2010 subject classifications. Primary 60F25; secondary 60B20.
Key words and phrases. Random Toeplitz matrices, maximum eigenvalue, spectral norm, sine kernel.

pinpointing the exact limit for the spectral norm of the Toeplitz matrix when there is no perturbation, that is, when the entries are mean zero, turned out to be much more challenging. This is partly due to the fact that, unlike the Wigner case where the limiting spectral distribution, the semicircular law has a compact support, and the top eigenvalue converges to the right endpoint of the support [this was proved by Bai and Yin (1988)], the limiting spectral distribution of Toeplitz matrices has infinite support. As a result, there is no natural guess to begin with. Another difficulty is that currently there are no useful estimates available for the trace of high powers of the Toeplitz matrix $\operatorname{tr}(\mathbf{T}_n^k)$ when $k = k(n)$ goes to infinity.

Meckes (2007) showed that if the entries have zero mean and uniformly subgaussian tail, then the expected spectral norm of an $n \times n$ random Toeplitz matrix is of the order of $\sqrt{n \log n}$, a significant departure from the standard $\sqrt{n}$ scaling of the Wigner case. Adamczak (2010) showed concentration, and more precisely, he proved that the spectral norm of random Toeplitz matrix normalized by the expected spectral norm converges almost surely to 1 if the entries are i.i.d. with zero mean and finite variance. Bose, Subhra Hazra and Saha (2010) gave an upper bound and a lower bound for the right tail probability of the spectral norm of Toeplitz matrix scaled by $n^{1/\alpha}$ when the entries are i.i.d. heavy-tailed random variables satisfying the following condition. There exist $p, q \geq 0$ with $p + q = 1$ and a slowly varying function $L(x)$ such that

$$\lim_{x\to\infty} \frac{\mathbb{P}\{X > x\}}{\mathbb{P}\{|X| > x\}} = p, \qquad \lim_{x\to\infty} \frac{\mathbb{P}\{X \leq -x\}}{\mathbb{P}\{|X| > x\}} = q$$

and

$$\mathbb{P}\{|X| > x\} \sim x^{-\alpha} L(x) \qquad \text{as } x \to \infty.$$

Throughout the paper, we will have the following standing assumption on the entries of our random Toeplitz matrix.

ASSUMPTION. For each n, $(a_i)_{0\leq i\leq n-1}$ is an array of independent real random variables (we suppress the dependence on n). There exists constants $\gamma > 2$ and C finite so that for each variable

$$\mathbb{E}a_i = 0, \qquad \mathbb{E}a_i^2 = 1 \quad \text{and} \quad \mathbb{E}|a_i|^\gamma < C.$$

Define the integral operator corresponding to the sine kernel by

$$\mathtt{Sin}(f)(x) := \int_{\mathbb{R}} \frac{\sin(\pi(x-y))}{\pi(x-y)} f(y)\,dy \qquad \text{for } f \in L^2(\mathbb{R}),$$

and its $2 \to 4$ operator norm as

$$\|\mathtt{Sin}\|_{2\to 4} := \sup_{\|f\|_2 \le 1} \|\mathtt{Sin}(f)\|_4,$$

where $\|f\|_p := (\int_{\mathbb{R}} |f(x)|^p)^{1/p}$ denotes the standard L^p-norm. $\|\mathtt{Sin}\|_{2\to 4}^2$ is the solution of the convolution optimization problem (see the Appendix),

$$\sup\{\|f \star f\|_2 : f \text{ even of } L^2\text{-norm 1 supported on } [-1/2, 1/2]\}. \tag{1}$$

For a Hermitian matrix $\mathbf{A}$, we denote by $\lambda_1(\mathbf{A})$ and $\lambda_n(\mathbf{A})$ the maximum and minimum eigenvalue of $\mathbf{A}$, respectively. The following theorem is the main result of our paper.

THEOREM 1. *Let $\mathbf{T}_n$ be a sequence of $n \times n$ symmetric random Toeplitz matrix as defined above with its entries satisfying the above assumption. Then*

$$\frac{\lambda_1(\mathbf{T}_n)}{\sqrt{2n \log n}} \xrightarrow{L^\gamma} \|\mathtt{Sin}\|_{2\to 4}^2 = 0.8288\ldots \qquad \text{as } n \to \infty.$$

Recall that a sequence of random variables converges in L^p to a constant c, denoted by $X_n \xrightarrow{L^p} c$, if $\mathbb{E}|X_n - c|^p \to 0$.

REMARK 2. Note that L^γ convergence is as best as we can hope for in Theorem 1. This is because of the fact that maximum eigenvalue of a symmetric matrix dominates the diagonal entries. So $\lambda_1(\mathbf{T}_n) \ge a_0$ and $\mathbb{E}|\lambda_1(\mathbf{T}_n)|^p$ can be infinite for any $p > \gamma$. Thus we cannot expect L^p convergence for $p > \gamma$.

By symmetry, the same theorem holds for $-\lambda_n$ and so for the spectral norm $\|\mathbf{T}_n\|_{\mathrm{sp}} = \max(\lambda_1, -\lambda_n)$ as well.

1.1. *Connection between Toeplitz and circulant matrices.* The starting point our analysis of the maximum eigenvalue is a connection between a Toeplitz matrix and a circulant matrix twice its size.

Observe that $\mathbf{T}_n$ is the $n \times n$ principal submatrix of a $2n \times 2n$ circulant matrix $\mathbf{C}_{2n} = (b_{j-i \bmod 2n})_{0 \le i,j \le 2n-1}$, where $b_j = a_j$ for $0 \le j < n$ and $b_j = a_{2n-j}$ for $n < j < 2n$ (choice of b_n is irrelevant at this point and it will be set later). We hope to relate the spectrum of Toeplitz matrix to that of the present circulant matrix twice its size, which can be easily diagonalized as follows:

$$(2n)^{-1/2}\mathbf{C}_{2n} = \mathbf{U}_{2n}\mathbf{D}_{2n}^{\dagger}\mathbf{U}_{2n}^{*},$$

where $\mathbf{U}_{2n}$ is the discrete Fourier transform, that is, a unitary matrix given by

$$\mathbf{U}_{2n}(j,k) = \frac{1}{\sqrt{2n}} \exp\left(\frac{2\pi i jk}{2n}\right), \qquad 0 \le j, k \le 2n-1,$$

and $\mathbf{D}_{2n}^{\dagger}$ is a diagonal matrix with

$$
\begin{aligned}
(\mathbf{D}_{2n}^{\dagger})_{j,j} &= \frac{1}{\sqrt{2n}} \sum_{k=0}^{2n-1} b_k \exp\left(\frac{2\pi i j k}{2n}\right) \\
&= \frac{1}{\sqrt{2n}} \left[a_0 + (-1)^j b_n + 2 \sum_{k=1}^{n-1} a_k \cos\left(\frac{2\pi j k}{2n}\right) \right].
\end{aligned} \tag{2}
$$

Clearly the j and $2n - j$ entries of $\mathbf{D}_{2n}^{\dagger}$ agree for all $n < j < 2n$. If we write

$$
\mathbf{Q}_{2n} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{pmatrix},
$$

then $\mathbf{T}_n$ and $\mathbf{Q}_{2n}\mathbf{C}_{2n}\mathbf{Q}_{2n}$ have the same nonzero eigenvalues by our observation. Moreover, the matrix $(2n)^{-1/2}\mathbf{Q}_{2n}\mathbf{C}_{2n}\mathbf{Q}_{2n}$ has the same eigenvalues as its conjugate

$$
(2n)^{-1/2}\mathbf{U}_{2n}^{*}\mathbf{Q}_{2n}\mathbf{C}_{2n}\mathbf{Q}_{2n}\mathbf{U}_{2n} = \mathbf{P}_{2n}\mathbf{D}_{2n}^{\dagger}\mathbf{P}_{2n},
$$

where

$$
\mathbf{P}_{2n} := \mathbf{U}_{2n}^{*}\mathbf{Q}_{2n}\mathbf{U}_{2n}. \tag{3}
$$

Consequently, we have a useful representation of the maximum eigenvalue of the Toeplitz matrix

$$
\lambda_1(n^{-1/2}\mathbf{T}_n) = \sqrt{2}\lambda_1(\mathbf{P}_{2n}\mathbf{D}_{2n}^{\dagger}\mathbf{P}_{2n}) \tag{4}
$$

as long as the right-hand side is not zero. We point out here that the matrix $\mathbf{Q}_{2n}\mathbf{C}_{2n}\mathbf{Q}_{2n}$ (and so $\mathbf{P}_{2n}\mathbf{D}_{2n}^{\dagger}\mathbf{P}_{2n}$) does not depend on the value of b_n, so we may replace it with an independent copy a_n of a_0. In addition, as we will show in Lemma 5 we can replace a_0 by $\sqrt{2}a_0$ in D_{2n} without changing the asymptotics. Dropping the subscript $2n$, we will study the matrix **PDP** as before and the entries of the diagonal matrix $\mathbf{D} = \mathrm{diag}(d_0, d_1, \ldots, d_{2n-1})$ given by

$$
d_j = \frac{1}{\sqrt{2n}} \left[\sqrt{2}a_0 + (-1)^j \sqrt{2}a_n + 2 \sum_{k=1}^{n-1} a_k \cos\left(\frac{2\pi j k}{2n}\right) \right], \qquad 0 \le j < 2n.
$$

The reason for choosing the "right" variance for diagonal and the "right" auxiliary variable b_n is that now the variables $d_j, 0 \le j \le n$ become uncorrelated; see Lemma 18. Thus in the special case when $\{a_j : 0 \le j \le n\}$ are i.i.d. Gaussian random variables with mean 0 and variance 1, it follows that $\{d_j : 0 \le j \le n\}$ are again independent Gaussian with mean 0 and have variance 1 except for d_0 and d_n, which have variance 2.

We have thus reduced our problem to studying the maximum eigenvalue of the matrix **PDP** where **P** is a deterministic Hermitian projection operator, and **D** is a random multiplication operator with defined on $\mathbb{C}^{2n}$ with uncorrelated entries. One can view this representation as a discrete analogue for Toeplitz operators defined on the Hardy space $\mathcal{H}^2$.

In the general case, to analyze the lower bound, we need to use an invariance principle. The difficulty is that the top eigenvalue does not come from the usual central limit theorem regime of $\mathbf{D}$, but from moderate deviations. We overcome this by extending the invariance principle of Chatterjee (2005) (based on Lindeberg's approach to the CLT) to the realm of moderate deviations.

1.2. *Heuristics and conjectures.* We first give a heuristic description of the origin of the limiting constant. Since $\mathbf{P}$ is a convolution with a decaying function, the matrix $\mathbf{PDP}$ in many ways behaves a like the diagonal matrix $\mathbf{D}$ itself. In particular, its top eigenvalue comes from a few nearby extreme values of $\mathbf{D}$. We can partition of the interval into short enough sections J, and show that the top eigenvalue is close to the maximum top eigenvalue of blocks $\mathbf{P}[J]\mathbf{D}[J]\mathbf{P}[J]$. This likes to be large when the entries of $\mathbf{D}[J]$ are large. By large deviations theory, the real constraint on the best $\mathbf{D}[J]$ is that the ℓ^2-norm the vector of entries is bounded by $\sqrt{2\log n}$.

Now the top eigenvalue is an ℓ^2 optimization problem, but now we have another ℓ^2 optimization over the entries of $\mathbf{D}[J]$. Thus an ℓ^4-norm appears, and that the solution of these two problems together are asymptotically given by the $2 \to 4$ norm of the limiting operator of $\mathbf{P}_{2n}$, which we then relate to the well-known sine kernel. The relation is natural, since $\mathbf{P}_{2n}$ and the sine kernel are both projections to an interval conjugated by a Fourier transform.

Before we prove Theorem 1, let us state some conjectures and open questions. Each of these conjectures can be split into parts (a) the Gaussian case, and (b) the general case where suitable moment conditions have to be imposed.

CONJECTURE 1. *Let v_n be the top eigenvector of $\mathbf{PD}^\dagger\mathbf{P}$. Then there exist random integers K_n so that for each $i \in \mathbb{Z}$, we have $v_n(K_n + i) \to \hat{g}(i)$, where $\hat{g}$ is the Fourier transform of the function $g(x) = \sqrt{2} f(2x - 1/2)$, and f is the* (*unique*) *optimizer in* (1).

CONJECTURE 2. *With high probability, all eigenvectors of $\mathbf{PD}^\dagger\mathbf{P}$ are localized: for each eigenvector, there exists a set of size $n^{o(1)}$ that supports $1 - o(1)$ proportion of the ℓ^2-norm.*

QUESTION 3. *What is the behavior of $n^{-1/2}\lambda_1(\mathbf{T}_n) - \sqrt{2\log n}\|\mathtt{Sin}\|_{2\to 4}^2$?*

CONJECTURE 4. *The top of the spectrum of $\mathbf{T}_n$, suitably shifted, and normalized, converges to a Poisson process with intensity $ce^{-\eta x}$ for some $c, \eta > 0$.*

Related to this, we have the following.

CONJECTURE 5. *The top eigenvalue of* $\mathbf{T}_n$, *suitably shifted and normalized*, *has a limiting Gumbel distribution*.

CONJECTURE 6. *The eigenvalue process of* $\mathbf{T}_n$, *away from the edge*, *after suitable normalization*, *converges to a standard Poisson point process on* $\mathbb{R}$.

1.3. *The proof of Theorem* 1. In this section, we break the proof of Theorem 1 into its components. This also serves as a guide to the rest of the paper.

PROOF OF THEOREM 1. Consider the discrete Fourier transform matrix,

$$\mathbf{U}(j,k) = \frac{1}{\sqrt{2n}} \exp\left(\frac{2\pi i jk}{2n}\right), \qquad 0 \le j,k \le 2n-1$$

and $\mathbf{D}^{\dagger}$, a $2n \times 2n$ diagonal matrix so that for $j = 0, \dots, 2n-1$

$$\begin{aligned} \mathbf{D}^{\dagger}(j,j) &= \frac{1}{\sqrt{2n}} \sum_{k=0}^{2n-1} b_k \exp\left(\frac{2\pi i jk}{2n}\right) \\ &= \frac{1}{\sqrt{2n}} \left[a_0 + (-1)^j b_n + 2\sum_{k=1}^{n-1} a_k \cos\left(\frac{2\pi jk}{2n}\right)\right]. \end{aligned} \tag{5}$$

Let

$$\mathbf{Q} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{0}_n \end{pmatrix}, \qquad \mathbf{P} := \mathbf{U}^* \mathbf{Q} \mathbf{U}.$$

Then, as argued in Section 1.1 we have the representation (4)

$$\lambda_1(n^{-1/2}\mathbf{T}_n) = \sqrt{2}\lambda_1(\mathbf{P}\mathbf{D}^{\dagger}\mathbf{P})$$

as long as the right-hand side is positive. In formula (5) for diagonal entries of $\mathbf{D}^{\dagger}$, we replace a_0 by $\sqrt{2}a_0$ (this is legal via Lemma 5) and choose $b_n = \sqrt{2}a_n$ where a_n is an identical copy of a_0 independent of $(a_i)_{0 \le i < n}$ (recall $\mathbf{P}\mathbf{D}^{\dagger}\mathbf{P}$ does not depend on b_n) to obtain a new diagonal matrix $\mathbf{D} = \operatorname{diag}(d_0, d_1, \dots, d_{2n-1})$ given by

$$d_j = \frac{1}{\sqrt{2n}} \left[\sqrt{2}a_0 + (-1)^j \sqrt{2}a_n + 2\sum_{k=1}^{n-1} a_k \cos\left(\frac{2\pi jk}{2n}\right)\right], \qquad 0 \le j < 2n.$$

The reason for choosing the "right" variance for diagonal and the "right" auxiliary variable b_n is that now the variables $d_j, 0 \le j \le n$ become uncorrelated; see Lemma 18. Thus in the special case when $\{a_j : 0 \le j \le n\}$ are i.i.d. Gaussian random variables with mean 0 and variance 1, it follows that $\{d_j : 0 \le j \le n\}$ are again independent Gaussian with mean 0 and have variance 1 except for d_0 and d_n who have variance 2.

In Lemma 5 we justify working with $\mathbf{PD}^\dagger\mathbf{P}$ by showing that

$$\mathbb{E}\big|\lambda_1(\mathbf{PDP}) - \lambda_1(\mathbf{PD}^\dagger\mathbf{P})\big|^\gamma = o(\sqrt{\log n}).$$

In Section 2, Corollary 6 we show that we can assume that the a_n are bounded by $n^{1/\gamma}$. In Lemma 7, we establish tightness, so it suffices to show convergence in probability. In Section 3, equation (16) we introduce a sparse version $\mathbf{D}^\varepsilon$ of the diagonal matrix $\mathbf{D}$, by considering the set

$$S = \{0 \le j \le 2n-1 : |d_j| \ge \varepsilon\sqrt{2\log n}\},$$

and setting $(\mathbf{D}^\varepsilon)_{jj} = d_j \mathbf{1}_{j\in S}$. Then we show that the eigenvalues are close,

$$\big|\lambda_1(\mathbf{PDP}) - \lambda_1(\mathbf{PD}^\varepsilon\mathbf{P})\big| \le \varepsilon\sqrt{2\log n}.$$

The matrix $\mathbf{D}^\varepsilon$ is sparse enough that the whole question can be reduced to a block-diagonal version, where the blocks are determined by a random partition Λ of $\{1,\dots,2n\}$. This is done in Lemma 9: with high probability,

$$\Big|\lambda_1(\mathbf{PD}^\varepsilon\mathbf{P}) - \max_{J\in\Lambda : J\cap S\neq\varnothing} \lambda_1(\mathbf{P}[J]\mathbf{D}^\varepsilon[J]\mathbf{P}[J])\Big| = O(1),$$

where $\mathbf{P}[J]$ refers to the minor of $\mathbf{P}$ corresponding to the subset of indices J. This is guaranteed by a careful choice of partition Λ, which ensures that the blocks J with $J\cap S\neq\varnothing$ are sufficiently far apart, so that the interaction between them is negligible. This interaction comes from off-diagonal elements of the matrix $\mathbf{P}$, whose entries decay with the distance from the diagonal.

The main idea of the last part of the proof is explained in Section 1.2. We proceed along those lines. In Propositions 10 and 14 (Sections 4, 5) we then establish the asymptotic upper and lower bounds for the for the block diagonal version. Together, they give that with high probability, there exists $\varepsilon_n \to 0$ so that

$$\max_{J\in\Lambda : J\cap S\neq\varnothing} \lambda_1(\mathbf{P}[J]\mathbf{D}^{\varepsilon_n}[J]\mathbf{P}[J]) = \sqrt{2\log n}\big(\|\Pi\|^2_{2\to4} + o(1)\big),$$

where Π is the $n\to\infty$ limit of $\mathbf{P}$ introduced in (14). Finally, in the Appendix, Lemma 19, we identify $\|\Pi\|^2_{2\to4} = \frac{1}{\sqrt{2}}\|\mathtt{Sin}\|^2_{2\to4}$. This completes the proof of Theorem 1. $\square$

1.4. *Notation*. We write that a sequence of events $(\mathcal{E}_n)_{n\ge1}$ occurs *with high probability* when $\mathbb{P}\{\mathcal{E}_n\}\to 1$. Let $\ell^2(\mathbb{C})$ [resp., $\ell^2(\mathbb{R})$] be the space of square summable sequences of complex (resp., real) numbers indexed by $\mathbb{Z}$. For square matrix $\mathbf{A}$ and a subset T of the index set, we denote by $\mathbf{A}[T]$, the principal submatrix of $\mathbf{A}$ which is obtained by keeping those rows and columns of $\mathbf{A}$ whose indices belong to T. We consider n as an asymptotic parameter tending to infinity. We will use the notation $f(n) = \Omega(g(n))$ or $g(n) = O(f(n))$ to denote the bound $g(n) \le Cf(n)$ for all sufficiently large n and for some constant C. Notation such as $f(n) = \omega(g(n))$ or $g(n) = o(f(n))$ means that $g(n)/f(n)\to 0$ as $n\to\infty$. We write $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $g(n) = O(f(n))$ hold.

2. Truncation and tightness.

2.1. *Truncation and changing the diagonal term.* Let n_0 be sufficiently large number. For $n \geq n_0$, define two arrays of truncated random variables by

$$\tilde{a}_i = \tilde{a}_i^{(n)} =: a_i \mathbf{1}_{\{|a_i| \leq n^{1/\gamma}\}} - \mathbb{E}[a_i \mathbf{1}_{\{|a_i| \leq n^{1/\gamma}\}}], \qquad 0 \leq i \leq n-1$$

and

$$\bar{a}_i = \bar{a}_i^{(n)} := \operatorname{Var}(\tilde{a}_i)^{-1/2} \tilde{a}_i, \qquad 0 \leq i \leq n-1.$$

Note that to define $\bar{a}_i^{(n)}$, we need that $\operatorname{Var}(\tilde{a}_i^{(n)}) > 0$, which holds for sufficiently large n.

We sometimes write $\mathbf{T}_n(a)$ for $\mathbf{T}_n$ to emphasize its dependence on the underlying sequence of random variables $(a_i)_{0 \leq i \leq n-1}$. Thus $\mathbf{T}_n(\tilde{a})$ and $\mathbf{T}_n(\bar{a})$ denote the Toeplitz matrices built with random variables $(\tilde{a}_i)_{0 \leq i \leq n-1}$ and $(\bar{a}_i)_{0 \leq i \leq n-1}$, respectively.

The next lemma says that the above truncation and the rescaling of the underlying random variables has a negligible effect in the study of the maximum eigenvalue of Toeplitz matrix.

LEMMA 3. *We have, as $n \to \infty$,*

(a)

$$\frac{\lambda_1(\mathbf{T}_n(a)) - \lambda_1(\mathbf{T}_n(\tilde{a}))}{\sqrt{n \log n}} \xrightarrow{L^\gamma} 0,$$

(b)

$$\frac{\lambda_1(\mathbf{T}_n(\tilde{a})) - \lambda_1(\mathbf{T}_n(\bar{a}))}{\sqrt{n \log n}} \xrightarrow{L^\gamma} 0.$$

PROOF. Define, for $n \geq n_0$,

$$\hat{a}_i := a_i \mathbf{1}_{\{|a_i| \leq n^{1/\gamma}\}}, \qquad 0 \leq i \leq n-1.$$

Recall that for a matrix $\mathbf{A}$, its spectral norm satisfies

$$\|\mathbf{A}\|_{\mathrm{sp}}^2 \leq \max_k \sum_l |\mathbf{A}(k,l)| \times \max_l \sum_k |\mathbf{A}(k,l)|. \tag{6}$$

In the special case when $\mathbf{A}$ is Hermitian, the above bound reduces to

$$\|\mathbf{A}\|_{\mathrm{sp}} \leq \max_k \sum_l |\mathbf{A}(k,l)|. \tag{7}$$

Then we have

$$\begin{aligned} \left| \frac{\lambda_1(\mathbf{T}_n(a)) - \lambda_1(\mathbf{T}_n(\hat{a}))}{\sqrt{n \log n}} \right| &\leq \frac{\|\mathbf{T}_n(a) - \mathbf{T}_n(\hat{a})\|_{\mathrm{sp}}}{\sqrt{n \log n}} \\ &\leq \frac{2 \sum_{i=0}^{n-1} |a_i| \mathbf{1}_{\{|a_i| > n^{1/\gamma}\}}}{\sqrt{n \log n}}, \end{aligned} \tag{8}$$

which follows from bound (7) and from the fact that the ℓ^1 norm of the each row of the Toeplitz matrix $\mathbf{T}_n(a) - \mathbf{T}_n(\hat{a})$ is bounded by $2\sum_{i=1}^n |a_i|\mathbf{1}_{\{|a_i|>n^{1/\gamma}\}}$. We quote a standard moment bound for sum of independent nonnegative random variables, commonly known as Rosenthal's inequality in the literature [see, e.g., Latała (1997), Corollary 3] which says that if $\xi_1, \dots, \xi_n$ are independent nonnegative random variables and $p \geq 1$, then there exists a universal constant C_p such that

$$\mathbb{E}\left[\left(\sum_{i=1}^n \xi_i\right)^p\right] \leq C_p \max\left(\left(\sum_{i=1}^n \mathbb{E}[\xi_i]\right)^p, \sum_{i=1}^n \mathbb{E}[\xi_i^p]\right).$$

Clearly, $\mathbb{E}[|a_i|^\gamma \mathbf{1}_{\{|a_i|>n^{1/\gamma}\}}] \leq \mathbb{E}[|a_i|^\gamma] \leq C$. On the other hand, by Hölder's inequality,

$$\mathbb{E}\big[|a_i|\mathbf{1}_{\{|a_i|>n^{1/\gamma}\}}\big] \leq \big(\mathbb{E}|a_i|^\gamma\big)^{1/\gamma} \cdot \big(\mathbb{P}\{|a_i|^\gamma > n\}\big)^{1-1/\gamma},$$

and by Markov's inequality, this is bounded above by

$$\big(\mathbb{E}|a_i|^\gamma\big)^{1/\gamma}\big(\mathbb{E}|a_i|^\gamma/n\big)^{1-1/\gamma} = \mathbb{E}\big[|a_i|^\gamma\big]n^{-1+1/\gamma} \leq Cn^{-1+1/\gamma}. \tag{9}$$

Therefore, by Rosenthal's inequality,

$$\mathbb{E}\left[\sum_{i=0}^{n-1} |a_i|\mathbf{1}_{\{|a_i|>n^{1/\gamma}\}}\right]^\gamma = O(n). \tag{10}$$

Combining (8) and (10), we obtain

$$\mathbb{E}\left|\frac{\lambda_1(\mathbf{T}_n(a)) - \lambda_1(\mathbf{T}_n(\hat{a}))}{\sqrt{n\log n}}\right|^\gamma = \frac{O(n)}{(n\log n)^{\gamma/2}} \to 0.$$

Next we see that

$$\begin{aligned}\mathbb{E}\left|\frac{\lambda_1(\mathbf{T}_n(\hat{a})) - \lambda_1(\mathbf{T}_n(\tilde{a}))}{\sqrt{n\log n}}\right|^\gamma &\leq \frac{\mathbb{E}[\|\mathbf{T}_n(\hat{a}) - \mathbf{T}_n(\tilde{a})\|_{\mathrm{sp}}^\gamma]}{(n\log n)^{\gamma/2}} \\ &\leq \frac{(2\sum_{i=0}^{n-1}\mathbb{E}|a_i|\mathbf{1}_{\{|a_i|>n^{1/\gamma}\}})^\gamma}{(n\log n)^{\gamma/2}} = \frac{O(n)}{(n\log n)^{\gamma/2}} \to 0,\end{aligned}$$

which completes the proof of part (a).

For part (b) we want to bound

$$\left|\frac{\lambda_1(\mathbf{T}_n(\tilde{a})) - \lambda_1(\mathbf{T}_n(\bar{a}))}{\sqrt{n\log n}}\right| \leq \frac{\|\mathbf{T}_n(\tilde{a}) - \mathbf{T}_n(\bar{a})\|_{\mathrm{sp}}}{\sqrt{n\log n}} \leq \sqrt{2}\cdot\frac{\|\mathbf{P}(\mathbf{D}^\dagger(\tilde{a}) - \mathbf{D}(\bar{a}))\mathbf{P}\|_{\mathrm{sp}}}{\sqrt{\log n}}.$$

Here we use representation (4). Since $\mathbf{P}$, being Hermitian projection matrix, has spectral norm equal to one, with $a^* = \tilde{a} - \bar{a}$ we have

$$\big\|\mathbf{P}\big(\mathbf{D}^\dagger(\tilde{a}) - \mathbf{D}(\bar{a})\big)\mathbf{P}\big\|_{\mathrm{sp}} = \big\|\mathbf{P}\mathbf{D}^\dagger(a^*)\mathbf{P}\big\|_{\mathrm{sp}} \leq \big\|\mathbf{D}^\dagger(a^*)\big\|_{\mathrm{sp}} = \max_{0\leq j\leq n}\big|d_j(a^*)\big|.$$

We have

$$1 - \mathrm{Var}(\tilde{a}_i) = \mathbb{E}\big[a_i^2 \mathbf{1}_{\{|a_i| > n^{1/\gamma}\}}\big] + \big(\mathbb{E}[a_i \mathbf{1}_{\{|a_i| > n^{1/\gamma}\}}]\big)^2.$$

We apply (9), and the similarly derived

$$\mathbb{E}\big[a_i^2 \mathbf{1}_{\{|a_i| > n^{1/\gamma}\}}\big] \le C n^{-1+2/\gamma}$$

to get $1 - \mathrm{Var}(\tilde{a}_i) = O(n^{-1+2/\gamma})$ uniformly in i. Note that $a_i^* = (1 - \mathrm{Var}(\tilde{a}_i)^{-1/2})\tilde{a}_i$ which implies $\mathbb{E}a_i^* = 0$ and $\mathrm{Var}(a_i^*) = (1 - \mathrm{Var}(\tilde{a}_i)^{1/2})^2 = O(n^{-2+4/\gamma})$ and $|a_i^*| \le 4n^{1/\gamma}$, for $n \ge n_0$.

Using the union bound and the identity $\mathbb{E}[X] = \int_0^\infty \mathbb{P}\{X > t\}\, dt$ which holds for any nonnegative random variable X, we can write

$$\begin{aligned}\mathbb{E}\Big[\max_{0\le j\le n} |d_j(a^*)|^\gamma\Big] &\le 1 + \int_1^\infty \mathbb{P}\Big\{\max_{0\le j\le n} |d_j(a^*)|^\gamma > t\Big\}\, dt \\ &\le 1 + (n+1) \max_{0\le j\le n} \int_1^\infty \mathbb{P}\{|d_j(a^*)| > t^{1/\gamma}\}\, dt.\end{aligned} \tag{11}$$

If $\xi_1, \xi_2, \dots, \xi_n$ be independent mean zero random variables, uniformly bounded by M, then the classical Bernstein inequality gives the following tail bound for their sum:

$$\mathbb{P}\Bigg\{\sum_{k=1}^n \xi_k > t\Bigg\} \le \exp\bigg(-\frac{t^2/2}{\sum_{k=1}^n \mathrm{Var}(\xi_k) + Mt/3}\bigg) \qquad \forall t > 0.$$

Using Bernstein's inequality, we obtain

$$\begin{aligned}(11) &\le 1 + (n+1) \int_1^\infty 2\exp\bigg(-\frac{t^{2/\gamma}(\sqrt{n})^2}{n \cdot O(n^{-1+2/\gamma}) + t^{1/\gamma}\sqrt{n}\, O(n^{1/\gamma})}\bigg)\, dt \\ &\le 1 + 2(n+1) \int_1^\infty \exp\big(-t^{1/\gamma} \cdot \Omega(n^{1/2-1/\gamma})\big)\, dt \\ &= 1 + n \cdot n^{1/\gamma - 1/2} e^{-\Omega(n^{1/2-1/\gamma})}.\end{aligned}$$

Hence, $(\log n)^{-1/2}\|\mathbf{D}(a^*)\| \xrightarrow{L^\gamma} 0$ and so, $\big|\frac{\lambda_1(\mathbf{T}_n(\tilde{a})) - \lambda_1(\mathbf{T}_n(\bar{a}))}{\sqrt{n\log n}}\big| \xrightarrow{L^\gamma} 0$. This completes the proof of part (b) of the lemma. $\square$

DEFINITION 4. Let $\mathbf{T}_n^\circ$ be the symmetric Toeplitz matrix which has $\sqrt{2}a_0$ on its diagonal instead of a_0.

LEMMA 5. *We have, as $n \to \infty$,*

$$\frac{\lambda_1(\mathbf{T}_n^\circ) - \lambda_1(\mathbf{T}_n)}{\sqrt{n \log n}} \xrightarrow{L^\gamma} 0.$$

PROOF. The proof is immediate from the following fact:

$$\mathbb{E}\|\mathbf{T}_n^\circ - \mathbf{T}_n\|_{\mathrm{sp}}^\gamma \le (\sqrt{2}-1)^\gamma \mathbb{E}\big[|a_0|^\gamma\big] = O(1). \qquad \square$$

COROLLARY 6. *It suffices to prove Theorem 1 for the symmetric random Toeplitz matrix $\mathbf{T}_n^\circ$ defined in Definition 4 where the random variables a_i are independent mean zero, variance one and bounded by $2n^{1/\gamma}$.*

PROOF. The proof is immediate from Lemmas 3 and 5. $\square$

Following (4), we can write

$$\lambda_1\big(n^{-1/2}\mathbf{T}_n^\circ\big) = \sqrt{2}\lambda_1(\mathbf{PDP}), \tag{12}$$

with $\mathbf{P}$ as before and the entries of the diagonal matrix $\mathbf{D} = \mathrm{diag}(d_0, d_1, \dots, d_{2n-1})$ given by

$$d_j = \frac{1}{\sqrt{2n}}\left[\sqrt{2}a_0 + (-1)^j\sqrt{2}a_n + 2\sum_{k=1}^{n-1} a_k \cos\left(\frac{2\pi jk}{2n}\right)\right], \qquad 0 \le j < 2n,$$

where $b_n := \sqrt{2}a_n$, a_n being an independent copy of a_0. The reason for choosing the "right" variance for diagonal and the "right" auxiliary variable b_n is that now the variables $d_j, 0 \le j \le n$ become uncorrelated; see Lemma 18. Thus in the special case when $\{a_j : 0 \le j \le n\}$ are i.i.d. Gaussian random variables with mean 0 and variance 1, it follows that $\{d_j : 0 \le j \le n\}$ are again independent Gaussian with mean 0 and have variance 1 except for d_0 and d_n which have variance 2.

2.2. *Tightness.*

LEMMA 7. *For each $n \ge 1$, let $a_0, a_1, \dots, a_{n-1}$ be a sequence of independent random variables that have mean zero, variance one and are bounded by $n^{1/\gamma}$. For any $p > 0$, we have*

$$\sup_{n\ge 1} \mathbb{E}\left[\left(\frac{\lambda_1(\mathbf{T}_n^\circ)}{\sqrt{2n\log n}}\right)^p\right] < \infty.$$

PROOF. The proof is a direct application of Bernstein's inequality similar to what we did in the proof of part (b) of Lemma 3. From representation (12), we know that $n^{-1/2}\lambda_1(\mathbf{T}_n^\circ)$ is bounded above by $\sqrt{2}\cdot\max_{0\le j\le n}|d_j|$. Therefore, for any $\alpha > 0$,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\lambda_1(\mathbf{T}_n^\circ)}{\sqrt{2n\log n}}\right)^p\right] &\le \mathbb{E}\left[\left(\frac{\max_{0\le j\le n}|d_j|}{\sqrt{\log n}}\right)^p\right] \\ &\le \alpha + (n+1)\max_{0\le j\le n-1}\int_\alpha^\infty \mathbb{P}\left\{\frac{|d_j|}{\sqrt{\log n}} > t^{1/p}\right\}dt.\end{aligned}$$

By the Bernstein inequality, for n sufficiently large,

$$\max_{0\le j\le n} \mathbb{P}\left\{\frac{|d_j|}{\sqrt{\log n}} > t^{1/p}\right\} \le 2\exp\left(-\frac{(1/2)\log n \cdot t^{2/p}}{2+(1/3)\sqrt{2}n^{1/\gamma-1/2}\sqrt{\log n}\cdot t^{1/p}}\right),$$

which implies that there exists a constant $c > 0$ such that for each n and each $0 \le j \le n$,

$$\int_\alpha^\infty \mathbb{P}\left\{\frac{|d_j|}{\sqrt{\log n}} > t^{1/p}\right\} dt \le \int_\alpha^{t_n} \exp(-ct^{2/p}\cdot \log n)\,dt + \int_{t_n}^\infty \exp(-ct^{1/p})\,dt,$$

where $t_n := n^{p(1/2-1/\gamma)}(\log n)^{-p/2}$. This particular choice of t_n is governed by the fact that $2+\frac{1}{3}\sqrt{2}n^{1/\gamma-1/2}\sqrt{\log n}\cdot t_n^{1/p} = O(1)$. The second integral above goes to zero faster than any polynomial power of n, whereas by choosing α sufficiently large we can make the first integral $O(n^{-1})$. The claim of the lemma follows. $\square$

3. Reduction to block diagonal form.

3.1. *Some facts about* $\mathbf{P}$. By definition (3) $\mathbf{P}:\mathbb{C}^{2n}\to\mathbb{C}^{2n}$ is a Hermitian projection matrix. The action of operator $\mathbf{P}$ can be described by the composition of the following three maps: For $x\in\mathbb{C}^{2n}$, we first take discrete Fourier transform of x, then project it to the first n Fourier frequencies and finally do the inverse discrete Fourier transform.

The entries of $\mathbf{P}$ are given by

$$\mathbf{P}(k,l) = \begin{cases} \dfrac{1}{2}, & \text{for } k=l,\\ 0, & \text{for } k\ne l,\ |k-l| \text{ is even},\\ \dfrac{1}{n}\times\dfrac{1}{1-\exp(-2\pi i(k-l)/(2n))}, & \text{for } |k-l| \text{ is odd}.\end{cases}$$

Note that $\mathbf{P}(k,l)$ is a function of $(k-l)$ only and that

$$|\mathbf{P}(k,l)| \le C_1/\min(|k-l|, 2n-|k-l|), \qquad k\ne l$$

for some constant C_1. Hence, the maximum of ℓ^1 norms of the rows or the columns of $\mathbf{P}$ has the following upper bound:

$$\max_k \sum_{l=0}^{2n-1} |\mathbf{P}(k,l)| \le C_2\log n, \qquad \max_l \sum_{k=0}^{2n-1} |\mathbf{P}(k,l)| \le C_2\log n, \tag{13}$$

where C_2 is some suitable constant.

Limiting operator for **P**. Let $\mathbb{T}$ be unit circle parametrized by $\mathbb{T} = \{e^{2\pi i x} : x \in (-1/2, 1/2]\}$ and $L^2(\mathbb{T}) := \{f : [-1/2, 1/2] \to \mathbb{C} : f(-1/2) = f(1/2)$ and $\int_{-1/2}^{1/2} |f(x)|^2\, dx < \infty\}$. We define an projection operator $\Pi : \ell^2(\mathbb{C}) \to \ell^2(\mathbb{C})$ as the composition of the following operators:

$$\Pi : \ell^2(\mathbb{C}) \xrightarrow{\psi} L^2(\mathbb{T}) \xrightarrow{\chi_{[0,1/2]}} L^2(\mathbb{T}) \xrightarrow{\psi^{-1}} \ell^2(\mathbb{C}), \tag{14}$$

where $\psi : \ell^2(\mathbb{C}) \to L^2(\mathbb{T})$ is the Fourier transform which sends the coordinate vector $e_m, m \in \mathbb{Z}$ to the periodic function $x \mapsto e^{2\pi i m x} \in L^2(\mathbb{T})$, ψ^{-1} is the inverse map of ψ and $\chi_{[0,1/2]} : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ is the projection map given by $f \mapsto f \times 1_{[0,1/2]}$. The operator Π is Hermitian and is defined on the entire $\ell^2(\mathbb{C})$ and hence is self-adjoint. It is easy to check that for any $k, l \in \mathbb{Z}$,

$$\langle e_k, \Pi e_l \rangle =: \Pi(k,l) = \begin{cases} \dfrac{1}{2}, & \text{if } k = l, \\ 0, & \text{if } k \neq l, |k-l| \text{ is even}, \\ \dfrac{-i}{\pi(k-l)}, & \text{if } |k-l| \text{ is odd}. \end{cases}$$

Here $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\ell^2(\mathbb{C})$.

Define $2 \to 4$ operator norm of Π as

$$\|\Pi\|_{2\to 4} := \sup\{\|\Pi \mathbf{v}\|_4 : \mathbf{v} \in \ell^2(\mathbb{C}), \|\mathbf{v}\|_2 \leq 1\},$$

where for any vector $\mathbf{v} \in \ell^2(\mathbb{C})$ and $p \geq 1$, $\|\mathbf{v}\|_p$ denotes the standard ℓ^p norm of $\mathbf{v}$.

We claim that

$$\left|\mathbf{P}_{2n}(k,l) - \Pi(k,l)\right| = O(n^{-1}) \qquad \text{as } n \to \infty \text{ and } |k-l| = o(n). \tag{15}$$

There is nothing to prove if $|k-l|$ is even. So assume that $|k-l|$ is odd and $|k-l| = o(n)$. In this case we can write $\mathbf{P}_{2n}(k,l) - \Pi(k,l)$ as $n^{-1} \times \frac{1}{x_n}[\frac{x_n}{1-e^{-ix_n}} + i]$ where $x_n = \pi(k-l)/n = o(1)$. Now (15) easily follows from the following limit, which is elementary:

$$\lim_{x\to 0} \frac{1}{x}\left[\frac{x}{1-e^{-ix}} + i\right] = \lim_{x \to 0} \frac{1 - \sin x/x + i(1-\cos x)/x}{(1-\cos x) + i \sin x} = \frac{1}{2}.$$

We will make use of (15) when we prove the upper bound for the top eigenvalue in Section 4.

3.2. *Allowing ε room.* As one might guess, the diagonal entries of **D** which have small absolute value should not have too much influence on determining the value of $\lambda_1(\mathbf{PDP})$. In this subsection, we will make this idea precise. For $\varepsilon > 0$, consider the random set

$$S = \{0 \leq j \leq 2n-1 : |d_j| \geq \varepsilon\sqrt{2\log n}\},$$

let $\mathbf{R} := \operatorname{diag}(1_{j\in S})$ and let $\mathbf{D}^{\varepsilon} := \mathbf{DR}$. Then

$$\begin{aligned}\left|\frac{\lambda_1(\mathbf{PDP})}{\sqrt{2\log n}} - \frac{\lambda_1(\mathbf{PD}^{\varepsilon}\mathbf{P})}{\sqrt{2\log n}}\right| &\le \frac{\|\mathbf{P}(\mathbf{D}-\mathbf{D}^{\varepsilon})\mathbf{P}\|_{\mathrm{sp}}}{\sqrt{2\log n}} \\ &\le \frac{\|\mathbf{P}\|_{\mathrm{sp}}\|\mathbf{D}-\mathbf{D}^{\varepsilon}\|_{\mathrm{sp}}\|\mathbf{P}\|_{\mathrm{sp}}}{\sqrt{2\log n}} \le \varepsilon.\end{aligned} \tag{16}$$

3.3. *Random partition of the interval.* For a set B, we denote by $\#B$ the cardinality of B.

Set $r = \lceil \log n \rceil^3$, and let $m = \lfloor n/2r \rfloor$. Divide the interval $\{0, 1, 2, \dots, 2n-1\}$ into $2m+1$ consecutive disjoint subintervals (called *bricks*) $L_{-m}, \dots, L_{-1}$, $L_0, L_1, \dots, L_m$ in such a way that:

- $0 \in L_{-m}$ and $n \in L_0$,
- the length of each L_i is between r and $4r$ and
- the subdivision is symmetric about n: $L_i = 2n - L_{-i}$ for $-m < i < m$ and $L_{-m} \setminus \{0\} = 2n - L_m$.

We define a *block* to be a (nonempty) union of consecutive bricks, say

$$J = L_j \cup L_{j+1} \cup \cdots \cup L_{k-1} \cup L_k,$$

with $-m \le j \le k \le m$. We fix $\varepsilon > 0$, and set $M = M(\varepsilon) := 4 + 12/\varepsilon^2$. We will call a block J *admissible* if:

(a) either $J \subseteq L_{-m+1} \cup L_{-m+2} \cup \cdots \cup L_{-1}$ or $J \subseteq L_1 \cup L_2 \cup \cdots \cup L_m$ and
(b) the number $1 + k - j$ of bricks that make up the block J is at most $M(\varepsilon)$.

The set of all *admissible blocks* will be denoted by $\mathcal{L}$. We mention here that the notion of admissible blocks depends on the fixed parameter ε. Notice that (a) is equivalent to requiring $L_{-m} \nsubseteq J$ and $L_0 \nsubseteq J$. Moreover, if $L \in \mathcal{L}$, then size of L is bounded below and above by r and $4Mr$, respectively.

We define a brick L_k to be *visible* if $L \cap S \neq \varnothing$. Otherwise L_k is called *invisible*. Clearly, for $1 \le k < m$, L_k is visible if and only if L_{-k} is visible. Given the random set S, partition $\{0, 1, 2, \dots, 2n-1\}$ into disjoint intervals J by subdividing between each pair of consecutive invisible bricks. Clearly, each such J is a block. Denote the collection of all those J's by Λ. Note that the Λ and the J's are at random. Note that in this random partition, each block $J \in \Lambda$ starts and ends with a "gap" of size at least $(\log n)^3$. More precisely, for any $[a, b] \in \Lambda$, we have $[a, a + (\log n)^3] \cap S = \varnothing$ (unless $a = 0$) and $[b - (\log n)^3, b] \cap S = \varnothing$ (unless $b = 2n-1$).

PROPOSITION 8. *For each $\varepsilon > 0$ and each division $\mathcal{L}$, the following holds with high probability: For each $J \in \Lambda$, if $J \cap S \neq \varnothing$, then:*

(1) *$J \in \mathcal{L}$ and*

(2) $\#(J \cap S) \leq M$ *for all* $J \in \mathcal{L}$.

PROOF. For any fixed $s \geq 1$ and $0 \leq j_1 < j_2 < \cdots < j_s \leq n$,

$$\begin{aligned}
&\mathbb{P}\{|d_{j_i}| > \varepsilon\sqrt{2\log n}, 1 \leq i \leq s\} \\
&\quad \leq \sum_{\beta_i \in \{-1,+1\}} \mathbb{P}\left\{\sum_{i=1}^{s} \beta_i d_{j_i} > s\varepsilon\sqrt{2\log n}\right\} \\
&\quad \leq 2^s \exp\left(-\frac{\varepsilon^2 s^2 \cdot \log n}{\operatorname{Var}(\sum_{i=1}^{s} \beta_i d_{j_i}) + O(sn^{1/\gamma - 1/2}) \cdot s\varepsilon\sqrt{2\log n}}\right) \\
&\quad = O(n^{-s\varepsilon^2/3}).
\end{aligned} \tag{17}$$

The second inequality in (17) is a consequence of Bernstein's inequality once we write $\sum_{i=1}^{s} \beta_i d_{j_i}$ as the linear sum of a_i's. Note that coefficient of each a_j in the sum is of the order of $s \times O(n^{-1/2})$, and on other hand, each a_j is bounded by $n^{1/\gamma}$. For the third inequality above we used the fact that $\operatorname{Var}(\sum_{i=1}^{s} \beta_i d_{j_i}) = \sum_{i=1}^{s} \operatorname{Var}(d_{j_i}) \leq (s+2)$ (by Lemma 18).

Note that if (1) or (2) fails, then one of the following events holds:

(i) either of the bricks L_0 and L_{-m} is visible;

(ii) there exists a stretch of M consecutive bricks from $L_{-m+1}, L_{-m+2}, \dots, L_{-1}$ such that at least $\lfloor M/2 \rfloor - 1$ of them are visible;

(iii) there exists a stretch of M consecutive bricks (say, $L_a, L_{a+1}, \dots, L_{a+M-1}$) from $L_{-m+1}, L_{-m+2}, \dots, L_{-1}$ such that $\sum_{i=0}^{M-1} \#(L_{a+i} \cap S) \geq M$.

By (17), we have $\mathbb{P}\{\text{event (i)}\} = O(n^{-\varepsilon^2/3})$. We observe that events (ii) and (iii) are both contained in the following event:

(iv) there exists a stretch of M consecutive bricks (say, $L_a, L_{a+1}, \dots, L_{a+M-1}$) from $L_{-m+1}, L_{-m+2}, \dots, L_{-1}$ such that $\sum_{i=0}^{M-1} \#(L_{a+i} \cap S) \geq \lfloor M/2 \rfloor - 1$.

Again by (17), if we fix a position of such M consecutive blocks $L_a, L_{a+1}, \dots, L_{a+M-1}$ and then fix $s = \lfloor M/2 \rfloor - 1$ positions $j_1, j_2, \dots$ within the blocks $L_a, L_{a+1}, \dots, L_{a+M-1}$, the probability that $j_1, j_2, \dots, j_s \in S$ is bounded above by $O(n^{-s\varepsilon^2/3}) = O(n^{-2})$. Hence by union bound, $\mathbb{P}\{\text{event (iv)}\} = O(n(\log n)^{3M/2} \times n^{-2})$. This is because we can choose the index a from the set $\{-m+1, -m+2, \dots, -1\}$ in at most $m \leq n$ ways and s positions $j_1, j_2, \dots$ can be selected in the blocks $L_a, L_{a+1}, \dots, L_{a+M-1}$ in at most $\binom{4Mr}{s} \leq O((\log n)^{3M/2})$ ways.

This implies that the probability that either of the events (i) or (iv) happens goes to 0 as $n \to \infty$. This completes the proof. $\square$

3.4. *Reduction to a block diagonal form.* Let $\mathbf{B}$ be the following block diagonal form of the matrix $\mathbf{P}$:

$$\mathbf{B}(k,l) = \begin{cases} \mathbf{P}(k,l), & \text{if } k,l \in J \text{ for some } J \in \Lambda, \\ 0, & \text{otherwise.} \end{cases}$$

LEMMA 9. *For each $\varepsilon > 0$, there exists $K > 0$ such that with high probability,*

$$\Bigl|\lambda_1(\mathbf{P}\mathbf{D}^\varepsilon\mathbf{P}) - \max_{J\in\Lambda: J\cap S\neq\varnothing} \lambda_1(\mathbf{P}[J]\mathbf{D}^\varepsilon[J]\mathbf{P}[J])\Bigr| \le K.$$

PROOF. The maximum above equals $\lambda_1(\mathbf{B}\mathbf{D}^\varepsilon\mathbf{B})$, so the left-hand side is bounded above by

$$\begin{aligned}\|\mathbf{P}\mathbf{D}^\varepsilon\mathbf{P} - \mathbf{B}\mathbf{D}^\varepsilon\mathbf{B}\|_{\mathrm{sp}} &\le \|(\mathbf{P}-\mathbf{B})\mathbf{D}^\varepsilon\mathbf{P}\|_{\mathrm{sp}} + \|\mathbf{B}\mathbf{D}^\varepsilon(\mathbf{P}-\mathbf{B})\|_{\mathrm{sp}} \\ &\le \|(\mathbf{P}-\mathbf{B})\mathbf{D}^\varepsilon\|_{\mathrm{sp}}\|\mathbf{P}\|_{\mathrm{sp}} + \|\mathbf{B}\|_{\mathrm{sp}}\|\mathbf{D}^\varepsilon(\mathbf{P}-\mathbf{B})\|_{\mathrm{sp}}.\end{aligned}$$

Note that since $\mathbf{P}$ is a projection matrix $\|\mathbf{P}\|_{\mathrm{sp}} = 1$ and by block-diagonality we have $\|\mathbf{B}\|_{\mathrm{sp}} = \max_{J\in\Lambda} \|\mathbf{P}[J]\|_{\mathrm{sp}}$. The matrix $\mathbf{P}[J]$ is just $\mathbf{P}$ conjugated by a coordinate projection, so it has norm at most 1.

Since $\mathbf{D}^\varepsilon = \mathbf{R}\mathbf{D}$ we first bound the spectral norm of $(\mathbf{P}-\mathbf{B})\mathbf{R}$. The maximal column sum of this matrix is bounded above by the maximal absolute row sum of $\mathbf{P}$, which by (13) is at most $C_2 \log n$. Note that $((\mathbf{P}-\mathbf{B})\mathbf{R})(k,l) = 0$ unless k,l are in different parts of Λ. This gives the upper bound for the maximal absolute row sum

$$\max_{J\in\Lambda, k\in J} \sum_{l\notin J} \mathbf{R}(l,l)|\mathbf{P}(k,l)| \le C \sum_{k=1}^{\#\Lambda-1} \frac{2M}{k(\log n)^3} = O((\log n)^{-2}),$$

which holds with high probability. We used the fact that with high probability each part in Λ has at most M elements j where $\mathbf{R}(j,j)$ is nonzero (Proposition 8), and different parts have gaps of size $(\log n)^3$ in between them. Since the spectral norm is bounded above by the geometric mean of the maximal row and column sums, we get $\|(\mathbf{P}-\mathbf{B})\mathbf{R}\|_{\mathrm{sp}} = O((\log n)^{-1/2})$.

By Bernstein's inequality, we can find a constant C_3 such that with high probability $\|\mathbf{D}\|_{\mathrm{sp}} = \max_j |d_j|$ is at most $C_3\sqrt{\log n}$. Therefore, $\|(\mathbf{P}-\mathbf{B})\mathbf{D}^\varepsilon\|_{\mathrm{sp}}^2 = O(1)$ with high probability. Similarly, $\|\mathbf{D}^\varepsilon(\mathbf{P}-\mathbf{B})\|_{\mathrm{sp}}^2 = O(1)$ with high probability, which yields the desired result. $\square$

4. Proof of the upper bound. Assume that $a_0, a_1, \ldots, a_n$ are independent mean zero, variance one random variables which are uniformly bounded by $n^{1/\gamma}$. This section consists of the proof of the upper bound.

PROPOSITION 10. *For each $\varepsilon > 0$, there exists $c_n = o(1)$ such that with high probability,*

$$\max_{J\in\Lambda: J\cap S\neq\varnothing} \frac{\lambda_1(\mathbf{P}[J]\mathbf{D}^\varepsilon[J]\mathbf{P}[J])}{\sqrt{2\log n}} \le \|\Pi\|_{2\to 4}^2 + c_n.$$

LEMMA 11. *Fix $k \geq 1$ and $\delta_1, \delta_2, \dots, \delta_k \geq 0$. Let $\|\delta\|_2 := (\delta_1^2 + \cdots + \delta_k^2)^{1/2}$. Then for sufficiently large n and for any $0 < j_1 < j_2 < \cdots < j_k < n$,*

$$\mathbb{P}\{|d_{j_1}| > \delta_1\sqrt{2\log n}, |d_{j_2}| > \delta_2\sqrt{2\log n}, \dots, |d_{j_k}| > \delta_k\sqrt{2\log n}\} \leq 2^{k+1}n^{-\|\delta\|_2^2}.$$

PROOF. To avoid triviality, assume that $\|\delta\|_2 > 0$. Clearly,

$$\begin{aligned}
&\mathbb{P}\{|d_{j_1}| > \delta_1\sqrt{2\log n}, \dots, |d_{j_k}| > \delta_k\sqrt{2\log n}\}\\
&\quad\leq \sum_{\beta_i \in \{\pm 1\}} \mathbb{P}\left\{\left(\sum_{i=1}^{k} \beta_i \delta_i d_{j_i}\right)\Big/ \|\delta\|_2 > \|\delta\|_2\sqrt{2\log n}\right\}.
\end{aligned}$$

By Lemma 18, for any choice of $\beta_i \in \{\pm 1\}$, the sum $(\beta_1\delta_1 d_{j_1} + \cdots + \beta_k\delta_k d_{j_k})/\|\delta\|_2$ has variance one and can be expressed as a linear combination independent random variables as $\sum_{i=0}^{n} \theta_i a_i$ for suitable real coefficients θ_i with $|\theta_i| \leq 2kn^{-1/2}$ and $\sum_{i=0}^{n} \theta_i^2 = 1$. Recall that $|a_i| \leq n^{1/\gamma}$ so that we can apply Bernstein's inequality to obtain

$$\begin{aligned}
&\mathbb{P}\left\{\left(\sum_{i=1}^{k} \beta_i \delta_i d_{j_i}\right)\Big/ \|\delta\|_2 > \|\delta\|_2\sqrt{2\log n}\right\}\\
&\quad\leq \exp\left(-\frac{\|\delta\|_2^2 \log n}{1 + \|\delta\|_2\sqrt{2\log n} \cdot 2kn^{1/\gamma - 1/2}/3}\right)\\
&\quad\leq 2n^{-\|\delta\|_2^2}
\end{aligned}$$

for sufficiently large n. This completes the proof of the lemma. $\square$

LEMMA 12. *Fix $\eta > 0, k \geq 1$. Then there exists a constant $C_4 = C_4(\eta, k)$ such that for sufficiently large n and for any $0 < j_1 < j_2 < \cdots < j_k < n$, we have*

$$\begin{aligned}
&\mathbb{P}\{|d_{j_1}| > \delta_1\sqrt{2\log n}, |d_{j_2}| > \delta_2\sqrt{2\log n}, \dots, |d_{j_k}| > \delta_k\sqrt{2\log n}\\
&\qquad \textit{for some } \delta_1, \dots, \delta_k > 0 \textit{ such that } \delta_1^2 + \delta_2^2 + \cdots + \delta_k^2 \geq 1 + \eta\}\\
&\quad\leq C_4 n^{-(1+\eta/2)}.
\end{aligned}$$

PROOF. Construct an $\frac{\eta}{4(1+\eta)k}$-net $\mathcal{N}$ for the interval $[0, 1+\eta]$ by choosing $\lfloor \frac{4(1+\eta)^2 k}{\eta} \rfloor + 1$ equally spaced points in $[0, 1+\eta]$ including both endpoints 0 and $1+\eta$. Therefore, given any $\delta_1, \dots, \delta_k \in [0, 1+\eta]$, we can find $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathcal{N}$ such that $\delta_i - \frac{\eta}{4(1+\eta)k} \leq \alpha_i \leq \delta_i$ for each i. This implies that $\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_k^2 > \sum_{i=1}^{k}(\delta_i^2 - 2\delta_i \frac{\eta}{4(1+\eta)k}) \geq \delta_1^2 + \delta_2^2 + \cdots + \delta_k^2 - \eta/2$. Clearly, the event $\{|d_{j_1}| > \delta_1\sqrt{2\log n}, \dots, |d_{j_k}| > \delta_k\sqrt{2\log n}$ for some $\delta_1, \dots, \delta_k \in [0, 1+\eta]$ such that $\delta_1^2 + \delta_2^2 + \cdots + \delta_k^2 \geq 1+\eta\}$ is contained in the finite union of events

$\{|d_{j_1}| > \alpha_1\sqrt{2\log n}, \ldots, |d_{j_k}| > \alpha_k\sqrt{2\log n}\}$ where the union is taken over for all possible choices of $\alpha_i \in \mathcal{N}$ for all i such that $\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_k^2 \geq 1 + \eta/2$. Now we use the union bound and Lemma 11 to conclude the probability of the given event is bounded by $2^{k+1}(\#\mathcal{N})^k n^{-(1+\eta/2)}$. On the other hand, it again follows from Lemma 11 that the probability of the event $\{|d_{j_i}| > (1+\eta)\sqrt{2\log n}$ for some $1 \leq i \leq k\}$ is bounded by $2kn^{-(1+\eta)^2} \leq 2kn^{-(1+\eta/2)}$.

We complete the proof by taking $C_4 = 2^{k+1}(\#\mathcal{N})^k + 2k$. $\square$

COROLLARY 13. *Let $\varepsilon > 0$, and let $M = M(\varepsilon)$ be as defined in Section 3.3. For every $\eta > 0$, with high probability, for all admissible blocks $L \in \mathcal{L}$ and all distinct $j_1, \ldots, j_M \in L$ we have $d_{j_1}^2 + \cdots + d_{j_M}^2 \leq (1+\eta)\sqrt{2\log n}$.*

PROOF. For a fixed admissible block and points j_i the probability that the claim is violated is at most $C_4 n^{-(1+\eta/2)}$ by the lemma. By union bound, the probability that the claim is violated is at most $n\binom{4M(\log n)^3}{M} \cdot C_4 n^{-(1+\eta/2)} = O((\log n)^{3M} n^{-\eta/2})$. This is because there can be at most n admissible blocks, the length of an admissible block can be at most $4M(\log n)^3$ and the number of ways M distinct indices can be chosen from an admissible block is at most $\binom{4M(\log n)^3}{M}$. $\square$

By part (2) of Proposition 8, with high probability, $\mathbf{D}^\varepsilon$ contains at most M nonzero entries in every admissible block. Thus it follows that for each $\eta > 0$ with high probability for all $L \in \mathcal{L}$,

$$\sum_{j \in L} (\mathbf{D}^\varepsilon_{j,j})^2 \leq 1 + \eta.$$

Therefore, we have

$$\begin{aligned}
&\max_{J \in \Lambda : J \cap S \neq \varnothing} \frac{\lambda_1(\mathbf{P}[J]\mathbf{D}^\varepsilon[J]\mathbf{P}[J])}{\sqrt{2\log n}} \\
&\quad \leq \sup\left\{\lambda_1\big(\mathbf{P}[L]\operatorname{Diag}(\delta_1, \delta_2, \ldots, \delta_{\#L})\mathbf{P}[L]\big) : L \in \mathcal{L}, \sum_{j=1}^{\#L} \delta_j^2 \leq 1 + \eta\right\},
\end{aligned} \tag{18}$$

which holds with high probability due to part (1) of Proposition 8. Now by (15), we have $\max_{L \in \mathcal{L}} \|\mathbf{P}[L] - \Pi[L]\|_{\mathrm{sp}} = O(\frac{(\log n)^3}{n})$. Therefore, if we take $q = q(n) := 4M\lceil\log n\rceil^3$ and $H = [1, q] \cap \mathbb{Z}$, then (18) can be bounded by

$$\begin{aligned}
(18) \leq {} & \sup\left\{\lambda_1\big(\Pi[L]\operatorname{Diag}(\delta_1, \delta_2, \ldots, \delta_{\#L})\Pi[L]\big) : L \in \mathcal{L}, \sum_{j=0}^{\#L} \delta_j^2 \leq 1 + \eta\right\} \\
& + O\left(\frac{(\log n)^3}{n}\right)
\end{aligned}$$

$$\leq (1+\eta)^{1/2} \sup\left\{\lambda_1\big(\Pi[H]\operatorname{Diag}(\delta_1,\delta_2,\dots,\delta_q)\Pi[H]\big) : \sum_{j=1}^{q}\delta_j^2 \leq 1\right\}$$
$$+ O\left(\frac{(\log n)^3}{n}\right).$$

By Rayleigh's characterization of the maximum eigenvalue of Hermitian matrices, the above supremum equals

$$\sup\left\{\langle \mathbf{v}, \Pi[H]\operatorname{Diag}(\delta_1,\delta_2,\dots,\delta_q)\Pi[H]\mathbf{v}\rangle : \sum_{j=1}^{q}\delta_j^2 \leq 1, \mathbf{v}\in\mathbb{C}^q, \|\mathbf{v}\|_2 \leq 1\right\}.$$

Denote by δ the infinite dimensional vector $(\dots,\delta_{-1},\delta_0,\delta_1,\dots)$ in $\ell^2(\mathbb{R})$. Now extending the range of optimization we get the upper bound

$$\sup\{\langle \mathbf{v}, \Pi \operatorname{Diag}(\delta)\Pi\mathbf{v}\rangle : \delta\in\ell^2(\mathbb{R}), \|\delta\|_2\leq 1, \mathbf{v}\in\ell^2(\mathbb{C}), \|\mathbf{v}\|_2\leq 1\}. \tag{19}$$

Now note that (with $\odot$ denoting coordinate-wise multiplication)

$$\langle \mathbf{v}, \Pi\operatorname{Diag}(\delta)\Pi\mathbf{v}\rangle = \langle \Pi\mathbf{v}, \operatorname{Diag}(\delta)\Pi\mathbf{v}\rangle = \langle \delta, \overline{\Pi\mathbf{v}}\odot\Pi\mathbf{v}\rangle,$$

so if we fix $\mathbf{v}$, this is maximized when δ equals $\overline{\Pi\mathbf{v}}\odot\Pi\mathbf{v}$ divided by its length. The maximum, for $\mathbf{v}$ fixed is $\|\overline{\Pi\mathbf{v}}\odot\Pi\mathbf{v}\|_2 = \|\Pi\mathbf{v}\|_4^2$, so expression (19) equals $\|\Pi\|_{2\to 4}^2$. This completes the proof of Proposition 10.

5. Proof of the lower bound. Assume that $a_0, a_1, \dots, a_n$ are independent mean zero, variance one random variables which are uniformly bounded by $n^{1/\gamma}$. This section consists of the proof of the lower bound.

PROPOSITION 14. *For each $\tau > 0$ and $\varepsilon > 0$, with high probability,*

$$\max_{J\in\Lambda : J\cap S\neq\varnothing} \frac{\lambda_1(\mathbf{P}[J]\mathbf{D}^{\varepsilon}[J]\mathbf{P}[J])}{\sqrt{2\log n}} \geq \|\Pi\|_{2\to 4}^2 - 2\tau.$$

Let $G_0, G_1, \dots, G_n$ be i.i.d. standard Gaussians independent of $(a_i)_{0\leq i\leq n}$. Recall

$$d_j(a) = \frac{1}{\sqrt{2n}}\left[\sqrt{2}a_0 + (-1)^j\sqrt{2}a_n + 2\sum_{k=1}^{n-1} a_k \cos\left(\frac{2\pi jk}{n}\right)\right], \qquad 0\leq j < 2n.$$

We will use $d_j(G)$ to refer to the above sum with a_i being replaced by G_i for each $0\leq i\leq n$. Suppose we are given some *nonzero* real numbers $u_1, u_2, \dots, u_k$ for some fixed $k\geq 1$ such that $u_1^2 + u_2^2 + \cdots + u_k^2 < 1$. Fix any $\eta > 0$ small such that $\eta < |u_i|$ for each i and if we set $u_s' := |u_s| - \eta > 0$ for $1\leq s\leq k$, then $(u_1')^2 +$

$\cdots + (u'_k)^2 < 1$. Take $p = p(n) := 100\lceil \log n \rceil^3$, $b = b(n) := 12\lceil \log n \rceil^3$ and $N = N(n) := \lfloor n/2p \rfloor$. For each $1 \le j \le N$, define the intervals

$$I_j = \begin{cases} (-\varepsilon, \varepsilon), & \text{for } -b+1 \le j \le 0, \\ (u_j - \eta, u_j + \eta), & \text{for } 1 \le j \le k, \\ (-\varepsilon, \varepsilon), & \text{for } k+1 \le j \le k+b, \end{cases}$$

and let A_i be the event that $d_{ip+j} \in \sqrt{2 \log n} I_j$ for all $j = -b+1, \dots, b+k$.

PROPOSITION 15. *Let A_i be as above. Then as $n \to \infty$, the probability that at least one of the events $A_1, A_2, \dots, A_N$ happens converges to one.*

PROOF. The proof is based on the second moment method. First of all, fix any smooth function $\psi : \mathbb{R} \to [0, 1]$ such that $\psi(x) = 0$ for $x \le 0$ and $\psi(x) = 1$ for $x \in [1, \infty)$. For any reals $a < b$, the indicator function of the interval $(a\sqrt{2 \log n}, b\sqrt{2 \log n})$ can be bounded below by the smooth function

$$\mathbf{1}_{(a\sqrt{2 \log n}, b\sqrt{2 \log n})}(x) \ge \zeta_{(a,b)}(x) := \psi(x - a\sqrt{2 \log n})\psi(b\sqrt{2 \log n} - x).$$

For a fixed $a < b$, the first, second and third derivatives of the function $\zeta_{(a,b)}$ are all bounded by some constant in the supremum norm. Define, for each i in $1 \le i \le N$,

$$W_i = \prod_{j=-b+1}^{b+k} \zeta_{I_j}(d_{ip+j}). \tag{20}$$

Clearly, for each i, $W_i \le \mathbf{1}_{A_i}$. Thus by the Paley–Zygmund inequality,

$$\mathbb{P}\left\{ \sum_{i=1}^N \mathbf{1}_{A_i} \ge 1 \right\} \ge \mathbb{P}\left\{ \sum_{i=1}^N W_i > 0 \right\} \ge \frac{\mathbb{E}[(\sum_{i=1}^N W_i)^2]}{(\mathbb{E}\sum_{i=1}^N W_i)^2}.$$

So, the proposition follows if we can show

$$\mathbb{E}\left[\left(\sum_{i=1}^N W_i \right)^2 \right] = (1 + o(1)) \left(\mathbb{E} \sum_{i=1}^N W_i \right)^2.$$

The rest of the proof is devoted to establishing this.

We will write W_i^G to denote the random variable corresponding to (20) with $d_j(G)$ in place $d_j(a)$ for all j. Recall that $(d_j(G))_{0<j<n}$ is a sequence of i.i.d. standard Gaussian random variables. Using the following well-known Gaussian tail estimates,

$$\frac{x}{x^2+1} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \le \mathbb{P}\{G_0 > x\} \le \frac{1}{x} \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \qquad x > 0,$$

we obtain, for all large n,

$$
\begin{aligned}
\mathbb{E}W_i^G &= \prod_{j=-b+1}^{b+k} \mathbb{E}\zeta_{I_j}\big(d_{ip+j}(G)\big) \\
&= \big(1 - O\big(n^{-\varepsilon^2}\big)\big)^{2b} \prod_{s=1}^{k} \Omega\bigg(\frac{1}{\sqrt{\log n} \cdot n^{u_s'^2}}\bigg) \qquad (21)\\
&= \Omega\big((\log n)^{-k/2} \cdot n^{-((u_1')^2+\cdots+(u_k')^2)}\big),
\end{aligned}
$$

uniformly over $1 \le i \le N$, which implies that $\sum_{i=1}^N \mathbb{E}W_i^G \to \infty$ as $n \to \infty$. Since the random variables W_i^G are bounded by 1, we have

$$
\begin{aligned}
\sum_{i=1}^N \operatorname{Var} W_i^G &\le \sum_{i=1}^N \mathbb{E}\big[\big(W_i^G\big)^2\big] \le \sum_{i=1}^N \mathbb{E}W_i^G \qquad (22)\\
&= o\Bigg(\bigg(\sum_{i=1}^N \mathbb{E}W_i^G\bigg)^2\Bigg).
\end{aligned}
$$

For any $1 \le i \ne i' \le N$, the two subsets of indices $\{ip-b+1, \dots, ip+k+b\}$ and $\{i'p-b+1, \dots, i'p+k+b\}$ are disjoint. Hence, W_i^G and $W_{i'}^G$ are independent of each other and therefore,

$$
\mathbb{E}\big[W_i^G W_{i'}^G\big] = \mathbb{E}\big[W_i^G\big]\mathbb{E}\big[W_{i'}^G\big]. \qquad (23)
$$

Now (22) and (23) yield

$$
\mathbb{E}\Bigg[\bigg(\sum_{i=1}^N W_i^G\bigg)^2\Bigg] = \big(1 + o(1)\big)\bigg(\mathbb{E}\sum_{i=1}^N W_i^G\bigg)^2. \qquad (24)
$$

Our next goal is to show that the differences $\mathbb{E}W_i - \mathbb{E}W_i^G$ and $\mathbb{E}[W_i W_{i'}] - \mathbb{E}[W_i^G W_{i'}^G]$ are of smaller order for each $i \ne i'$ using the invariance principle given in Lemma 17.

We apply Lemma 17 with $r = n+1, m = k+2b, \mathbf{X} = (a_0, \dots, a_n)$, $\mathbf{Y} = (G_0, \dots, G_n)$, $f = (d_{ip-b+1}, \dots, d_{ip+k+b})$ and

$$
g(\mathbf{z}) = \prod_{j=-b+1}^{b+k} \zeta_{I_j}(z_{b+j}),
$$

where $\mathbf{z} = (z_1, z_2, \dots, z_{2b+k})$ to obtain

$$
\big|\mathbb{E}[W_i] - \mathbb{E}\big[W_i^G\big]\big| \le \sum_{j=1}^r \mathbb{E}[R_j] + \sum_{j=1}^r \mathbb{E}[T_j], \qquad (25)
$$

where R_j and T_j are as defined in Lemma 17 with

$$h_j(\mathbf{x}) = \sum_{\ell,p,q=1}^{m} \partial_\ell\,\partial_p\,\partial_q g\bigl(f(\mathbf{x})\bigr)\,\partial_j f_\ell(\mathbf{x})\,\partial_j f_p(\mathbf{x})\,\partial_j f_q(\mathbf{x}) + 3\sum_{\ell,p=1}^{m} \partial_\ell\,\partial_p g\bigl(f(\mathbf{x})\bigr)\,\partial_j^2 f_\ell(\mathbf{x})\,\partial_j f_p(\mathbf{x}) + \sum_{\ell=1}^{m} \partial_\ell g\bigl(f(\mathbf{x})\bigr)\,\partial_j^3 f_\ell(\mathbf{x}). \tag{26}$$

We will first find an upper bound for $\mathbb{E}[R_j]$. The bound for $\mathbb{E}[T_j]$ will be similar. Note that $\|\partial_\ell f_t\|_\infty \le 2n^{-1/2}$ for each ℓ and t. Since each function f_t is linear, its higher derivatives all vanish and hence the second and the third term of h_j in (26) disappear. Also, $\|\partial_\ell\,\partial_{\ell'}\,\partial_{\ell''} g\|_\infty = O(1)$ for each ℓ, ℓ' and ℓ'' and $\partial_\ell\,\partial_{\ell'}\,\partial_{\ell''} g(\mathbf{z}) \ne 0$ only if $z_{b+j} \in \sqrt{2\log n}\, I_j$ for $j = 1,\dots,k$. For $x \in \mathbb{R}$, define a random vector $\mathbf{Z}^{(j)}(x) := (a_0,\dots,a_{j-2}, x, G_j,\dots,G_n)$. Therefore, we have

$$\mathbb{E}R_j \le \frac{cm^3}{n^{3/2}}\mathbb{E}\Bigl[|a_{j-1}|^3 \sup_{x:|x|\le|a_{j-1}|} \mathbf{1}\bigl\{d_{ip+s}\bigl(\mathbf{Z}^{(j)}(x)\bigr) \in \sqrt{2\log n}\, I_s,\ 1 \le s \le k\bigr\}\Bigr]. \tag{27}$$

Note that d_j depends on a_{j-1} linearly with absolute coefficient at most $\sqrt{2/n}$, and that the random variable a_{j-1} is bounded by $n^{1/\gamma}$. Thus the supremum above is at most

$$\mathbf{1}\bigl\{\bigl|d_{ip+s}\bigl(\mathbf{Z}^{(j)}(0)\bigr)\bigr| \ge \bigl(|u_s| - \eta\bigr)\sqrt{2\log n} - \sqrt{2}n^{1/\gamma-1/2},\ 1 \le s \le k\bigr\},$$

which is independent of a_{j-1}. Since $\mathbb{E}a_{j-1}^2 = 1$, we have, with $u_s' = |u_s| - \eta$

$$(27) \le \frac{cm^3}{n^{3/2-1/\gamma}}\mathbb{P}\bigl\{\bigl|d_{ip+s}\bigl(\mathbf{Z}^{(j)}(0)\bigr)\bigr| \ge u_s'\sqrt{2\log n} - \sqrt{2}n^{1/\gamma-1/2},\ 1 \le s \le k\bigr\}. \tag{28}$$

Note that if we truncate the random variables $G_j,\dots,G_n$ at level $n^{1/\gamma}$, then we can bound (28) using Bernstein's inequality exactly as we did in proving Lemma 11. Toward this end, we define

$$\hat{\mathbf{Z}}^{(j)} = \bigl(a_0,\dots,a_{j-2}, 0, G_j\mathbf{1}_{\{|G_j|\le n^{1/\gamma}\}},\dots,G_n\mathbf{1}_{\{|G_n|\le n^{1/\gamma}\}}\bigr).$$

Then

$$
\begin{aligned}
(28) \le \frac{cm^3}{n^{3/2-1/\gamma}} \big[& \mathbb{P}\{|d_{ip+s}(\hat{\mathbf{Z}}^{(j)})| \ge u'_s \sqrt{2\log n} - \sqrt{2} n^{1/\gamma-1/2}, 1 \le s \le k\} \\
& + \mathbb{P}\{|G_\ell| > n^{1/\gamma} \text{ for some } \ell\}\big] \qquad (29) \\
\le \frac{cm^3}{n^{1/2-1/\gamma}} \big[& O\big(n^{-((u'_1)^2+\cdots+(u'_k)^2)}\big) + O\big(n \exp(-n^{2/\gamma}/2)\big)\big].
\end{aligned}
$$

Hence, by combining (27), (28) and (29), we obtain

$$
\sum_{j=1}^{r} \mathbb{E}[R_j] \le O\big((\log n)^9 n^{1/\gamma-1/2} \cdot n^{-((u'_1)^2+\cdots+(u'_k)^2)}\big),
$$

where the constant hidden inside the big-O notation above does not depend on i. Similar computation yields the same asymptotic bound for $\sum_{j=1}^{r} \mathbb{E}[T_j]$. Therefore, by (25) and (21), we have

$$
\begin{aligned}
\mathbb{E}[W_i] &= \mathbb{E}[W_i^G] + O\big((\log n)^9 n^{1/\gamma-1/2} \cdot n^{-((u'_1)^2+\cdots+(u'_k)^2)}\big) \\
&= (1+o(1))\mathbb{E}[W_i^G],
\end{aligned} \tag{30}
$$

which hold uniformly in $1 \le i \le N$. By similar argument as above, we can also show that

$$
\begin{aligned}
\mathbb{E}[W_i W_{i'}] &= \mathbb{E}[W_i^G W_{i'}^G] + O\big((\log n)^9 n^{1/\gamma-1/2} \cdot n^{-2((u'_1)^2+\cdots+(u'_k))^2}\big) \\
&= (1+o(1))\mathbb{E}[W_i^G W_{i'}^G],
\end{aligned} \tag{31}
$$

uniformly in $1 \le i \ne i' \le N$. From (30) arguing similarly as we did in (22), we deduce

$$
\sum_{i=1}^{N} \operatorname{Var} W_i \le o\left(\left(\sum_{i=1}^{N} \mathbb{E} W_i\right)^2\right). \tag{32}
$$

Finally, combining (23), (31) and (32) together, we have

$$
\sum_{i,i'=1}^{N} \mathbb{E}[W_i W_{i'}] = (1+o(1))\left(\sum_{i=1}^{N} \mathbb{E}[W_i]\right)^2.
$$

This implies that $\mathbb{P}\{\sum_{i=1}^{N} \mathbf{1}_{A_i} \ge 1\} = 1 - o(1)$ which completes the proof. $\square$

For any finite $k \ge 1$, we write Π_k as a shorthand for $k \times k$ matrix $\Pi[\{1, 2, \dots, k\}]$. Arguing along the line of the proof of the fact $\|\Pi\|_{2\to 4}^2 = \sup\{\langle \mathbf{v}, \Pi \operatorname{diag}(\delta) \Pi \mathbf{v}\rangle : \delta \in \ell^2(\mathbb{R}), \|\delta\|_2 \le 1, \mathbf{v} \in \ell^2(\mathbb{C}), \|\mathbf{v}\|_2 \le 1\}$, we can also show that

$$
\|\Pi_k\|_{2\to 4}^2 = \sup\{\lambda_1(\Pi_k \operatorname{diag}(\delta) \Pi_k) : \delta \in \mathbb{R}^k, \|\delta\|_2 \le 1\}. \tag{33}
$$

Next we prove

LEMMA 16. $\lim_{k\to\infty} \|\Pi_k\|_{2\to 4} = \|\Pi\|_{2\to 4}$.

PROOF. Since the operators ψ, ψ^{-1}, $\chi_{[0,1/2]}$ are all bounded and the inclusion map $\iota : \ell^2(\mathbb{C}) \to \ell^4(\mathbb{C})$ is also a bounded operator, we have $\|\Pi\|_{2\to 4} < \infty$.

It will be convenient to think of Π_{2k+1} as a linear operator acting on the space $\ell^2(\mathbb{C})$ with the representation $\Pi_{2k+1}(i, j) = \Pi(i, j)$ for $|i|, |j| \le k$ and 0 otherwise. Clearly, $\|\Pi_k\|_{2\to 4}$ is increasing and bounded above by $\|\Pi\|_{2\to 4}$.

For the other direction, consider a sequence of unit vectors $\mathbf{v}_m \in \ell^2(\mathbb{C})$ supported on $[-m, m]$ so that $\|\Pi\mathbf{v}_m\|_4 \to \|\Pi\|_{2\to 4}$. Then for $k \ge m$ we have

$$\|\Pi\mathbf{v}_m\|_4^4 - \|\Pi_{2k+1}\mathbf{v}_m\|_4^4 = \sum_{i:|i|>k} \left|(\Pi\mathbf{v}_m)(i)\right|^4 = \sum_{i:|i|>k} \left| \sum_{j:|j|\le m} \Pi(i, j)\mathbf{v}_m(j)\right|^4.$$

Since $|\Pi(i, j)| \le |i - j|^{-1}$ and $|\mathbf{v}_m(j)| \le 1$, the inside sum is bounded above by $(2m + 1)(|i| - m)^{-1}$. This gives the upper bound

$$(2m + 1)^4 \sum_{\ell:|\ell|>k-m} \frac{1}{\ell^4} \to 0 \qquad \text{as } k \to \infty.$$

Letting $k \to \infty$ and then $m \to \infty$ completes the proof. $\square$

Given $\tau > 0$, by (33) and Lemma 16, we can find $k \ge 1$ sufficiently large and a vector $\mathbf{u} = (u_1, u_2, \ldots, u_k) \in \mathbb{R}^k$ with $\|\mathbf{u}\|_2 < 1$ such that $\lambda_1(\Pi_k \operatorname{diag}(\mathbf{u})\Pi_k) > \|\Pi\|_{2\to 4}^2 - \tau/2$. By perturbing the coordinates a little, if necessary, we can also assume $u_s \ne 0$ for each $1 \le s \le k$. Now choose $\eta > 0$ sufficiently small such that:

(1) $0 \notin (u_s - \eta, u_s + \eta)$ for each $1 \le s \le k$.

(2) If $\mathbf{v} = (v_1, v_2, \ldots, v_k) \in [u_1 - \eta, u_1 + \eta] \times \cdots \times [u_k - \eta, u_k + \eta]$, then $\lambda_1(\Pi_k \operatorname{diag}(\mathbf{v})\Pi_k) > \|\Pi\|_{2\to 4}^2 - \tau$.

(3) $\sup\{\|\mathbf{v}\|_2 : \mathbf{v} \in [u_1 - \eta, u_1 + \eta] \times \cdots \times [u_k - \eta, u_k + \eta]\} < 1$.

Finally choose $\varepsilon > 0$ small such that $[-\varepsilon, \varepsilon] \cap (u_s - \eta, u_s + \eta) = \varnothing$ for each $1 \le s \le k$.

By Proposition 15 we know that one of the events A_i, $1 \le i \le N$ happens with high probability. If A_i occurs, then the points $ip - \lceil \log n\rceil^3, \ldots, ip + k + \lceil \log n\rceil^3$ are contained in a single partition block $J \in \Lambda$. This is because when A_i occurs, $d_{ip+j} \in (u_s - \eta, u_s + \eta)$ and hence $|d_{ip+j}| > \varepsilon$ for each $j = 1, \ldots, k$ and because of the property of our random partition which guarantees that each (random) partition block always has a padding of two invisible bricks (of length between $\lceil \log n\rceil^3$ and $4\lceil \log n\rceil^3$) from each side. On the other hand, since two consecutive invisible bricks cannot belong to the same partitioning block, J has no other point from S except the k points $ip + 1, ip + 2, \ldots, ip + k$. Write

$F := \{ip+1, ip+2, \dots, ip+k\}$. Therefore, if A_i happens, then

$$\begin{aligned}
&\frac{\lambda_1(\mathbf{P}[J]\mathbf{D}^\varepsilon[J]\mathbf{P}[J])}{\sqrt{2\log n}} \\
&\quad \ge \frac{\lambda_1(\mathbf{P}[F]\mathbf{D}^\varepsilon[F]\mathbf{P}[F])}{\sqrt{2\log n}} \\
&\quad \ge \inf\{\lambda_1(\mathbf{P}[F]\operatorname{diag}(\mathbf{v})\mathbf{P}[F]) : \mathbf{v} \in [u_1-\eta, u_1+\eta] \times \cdots \times [u_k-\eta, u_k+\eta]\}.
\end{aligned}$$

By the convergence (15) of $\mathbf{P}$ to Π this equals

$$\begin{aligned}
&\inf\{\lambda_1(\Pi_k \operatorname{diag}(\mathbf{v})\Pi_k) : \mathbf{v} \in [u_1-\eta, u_1+\eta] \times \cdots \times [u_k-\eta, u_k+\eta]\} - O(n^{-1}) \\
&\quad \ge \|\Pi_k\|^2_{2\to 4} - \tau - O(n^{-1}).
\end{aligned}$$

So, by Lemma 16 we get with high probability,

$$\max_{J\in\Lambda : J\cap S\neq\varnothing} \frac{\lambda_1(\mathbf{P}[J]\mathbf{D}^\varepsilon[J]\mathbf{P}[J])}{\sqrt{2\log n}} \ge \|\Pi_k\|^2_{2\to4} - \tau - O(n^{-1}) = \|\Pi\|^2_{2\to4} - \tau - o(1).$$

This yields the claim of Proposition 14.

APPENDIX

A.1. Invariance principle. Let $\mathbf{X} = (X_1, \dots, X_r)$ and $\mathbf{Y} = (Y_1, \dots, Y_r)$ be two vectors of independent random variables with finite second moments, taking values in some open interval I and satisfying, for each i, $\mathbb{E}[X_i] = \mathbb{E}[Y_i]$ and $\mathbb{E}[X_i^2] = \mathbb{E}[Y_i^2]$. We shall also assume that $\mathbf{X}$ and $\mathbf{Y}$ are defined on the same probability space and are independent. The following lemma is an immediate generalization of Theorem 1.1 of Chatterjee (2005) (based on Lindeberg's approach to the CLT). We need this more detailed version because we will use the invariance principle under the moderate deviation regime.

LEMMA 17. *Let $f = (f_1, f_2, \dots, f_m) : I^r \to \mathbb{R}^m$ be thrice continuously differentiable. If we set $\mathbf{U} = f(\mathbf{X})$ and $\mathbf{V} = f(\mathbf{Y})$, then for any thrice continuously differentiable $g : \mathbb{R}^m \to \mathbb{R}$,*

$$\left|\mathbb{E}[g(\mathbf{U})] - \mathbb{E}[g(\mathbf{V})]\right| \le \sum_{i=1}^r \mathbb{E}[R_i] + \sum_{i=1}^r \mathbb{E}[T_i],$$

where

$$R_i := \frac{1}{6}|X_i|^3 \times \sup_{x\in[\min(0,X_i),\max(0,X_i)]} \left|h_i(X_1, \dots, X_{i-1}, x, Y_{i+1}, \dots, Y_r)\right|,$$

$$T_i := \frac{1}{6}|Y_i|^3 \times \sup_{y\in[\min(0,Y_i),\max(0,Y_i)]} \left|h_i(X_1, \dots, X_{i-1}, y, Y_{i+1}, \dots, Y_r)\right|,$$

$$h_i(\mathbf{x}) := \sum_{\ell,p,q=1}^{m} \partial_\ell\,\partial_p\,\partial_q g\big(f(\mathbf{x})\big)\partial_i f_\ell(\mathbf{x})\,\partial_i f_p(\mathbf{x})\,\partial_i f_q(\mathbf{x})$$
$$+ 3\sum_{\ell,p=1}^{m} \partial_\ell\,\partial_p g\big(f(\mathbf{x})\big)\,\partial_i^2 f_\ell(\mathbf{x})\,\partial_i f_p(\mathbf{x}) + \sum_{\ell=1}^{m} \partial_\ell g\big(f(\mathbf{x})\big)\,\partial_i^3 f_\ell(\mathbf{x}).$$

PROOF. The lemma can be proved by merely imitating the steps of Chatterjee [(2005), Theorem 1.1], but we include a proof here for sake of completeness. Let $H : I^r \to \mathbb{R}$ be the function $H(\mathbf{x}) := g(f(\mathbf{x}))$. It is a routine computation to verify that $\partial_i^3 H(\mathbf{x}) = h_i(\mathbf{x})$ for all i. For $0 \le i \le r$, define the random vectors $\mathbf{Z}_i := (X_1, X_2, \dots, X_{i-1}, X_i, Y_{i+1}, \dots, Y_n)$ and $\mathbf{W}_i := (X_1, X_2, \dots, X_{i-1}, 0, Y_{i+1}, \dots, Y_n)$ with obvious meanings for $i = 0$ and $i = r$. For $1 \le i \le r$, define

$$\mathrm{Error}_i^{(1)} = H(\mathbf{Z}_i) - X_i\,\partial_i H(\mathbf{W}_i) - \tfrac{1}{2}X_i^2\,\partial_i H(\mathbf{W}_i)$$

and

$$\mathrm{Error}_i^{(2)} = H(\mathbf{Z}_{i-1}) - Y_i\,\partial_i H(\mathbf{W}_i) - \tfrac{1}{2}Y_i^2\,\partial_i H(\mathbf{W}_i).$$

Hence, by Taylor's remainder theorem and the above observation about the third partial derivatives of H, it follows that

$$\big|\mathrm{Error}_i^{(1)}\big| \le R_i \quad \text{and} \quad \big|\mathrm{Error}_i^{(2)}\big| \le T_i.$$

For each i, X_i, Y_i and $\mathbf{W}_i$ are independent. Hence,

$$\mathbb{E}\big[X_i\,\partial_i H(\mathbf{W}_i)\big] - \mathbb{E}\big[Y_i\,\partial_i H(\mathbf{W}_i)\big] = \mathbb{E}[X_i - Y_i]\cdot\mathbb{E}\big[H(\mathbf{W}_i)\big] = 0.$$

Similarly, $\mathbb{E}[X_i^2\,\partial_i H(\mathbf{W}_i)] = \mathbb{E}[Y_i^2\,\partial_i H(\mathbf{W}_i)]$. Combining all these ingredients, we obtain

$$\begin{aligned}
\big|\mathbb{E}[g(\mathbf{U})] - \mathbb{E}[g(\mathbf{V})]\big| &= \left|\sum_{i=1}^{r}\big(\mathbb{E}\big[H(\mathbf{Z}_i)\big] - \mathbb{E}\big[H(\mathbf{Z}_{i-1})\big]\big)\right| \\
&\le \left|\sum_{i=1}^{r}\mathbb{E}\big[X_i\,\partial_i H(\mathbf{W}_i)\big] + \frac{1}{2}X_i^2\,\partial_i H(\mathbf{W}_i) + \mathrm{Error}_i^{(1)}\right. \\
&\qquad \left. - \sum_{i=1}^{r}\mathbb{E}\big[Y_i\,\partial_i H(\mathbf{W}_i)\big] + \frac{1}{2}Y_i^2\,\partial_i H(\mathbf{W}_i) + \mathrm{Error}_i^{(2)}\right| \\
&\le \sum_{i=1}^{r}\mathbb{E}[R_i] + \sum_{i=1}^{r}\mathbb{E}[T_i],
\end{aligned}$$

which completes the proof. $\square$

A.2. Covariances between the eigenvalues of random circulant.

LEMMA 18. *Let $a_0, a_1, \dots, a_n$ be independent mean zero, variance one random variables. Define*

$$d_j = \frac{1}{\sqrt{2n}}\left[\sqrt{2}a_0 + (-1)^j\sqrt{2}a_n + 2\sum_{k=1}^{n-1} a_k \cos\left(\frac{2\pi jk}{2n}\right)\right], \qquad 0 \le j < 2n.$$

Then $d_j = d_{2n-j}$ for $0 < j < 2n$. Moreover, the random variables $d_j, 0 \le j \le n$ have mean 0, *and their covariances are given by*

$$\mathbb{E}[d_j d_k] = \begin{cases} 2, & \text{if } j = k \in \{0, n\}, \\ 1, & \text{if } 0 < j = k < n, \\ 0, & \text{if } 0 \le j \ne k \le n. \end{cases}$$

PROOF. The fact that $d_j = d_{2n-j}$ for $0 < j < 2n$ and zero mean property is immediate from the definition of d_j. Since $a_0, a_1, \dots, a_n$ be independent with variance one,

$$\begin{aligned}\mathbb{E}[d_j d_k] &= \frac{1}{2n}\left[2 + 2(-1)^{j+k} + 4\sum_{\ell=1}^{n-1}\cos\left(\frac{2\pi j\ell}{2n}\right)\cos\left(\frac{2\pi k\ell}{2n}\right)\right] \\ &= \frac{1}{2n}\left[2 + 2(-1)^{j+k} + 2\sum_{\ell=1}^{n-1}\cos\left(\frac{2\pi(j-k)\ell}{2n}\right)\right. \\ &\qquad\qquad \left. + 2\sum_{\ell=1}^{n-1}\cos\left(\frac{2\pi(j+k)\ell}{2n}\right)\right].\end{aligned}$$

Plugging in $x = \frac{2\pi m}{2n}$, $m = 0, 1, 2, \dots, 2n$ into the well-known Dirichlet kernel formula,

$$1 + 2\sum_{\ell=1}^{n-1}\cos(\ell x) = \frac{\sin((n-1/2)x)}{\sin(x/2)},$$

we obtain

$$\sum_{\ell=1}^{n-1}\cos\left(\frac{2\pi m\ell}{2n}\right) = -\frac{1+(-1)^m}{2},$$

unless $m = 0$ or $2n$ when the sum equals to $(n-1)$. Using the above formula, the covariances $\mathbb{E}[d_j d_k]$ can be easily computed. $\square$

A.3. Optimization problem and connection to the sine kernel. For any $\delta > 0$, let υ_δ be the indicator function $\mathbf{1}_{[-\delta/2,\delta/2]}$. For a complex valued function f defined on $\mathbb{R}$, we define the involution f^* by $f^*(x) = \overline{f(-x)}$. Given two functions f and g on $\mathbb{R}$ their convolution $f \star g$ is defined as $(f \star g)(x) = \int_{\mathbb{R}} f(x-y)g(y)\,dy$ provided the integral makes sense. Let $f \star_{\mathbb{T}} g(x) = \int_{-1/2}^{1/2} \tilde{f}(x-y)\tilde{g}(y)\,dy$ denote the convolution of the two functions f and g are in $L^2(\mathbb{T})$ where $\tilde{f}$ and $\tilde{g}$ are the periodic extension of f and g, respectively, on the whole real line. Let $\hat{f}(t) = \int_{\mathbb{R}} e^{-2\pi i x t} f(x)\,dx$ be the usual Fourier transform of $f \in L^2(\mathbb{R})$. Let ψ^{-1} be the discrete Fourier transform of from $L^2(\mathbb{T})$ to $\ell^2(\mathbb{C})$. Below we collect some basic facts of the usual and discrete Fourier transform which we will need later:

(1) $\psi^{-1} : L^2(\mathbb{T}) \to \ell^2(\mathbb{C})$ and $\hat{}: L^2(\mathbb{R}) \to L^2(\mathbb{R})$ are isometries.

(2) $\overline{\psi^{-1}(f)} = \psi^{-1}(f^*)$ and $\overline{\hat{f}} = \widehat{f^*}$.

(3) $\psi^{-1}(f \star_{\mathbb{T}} g)(k) = \psi^{-1}(f)(k)\psi^{-1}(g)(k)$ for all $k \in \mathbb{Z}$ and for all $f, g \in L^2(\mathbb{T})$. When $f, g \in L^2(\mathbb{T})$, then $\widehat{f \star g} = \hat{f} \cdot \hat{g}$.

(4) If f and g are supported on $[0, 1/2]$, then $f \star_{\mathbb{T}} g = f \star g$.

(5) $\hat{\upsilon}_1(t) = \frac{\sin(\pi t)}{\pi t}$.

Note that $\mathtt{Sin}(f)(x) = \hat{\upsilon}_1 \star f(x)$ for $f \in L^2(\mathbb{R})$. The next lemma establishes the connection between the $2 \to 4$ norm of the operator Π (defined in Section 3.1) and the $2 \to 4$ norm of the integral operator $\mathtt{Sin}$.

LEMMA 19. *The following holds true*:

$$\|\Pi\|_{2\to4}^2 = \frac{1}{\sqrt{2}} \|\mathtt{Sin}\|_{2\to4}^2.$$

Let us define for $\delta > 0$,

$$K_\delta := \sup\{ \|(f\upsilon_\delta)^* \star (f\upsilon_\delta)\|_2 : f \in L^2(\mathbb{R}), \|f\|_2 \le 1 \}. \tag{34}$$

LEMMA 20. *Let K_δ be as above. Then $K_\delta = \delta^{1/2} K_1$.*

PROOF. For any $f \in L^2(\mathbb{R})$,

$$\|(f\upsilon_\delta)^* \star (f\upsilon_\delta)\|_2^2 = \int \left| \int f\upsilon_\delta(x+t)\overline{f\upsilon_\delta(t)}\,dt \right|^2 dx.$$

After a change of variables $s = t/\delta$ and $y = x/\delta$ the above integral is same as

$$\delta^3 \int \left| \int f\upsilon_\delta(\delta(s+y))\overline{f\upsilon_\delta(\delta s)}\,ds \right|^2 dy. \tag{35}$$

Keeping in mind that $\upsilon_\delta(\delta s) = \upsilon_1(s)$ and replacing f by $g(x) := \delta^{1/2} f(\delta x)$ in (35), we obtain

$$(35) = \delta \int \left| \int g\upsilon_1((s+y))\overline{g\upsilon_1(s)}\,ds \right|^2 dy.$$

Since $\|g\|_2 = \|f\|_2$, it follows that $K_\delta^2 = \delta K_1^2$ which completes the proof of the lemma. $\square$

PROOF OF LEMMA 19. The proof consists of a series of elementary steps. Let κ be the indicator function $\mathbf{1}_{[0,1/2]}$. Applying the Fourier transform from $\ell^2(\mathbb{Z})$ to $L^2(\mathbb{T})$, we get

$$\begin{aligned}\|\Pi\|_{2\to4}^2 &= \sup\{\|\overline{\Pi\mathbf{v}} \odot \Pi\mathbf{v}\|_2 : \mathbf{v} \in \ell^2(\mathbb{Z}), \|\mathbf{v}\|_2 \le 1\}\\ &= \sup\{\|\overline{\psi^{-1}(f\cdot\kappa)} \odot \psi^{-1}(f\cdot\kappa)\|_2 : f \in L^2(\mathbb{T}), \|f\|_2 \le 1\}.\end{aligned}$$

The properties of the Fourier transform further imply

$$\overline{\psi^{-1}(f\cdot\kappa)} \odot \psi^{-1}(f\cdot\kappa) = \psi^{-1}\big((f\cdot\kappa)^*\big) \odot \psi^{-1}(f\cdot\kappa) = \psi^{-1}\big((f\cdot\kappa)^* \star_{\mathbb{T}} (f\cdot\kappa)\big),$$

and since we convolve functions supported on $[0, 1/2]$, we might as well do the entire optimization on the real line to get

$$\|\Pi\|_{2\to4}^2 = \sup\{\|(f\cdot\kappa)^* \star (f\cdot\kappa)\|_2 : f \in L^2(\mathbb{R}), \|f\|_2 \le 1\}.$$

By translating the function $f \in L^2(\mathbb{R})$ via the map $f \mapsto f(\cdot + 1/4)$ in the above optimization problem, we see that $\|\Pi\|_{2\to4}^2 = K_{1/2}$. This equals $K_1/\sqrt{2}$ by Lemma 20. Now note that

$$\|(f\upsilon_1)^* \star (f\upsilon_1)\|_2 = \|\overline{\widehat{f\upsilon_1}} \cdot \widehat{f\upsilon_1}\|_2 = \|\widehat{f\upsilon_1}\|_4^2 = \|\mathtt{Sin}(\hat{f})\|_4^2$$

and so

$$K_1 = \sup\{\|\mathtt{Sin}(\hat{f})\|_4^2 : \hat{f} \in L^2(\mathbb{R}), \|\hat{f}\|_2 \le 1\} = \|\mathtt{Sin}\|_{2\to4}^2.$$

This completes the proof of the lemma. $\square$

As far as we know, the explicit value of constant K_1 of (34) is not known in the literature. However, the maximization problem given in (34) has been studied in Garsia, Rodemich and Rumsey (1969). We list below some of the interesting results from Garsia, Rodemich and Rumsey (1969):

- $K_1 = \sup\{(\int_{\mathbb{R}}(\int_{\mathbb{R}} f(x+t)f(t)\,dt)^2\,dx)^{1/2} : f \in \mathcal{F}\}$ where $\mathcal{F}$ is the class of all real-valued functions f satisfying $f(x) \ge 0$, $f(x) = f(-x)$ for all $x \in \mathbb{R}$, $f(x) \ge f(y)$ for $0 \le x \le y$ and $f(x) = 0$ for $|x| \ge 1/2$ and $\int_{-1/2}^{1/2} f^2(x)\,dx = 1$.
- There exists a unique $f \in \mathcal{F}$ such that $(\int_{\mathbb{R}}(\int_{\mathbb{R}} f(x+t)f(t)\,dt)^2\,dx)^{1/2} = K_1$.
- $K_1^2 = 0.686981293033114600949413\ldots!$

Acknowledgments. A. Sen thanks Manjunath Krishnapur for pointing out the usefulness of the invariance principle in the world of random matrices. Most of the research has been conducted while A. Sen visited University of Toronto and Technical University of Budapest in June 2010 and July 2011, respectively. B. Virág thanks Máté Matolcsi for many interesting discussions about the final optimization problem. We also thank him and Mihalis Kolountzakis for the reference Garsia, Rodemich and Rumsey (1969).

REFERENCES

ADAMCZAK, R. (2010). A few remarks on the operator norm of random Toeplitz matrices. *J. Theoret. Probab.* **23** 85–108. MR2591905

BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. MR1711663

BAI, Z. D. and YIN, Y. Q. (1988). Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* **16** 1729–1741. MR0958213

BOSE, A. and SEN, A. (2007). Spectral norm of random large dimensional noncentral Toeplitz and Hankel matrices. *Electron. Commun. Probab.* **12** 29–35 (electronic). MR2284045

BOSE, A. and SEN, A. (2008). Another look at the moment method for large dimensional random matrices. *Electron. J. Probab.* **13** 588–628. MR2399292

BOSE, A., SUBHRA HAZRA, R. and SAHA, K. (2010). Spectral norm of circulant type matrices with heavy tailed entries. *Electron. Commun. Probab.* **15** 299–313. MR2670197

BÖTTCHER, A. and GRUDSKY, S. M. (2000). *Toeplitz Matrices, Asymptotic Linear Algebra, and Functional Analysis*. Birkhäuser, Basel. MR1772773

BÖTTCHER, A. and SILBERMANN, B. (1999). *Introduction to Large Truncated Toeplitz Matrices*. Springer, New York. MR1724795

BÖTTCHER, A. and SILBERMANN, B. (2006). *Analysis of Toeplitz Operators*, 2nd ed. Springer, Berlin. MR2223704

BRYC, W., DEMBO, A. and JIANG, T. (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.* **34** 1–38. MR2206341

CHATTERJEE, S. (2005). A simple invariance theorem. Unpublished manuscript. Available at arXiv:math/0508213.

CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121

GARSIA, A., RODEMICH, E. and RUMSEY, H. (1969). On some extremal positive definite functions. *J. Math. Mech.* **18** 805–834. MR0251454

GRENANDER, U. and SZEGŐ, G. (1984). *Toeplitz Forms and Their Applications*, 2nd ed. Chelsea Publishing, New York. MR0890515

HAMMOND, C. and MILLER, S. J. (2005). Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *J. Theoret. Probab.* **18** 537–566. MR2167641

KARGIN, V. (2009). Spectrum of random Toeplitz matrices with band structure. *Electron. Commun. Probab.* **14** 412–421. MR2551851

LATAŁA, R. (1997). Estimation of moments of sums of independent real random variables. *Ann. Probab.* **25** 1502–1513. MR1457628

MECKES, M. W. (2007). On the spectral norm of a random Toeplitz matrix. *Electron. Commun. Probab.* **12** 315–325 (electronic). MR2342710

DEPARTMENTS OF MATHEMATICS
UNIVERSITY OF MINNESOTA
127 VINCENT HALL, 206 CHURCH ST. SE
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: arnab@math.umn.edu

DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF TORONTO
40 ST GEORGE ST.
TORONTO, ONTARIO
M5S 2E4, CANADA
E-MAIL: balint@math.toronto.edu