SPECTRUM OF RANDOM PERTURBATIONS OF TOEPLITZ MATRICES WITH
FINITE SYMBOLS

ANIRBAN BASAK∗, ELLIOT PAQUETTE‡, AND OFER ZEITOUNI§

Abstract. Let $T_N$ denote an $N \times N$ Toeplitz matrix with finite, $N$ independent symbol $a$. For $E_N$ a noise matrix satisfying mild assumptions (ensuring, in particular, that $N^{-1/2}\|E_N\|_{HS} \to N \to \infty 0$ at a polynomial rate), we prove that the empirical measure of eigenvalues of $T_N + E_N$ converges to the law of $a(U)$, where $U$ is uniformly distributed on the unit circle in the complex plane. This extends results from [2] to the non-triangular setup and non complex Gaussian noise, and confirms predictions obtained in [16] using the notion of pseudospectrum.

1. INTRODUCTION

Let $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ denote the unit circle in the complex plane. Let $a : S^1 \to \mathbb{C}$ be a function given by

$$a(\lambda) := \sum_{k=-\infty}^{\infty} a_k \lambda^k, \quad \lambda \in S^1,$$

where $\{a_k\}_{k=-\infty}^{\infty}$ is an absolutely summable complex valued sequence. We denote by $T_N := T_N(a)$ the Toeplitz matrix of dimension $N \times N$ with symbol $a$, given by

$${\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & \cdots & a_{N-1} \\ a_{-1} & a_0 & a_1 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ a_{-(N-1)} & \cdots & \cdots & a_{-2} & a_{-1} & a_0 \end{bmatrix}}.$$

From the definition it is clear that when $a$ is a Laurent polynomial, i.e.

$$a(\lambda) = \sum_{k=-d_1}^{d_2} a_k \lambda^k, \quad \text{for some } d_1, d_2 \geq 0,$$

then $T_N$ is a (finitely) banded Toeplitz matrix which can be thought of as a piece from an infinite Toeplitz matrix; we refer to such matrices as Toeplitz matrices with finite symbols.

For any $N \times N$ matrix $A_N$ we denote the empirical measure of its eigenvalues, or equivalently esd, the empirical spectral distribution, by $L^A_N$. That is,

$$L^A_N := \frac{1}{N} \sum_{i=1}^{n} \delta_{\lambda_i},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_N$ are the eigenvalues of $A_N$. In this paper, we find the limit of the empirical spectral distribution (esd) of random perturbations of Toeplitz matrices with finite symbols. This generalizes those results in [2] that deal with triangular Toeplitz matrices with finite symbols (and also with twisted Toeplitz matrices).
matrices, which we cannot generalize to the non-triangular case, see Remarks 1.7 and 1.9 below). In contrast with [2], we allow for rather general perturbations, as codified in Assumption 1.1.

**Assumption 1.1.** Let \( \{E_N\}_{N \in \mathbb{N}} \) be a sequence of matrices, with possibly complex valued entries, such that the followings hold:

(i) \[ \mathbb{E} \left[ \sum_{i,j=1}^{N} |e_{i,j}|^2 \right] = O(N^2), \]

where \( \{e_{i,j}\}_{i,j=1}^{N} \) are the entries of \( E_N \).

(ii) For any \( \alpha \in (0, \infty) \), there exists a \( \beta \in (0, \infty) \), depending only on \( \alpha \), so that for any fixed deterministic matrix \( M_N \) with \( \|M_N\| = O(N^\alpha) \), we have

\[ \mathbb{P} \left( s_{\min}(E_N + M_N) \leq N^{-\beta} \right) = o(1). \]

Let \( \mu_a \) denote the law of \( a(U) \), where \( U \) is a random variable uniformly distributed on the unit circle in the complex plane. Equipped with Assumption 1.1 we now state the main result of this paper.

**Theorem 1.2.** Let \( T_N \) be any \( N \times N \) Toeplitz matrix with a symbol \( a \), where \( a \) is a Laurent polynomial. Assume that \( \{E_N\}_{N \in \mathbb{N}} \) satisfy Assumption 1.1. Then, for any \( \gamma > \frac{1}{2} \), the esd \( L_N^{T_N + N^{-\gamma}E} \) of \( T_N + N^{-\gamma}E_N \) converges weakly, in probability, to \( \mu_a \).

Assumption 1.1(i) holds as soon as the second moment of each of the entries (of both complex and real parts) is uniformly bounded. By [15, Theorem 2.1], whenever the entries of \( E_N \) are i.i.d. (complex or real) with common \( N \)-independent distribution having a finite variance, Assumption 1.1(ii) holds. Therefore, Theorem 1.2 holds in that setup. In the next remark, we summarize other cases where Assumption 1.1, and hence Theorem 1.2, hold.

**Remark 1.3.** Assumption 1.1 holds under various relaxed assumptions on the noise matrix \( E_N \), which we list below.

1. When the entries of \( E_N \) are independent and dominated by a single distribution (in the Fourier-analytic sense) that has a \( \kappa \)-controlled second moment for some \( \kappa > 0 \), see [15, Definition 2.2 and Remark 2.8].
2. When the entries of \( E_N \) are independent, satisfy a uniform anti-concentration bound near 0, and have uniform lower bound on the truncated variance, see [4, Lemma A.1]. Furthermore, [15, Theorem 2.9] and [4, Lemma A.1] allow \( E_N \) to be a sparse random matrix.
3. When the entries of \( E_N \) have an inhomogeneous variance profile satisfying appropriate assumptions, by a recent result of Cook [6]. Specifically, by [6, Theorem 1.24], the assumption is satisfied when the variance profile is super-regular, see [6, Definition 1.23] for a precise formulation.
4. When \( E_N = \sqrt{N}U_N \), where \( U_N \) is a Haar distributed unitary matrix, see [18, Theorem 1.1].

**Remark 1.4.** We believe that the sequence \( N^{-\gamma} \) in Theorem 1.2 can be replaced by any sequence \( a_N \) satisfying \( \sqrt{N}a_N \to_{N \to \infty} 0 \). We chose to work with \( N^{-\gamma} \) in order to somewhat simplify the proofs.

**Remark 1.5.** A general notion developed to deal with perturbations of non-normal matrices is that of pseudospectrum, see [17] for an extensive review. This notion provides worse-case estimates and does not focus on the evaluation of limits of empirical measures under random perturbation. However, Theorem 1.2 is consistent with predictions based on pseudospectrum. For a thorough discussion of how pseudospectrum relates to Theorem 1.2, see [2, Section 1.3] and [16].

Our approach to the proof of Theorem 1.2 differs from the one employed in [2], which derived a deterministic equivalence that worked only for complex i.i.d. Gaussian perturbations (in particular, even real Gaussian perturbations are not covered by [2]). Instead, our approach is based on a perturbation idea that can be traced back in this context to [9]. See Section 1.1 below for a further discussion on this.

To describe the approach of this paper we first recall the important notion of logarithmic potential associated with a probability measure \( \mu \).
Definition 1.1 (Log-potential). For a probability measure \( \mu \) supported on the complex plane define its log-potential as follows:

\[
\mathcal{L}_\mu(z) := \int \log |z - x|d\mu(x), \quad z \in \mathbb{C}.
\]

As a first step, we will show that there exists a random matrix \( \Delta_N \), with a polynomially decaying spectral norm, such that the conclusion of Theorem 1.2 holds with \( N^{-\gamma}E_N \) replaced by \( \Delta_N \).

Theorem 1.6. Let \( T_N \) be any \( N \times N \) Toeplitz matrix with a symbol \( a \), where \( a \) is a Laurent polynomial. Then, there exists a random matrix \( \Delta_N \) with

\[
\mathbb{P}(\|\Delta_N\| \geq N^{-\gamma_0}) = o(1),
\]

for some \( \gamma_0 > 0 \), so that \( L_N^{T+\Delta} \) converges weakly, in probability, to \( \mu_a \). Equivalently, for Lebesgue almost every \( z \in \mathbb{C} \), \( \mathcal{L}_{L_N^{T+\Delta}}(z) \to \mathcal{L}_{\mu_a}(z) \), in probability.

Remark 1.7. We do not know the analogue of Theorem 1.6 for the twisted Toeplitz matrices considered in [2], and their non-triangular generalizations. For this reason, we cannot extend Theorem 1.2 to the general banded twisted case. See however Remark 1.9 below for the case of upper triangular twisted Toeplitz matrices.

We next state the replacement principle alluded to above. Here and in the sequel, \( B_C(c,R) \) denotes the open ball in the complex plane of center \( c \) and radius \( R \).

Theorem 1.8 (Replacement principle). Let \( A_N \) be any deterministic matrix with a bounded operator norm. Suppose \( \Delta_N \) and \( E_N \) are random matrices. Let \( \mu \) be a probability measure on \( \mathbb{C} \) whose support is contained in \( B_C(0,R_0/2) \) for some \( R_0 < \infty \). Assume the following.

(a) \( E_N \) and \( \Delta_N \) are independent. \( \Delta_N \) satisfies (1.1) and \( E_N \) satisfies Assumption 1.1.
(b) For Lebesgue a.e. \( z \in B_C(0,R_0) \), the empirical distribution of the singular values of \( A_N - z\text{Id}_N \) converges weakly, to the law induced by \( |X - z| \), where \( X \sim \mu \).
(c) For Lebesgue a.e. every \( z \in B_C(0,R_0) \),

\[
\mathcal{L}_{L_N^{A+\Delta}}(z) \to \mathcal{L}_{\mu}(z), \quad \text{as } N \to \infty, \quad \text{in probability.}
\]

Then, for any \( \gamma > \frac{1}{2} \), for Lebesgue a.e. every \( z \in B_C(0,R_0) \),

\[
\mathcal{L}_{L_N^{A+N^{-\gamma}E}}(z) \to \mathcal{L}_{\mu}(z), \quad \text{as } N \to \infty, \quad \text{in probability.}
\]

Theorem 1.8 is a generalization of the replacement lemma in [9, Theorem 5], with the advantage that it allows for more general noise models \( E_N \) and that it is stated directly in terms of logarithmic potentials and avoids the need to realize the \( * \)-limit of \( A_N \) as a regular element of a non-commutative probability space. It may be of independent interest beyond the study of perturbations of Toeplitz matrices.

Remark 1.9. Theorem 1.8 shows that [2, Theorem 4.1] remains true if one replaces there the complex Gaussian noise \( G_N \) by a noise \( E_N \) satisfying Assumption 1.1. This can be seen by using in Theorem 1.8 \( \Delta_N = N^{-\gamma}G_N \), and using [2, Lemma 4.6] to verify condition (b) of the theorem.

1.1. Related results and extensions. The study of the limiting esd of random perturbations of Toeplitz matrices can be traced back to [7] where in the simplest case of \( a(\lambda) = \lambda \), i.e. when the Toeplitz matrix is the standard Jordan matrix, they derive the limit by studying a relevant Grushin problem. On the other hand [9] derives the limit in the same set-up by first analyzing the limit of the log-potential of the esd of a specific (deterministic) perturbation of the Jordan matrix. Then they use an argument similar in spirit to that of Theorem 1.8 which allows them to replace that specific perturbation by a polynomially vanishing Gaussian perturbation. When the Toeplitz matrix is non-triangular with an arbitrary symbol it is not straightforward to find the required perturbation. Furthermore, it is not clear whether there exists at all some deterministic perturbation allowing one to apply [9, Theorem 5]. Theorem 1.6 of this paper shows that one can indeed find a random perturbation which does that job. Moreover, instead of appealing to [9, Theorem 5] we use Theorem 1.8 which enables us to consider a broad class of random perturbations.
Recently in [2] the limiting spectral distribution of Gaussian perturbation of triangular Toeplitz matrices has been derived by adopting a different strategy. The key to the proof in [2] lies in the following observation: If for Lebesgue a.e. $z \in \mathbb{C}$ the number of polynomially small singular values of $M_N - z \text{Id}_N$ is not too large, where $\{M_N\}_{N \in \mathbb{N}}$ is some sequence of matrices and $\text{Id}_N$ is the identity matrix, then the limiting esd of Gaussian perturbations of $M_N$ can be described by the Brown measure associated with the limiting operator. So it boils down to finding estimates on the number of small singular values. When $M_N = T_N$, a triangular Toeplitz (or a twisted Toeplitz) matrix, this task has been accomplished in [2]. If $T_N$ is a non-triangular matrix then the approach to finding bounds on the number of small singular values that is used in [2] fail.

Let us add that recent works of Sjöstrand and Vogel [12, 13] also deal with the limiting spectrum of Gaussian perturbations of general Toeplitz matrices. They use yet another strategy which is similar in spirit to the one adopted in [7]. In particular, their methods are robust enough that in [13] they apply them to Toeplitz operators with unbounded symbols.

There are several possible extensions of this paper that one can pursue. For example, one may be interested in understanding finer details of the spectrum, such as the behavior of the outliers of random perturbations of Toeplitz matrices. Building on the ideas of this paper the behavior of the outliers has been studied in a follow-up work [3].

Another interesting question would be to study the limiting esd of random perturbations of Toeplitz matrices with infinite symbols; as mentioned above, for certain perturbations this was achieved in [13]. A careful inspection of the proof of Theorem 1.2 of this paper reveals that one can build on the strategies developed in this paper to consider the case of Toeplitz matrices with a slowly growing bandwidth. For ease of writing and explanation we chose to work with a fixed bandwidth. The case of Toeplitz matrices with a general infinite symbol is at present beyond the scope of our methods.

Outline of the rest of the paper. We will show in Section 3 that Theorem 1.2 is an immediate consequence of Theorems 1.6 and 1.8. In Section 2 we provide the outlines of the proofs of Theorems 1.6 and 1.8. The proofs of these two theorems are carried out in Sections 4 and 5, respectively. Appendix A contains some algebraic results that are used in the proofs.

Acknowledgements. AB is partially supported by a Start-up Research Grant (SRG/2019/001376) from Science and Engineering Research Board of Govt. of India, and ICTS–Infosys Excellence Grant. OZ is partially supported by Israel Science Foundation grant 147/15 and funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (grant agreement number 692452). We thank the anonymous referees for helpful comments that enhanced the presentation of this paper.

2. Outlines of proofs of Theorems 1.6 and 1.8

We begin with an outline of the proof of Theorem 1.6. From [14, Theorem 2.8.3] and the fact that the support of $\mu_a$ is compact, it suffices to show that for Lebesgue a.e. $z$ in some large compact subset of the complex plane, $\mathcal{L}_{L_N^{T,a}}(z) \to \mathcal{L}_{\mu_a}(z)$ in probability. Toward this goal, it is useful first to obtain a different representation of the limit.

Lemma 2.1. Let

$$a(\lambda) = \sum_{k=-d}^{d_2} a_k \lambda^k,$$

for some $d_1, d_2 \in \mathbb{N}$. For any $z \in \mathbb{C}$ let $\lambda_1(z), \lambda_2(z), \ldots, \lambda_d(z)$ be the roots of the polynomial equation

$$P_{z,a}(\lambda) := (a(\lambda) - z) \cdot \lambda^d = 0,$$

where $d := d_1 + d_2$. Then, for any $z \in \mathbb{C}$,

$$\mathcal{L}_{\mu_a}(z) = \log |a_{d_1}| + \sum_{k=1}^{d} \log_+ |\lambda_k(z)|,$$

where for $x \geq 0$, $\log_+ (x) := \max\{\log x, 0\}$. 

The proof of Lemma 2.1 is a straightforward modification of that of [2, Lemma 4.3]. We omit the details.

We next sketch the proof of Theorem 1.6, in the special case where \(T_N\) is the Toeplitz matrix with symbol \(a(\lambda) = \lambda + \lambda^2\). Set \(T_N(z) := T_N - z1_{N}1_{N}\), where \(z \in \mathbb{C}\). By Lemma 2.1, the form of limiting log potential depends on the number of roots of the polynomial \(P_{2, a}(\lambda)\) greater than one in modulus. This yields (open) regions \(R_{\ell} \subset \mathbb{C}, \ell = 0, 1, 2\), whose boundaries have zero Lebesgue measure and the closure of whose union is \(\mathbb{C}\), so that for all \(z \in R_{\ell}\) there are exactly \(\ell\) roots of the equation \(\lambda + \lambda^2 - z = 0\) that are greater than one in modulus. Thus, to establish Theorem 1.6 we need to find a noise matrix \(\Delta_N\) supported on the lower left elements \(\Delta_N(1, 1), \Delta_N(N, 2), \Delta_N(N - 1, 1)\).

To derive (2.2), we expand the determinant of \(T_N(z) + \Delta_N\). The latter can be written as a linear combination of products of determinants of various sub-matrices of \(T_N(z)\) and \(\Delta_N\) (see Lemma A.1 below). We identify the dominant term in this expansion, as follows. Let \(A_N[X; Y]\) denote the sub-matrix of \(A_N\) induced by the rows and the columns indexed by \(X\) and \(Y\), respectively. Recalling Widom’s theorem concerning the determinant of a finitely banded Toeplitz matrix (see [11, 19]), we obtain

\[
\det(T_N(z)[X; Y]) \sim \begin{cases} |\lambda_1(z)|^N \cdot |\lambda_2(z)|^N & \text{if } X = Y = [N] \\ |\lambda_1(z)|^N & \text{if } X = [N - 1], Y = [N] \setminus \{1\} \\ 1 & \text{if } X = [N - 2], Y = [N] \setminus \{1, 2\} \end{cases},
\]

where we write \(a_N \sim b_N\) to indicate that there exists some absolute constant \(C > 0\) such that \(N^{-C}a_N \leq b_N \leq N^C a_N\), for all large \(N\).

From (2.3) we see that if \(z \in R_0\) or \(R_1\) then there are sub-matrices of \(T_N(z)\) whose determinants are of larger magnitude than that of \(T_N(z)\). We also note that the expansion of \(\det(T_N(z) + \Delta_N)\) has terms that are products of determinants of these sub-matrices and the determinant of relevant sub-matrices of the noise matrix \(\Delta_N\) (of fixed dimension), where the latter can be chosen to be non-zero and only polynomially (in \(N\)) decaying. It follows that if the determinants of those sub-matrices of \(T_N(z)\) are of maximal exponential growth among the determinants of all possible sub-matrices of \(T_N(z)\), then \(\frac{1}{N} \log |\det(T_N(z) + \Delta_N)|\) converges...
to the limit in (2.2). This not only explains how the limit arises but also identifies potential candidates for the dominant terms (depending on the location of $z$ in the complex plane) in the expansion of the determinant, and gives a heuristic for the proof of Theorem 1.6.

To justify this heuristic and obtain an actual proof of Theorem 1.6 in the case under consideration, it is natural to extend (2.3) and claim that

$$\sum_{\ell \neq k} P_{\ell}(z) = o \left( N^{-C} \cdot \prod_{j=1}^{k} |\lambda_j(z)|^N \right) = \Omega(|P_k(z)|), \quad z \in \mathcal{R}_k, \ k = 0, 1, 2,$$

for some large absolute constant $C$, with large probability, where $P_{\ell}(z)$ is the homogeneous polynomial of degree $\ell$ in the entries of $\Delta_N$, in the expansion of the determinant of $T_N(z) + \Delta_N$. In (2.4) we have used the standard notations $a_n = o(b_n)$ and $a_n = \Omega(b_n)$ to denote $\lim_{n \to \infty} a_n/b_n = 0$ and $\lim \inf_{n \to \infty} a_n/b_n > 0$, respectively. Finding bounds on $P_{\ell}(z)$ requires the same for $\det(T_N(z)[X;Y])$ for all subsets $X, Y \subseteq [N]$ such that $|X| = |Y| = N - \ell$. As $T_N(z)[X,Y]$ is not necessarily a Toeplitz matrix for arbitrary choices of $X, Y \subset [N]$ we can no longer rely on Widom’s result. We overcome this obstacle by noting that any upper triangular finitely banded Toeplitz matrix $T_N$ can be represented as a product of bidiagonal matrices, where the bidiagonal matrices depend on the roots of polynomial equation associated with the symbol of the Toeplitz matrix in context. Since the determinant of any sub-matrix of a bidiagonal matrix is easily computable (see Lemma A.3) one can then use the Cauchy-Binet theorem to find a bound on $\det(T_N(z)[X;Y])$. Using this and some combinatorial arguments, we then obtain the desired bound on $P_{\ell}(z)$ whenever the entries of $\Delta_N$ are uniformly polynomially vanishing.

We emphasize that the approach described above generalizes easily to triangular finitely banded Toeplitz matrix. The general case requires a modification, since non-triangular Toeplitz matrices cannot be decomposed into a product. We resolve this issue by using the following simple key observation: any Toeplitz matrix with finite symbol can be viewed as a sub-matrix of an upper triangular Toeplitz matrix with an another finite symbol of a slightly larger dimension. Using this observation, we can then follow the same scheme as described above to find an upper bound on $P_{\ell}(z)$.

To complete the proof of (2.4) we then need to find a lower bound on the predicted dominant term, $P_k(z)$. This is obtained using an anti-concentration estimate, which is shown to hold whenever the entries of $\Delta_N$ are assumed to have a bounded density, which we will impose since the matrix $\Delta_N$ is an auxiliary matrix and does not appear in the statement of our main theorem, Theorem 1.2. See Lemma 4.1 and Proposition 4.5. This will prove (2.4). To finish the proof of Theorem 1.6, we then obtain an (easy) matching upper bound on $P_k(z)$.

We next outline the proof of Theorem 1.8. It suffices to show that for Lebesgue a.e. $z$ in a compact subset of $\mathbb{C}$,

$$\lim_{N \to \infty} \left| \mathcal{L}_{L_N^{N+\Delta}}(z) - \mathcal{L}_{L_N^{N+\gamma}}(z) \right| = 0, \quad \text{in probability.}$$

Using the assumptions of Theorem 1.8 and standard perturbation results for the spectrum of Hermitian matrices, it readily follows that $\nu_{A_N+\Delta_N}^\mu$ and $\nu_{A_N+\gamma}^\mu$, the empirical distributions of the singular values of $A_N(z) + \Delta_N$ and $A_N(z) + N^{-\gamma}E_N$, respectively, have the same limit, and that limit is $\mu_z$, the law of $|X-z|$ where $X \sim \mu$. As $\log(\cdot)$ is unbounded both near 0 and $\infty$, the limit in (2.5) is not immediate from this. Using bounds on the Hilbert-Schmidt norms of the relevant matrices the singularity near $\infty$ can be taken care of. Treating the singularity of $\log(\cdot)$ near 0 involves two steps. As the integral of $\log(\cdot)$ near zero, with respect to $\mu_z$ is negligible, using assumptions (b)-(c) the same can be shown to hold for $\nu_{A_N+\gamma}^{\Delta_N}$. Hence, it suffices to show that the integral of $\log(\cdot)$ on the interval $(0, \varepsilon)$ with respect to $\nu_{A_N+\gamma}^{\Delta_N}$ goes to zero as $\varepsilon \downarrow 0$.

The latter is obtained by standard arguments, as follows. We use Assumption 1.1(ii) to deduce that it is enough to integrate $\log(\cdot)$ in $(N^{-\kappa}, \varepsilon)$ for some small constant $\kappa$. Now, using bounds on Hilbert-Schmidt norms of $E_N$ and $\Delta_N$ one can derive a bound on the difference of the Stieltjes transforms of $\nu_{A_N+\gamma}^{\Delta_N}$ and $\nu_{A_N+\gamma}^{\Delta_N}$. Using this, one obtains that the difference of the total mass of any interval near zero, under
\(\nu_{\Delta_N/N-\gamma E}\) and \(\nu_{\Delta_N/N}\) is negligible. Upon using an integration by parts, this gives the required control on the integral of \(\log(\cdot)\) near 0 under \(\nu_{\Delta_N/N-\gamma E}\) and completes the proof.

3. Proof of Theorem 1.2 using Theorems 1.6 and 1.8

We will take \(\Delta_N\) provided by Theorem 1.6, set \(\mu = \mu_a\) in Theorem 1.8, and verify that the hypotheses of the latter hold. Clearly, \(A_N\) has uniformly bounded operator norm. The assumption (a) is obvious. To see that assumption (b) holds, it is enough to check that for \(k\) positive integer,

\[
\lim_{N \to \infty} \frac{1}{N} \operatorname{tr}((z \text{Id}_N - T_N)(z \text{Id}_N - T_N)^*)^k - \mathbb{E}(|z - \sum_{i=-d_2}^{d_1} a_i U^i|^{2k}) = 0.
\]

To check (3.1) we first note that

\[
T_N = \sum_{m=0}^{d_1} a_m J_N^m + \sum_{n=1}^{d_2} a_{-n}(J_N^n),
\]

where \(J_N\) is the nilpotent matrix given by\( (J_N^i)_{i,j} = 1_{j = i+1} \). Using this observation we then expand \(\operatorname{tr}((z \text{Id}_N - T_N)(z \text{Id}_N - T_N)^*)^k\) and find out the limit of each term in term in that expansion. To work out this step we need to introduce some notation.

Let \(m := (m_1, \ldots, m_{2k})\) and \(n := (n_1, \ldots, n_{2k})\) with \(n_i, m_i\) non-negative integers bounded by \(\max(d_1, d_2)\), and set \(M := M_{m,n} := \sum m_i - \sum n_i\). We say that \((m, n)\) is balanced if \(M_{m,n} = 0\). Using (3.2) we find that

\[
\frac{1}{N} \operatorname{tr}((z \text{Id}_N - T_N)(z \text{Id}_N - T_N)^*)^k = \frac{1}{N} \sum_{m,n} b_{m,n}(z) \mathbb{E}[|J_N^{m_1}(J_N^{n_1})^{n_1} \cdots J_N^{m_{2k}}(J_N^{n_{2k}})^{n_{2k}}|]
\]

for appropriate coefficients \(b_{m,n}(z)\), while

\[
\mathbb{E}(|z - \sum_{i=-d_2}^{d_1} a_i U^i|^{2k}) = \sum_{m,n} b_{m,n}(z) \mathbb{E}[|U^{M_{m,n}}|],
\]

with the same coefficients \(b_{m,n}(z)\). Note that

\[
\operatorname{tr} [J_N^{m_1}(J_N^{n_1})^{n_1} \cdots J_N^{m_{2k}}(J_N^{n_{2k}})^{n_{2k}}] = 0
\]

if \((m, n)\) is not balanced, while

\[
\lim_{N \to \infty} \frac{1}{N} \operatorname{tr} [J_N^{m_1}(J_N^{n_1})^{n_1} \cdots J_N^{m_{2k}}(J_N^{n_{2k}})^{n_{2k}}] = 1
\]

if \((m, n)\) is balanced. Similarly, \(\mathbb{E}[|U^{M_{m,n}}|]\) equals 1 if \((m, n)\) is balanced, and vanishes otherwise. Combining these facts, we obtain (3.1), and thus verify that assumption (b) holds.

Assumption (c) holds because, from Theorem 1.6, we see that for Lebesgue a.e. \(z \in B_C(0, R_0)\),

\[
\mathcal{L}_{L_N^{1+\Delta}}(z) \to \mathcal{L}_{\mu_a}(z), \quad \text{in probability.}
\]

We have checked all assumptions of Theorem 1.8; applying the latter we conclude that for Lebesgue a.e. \(z \in B_C(0, R_0)\) and for any \(\gamma > \frac{1}{2}\),

\[
\mathcal{L}_{L_N^{1+\Delta}}(z) \to \mathcal{L}_{\mu_a}(z), \quad \text{in probability.}
\]

By the proof of [14, Theorem 2.8.3] and the fact that the support of \(\mu_a\) is compact, this implies the convergence in probability of \(L_N^{1+\Delta} - \gamma E\) to \(\mu_a\) in the vague topology, and hence in the weak topology. \(\square\)
4. Proof of Theorem 1.6

In this section we prove Theorem 1.6. As outlined in Section 2, the key is to establish (2.4). Turning to this task, introduce, for any $k \in [N]$,  
\begin{equation}
    P_k(z) := \sum_{X,Y \subset [N] \atop |X| = |Y| = k} (-1)^{\text{sgn}(\sigma_X)\sgn(\sigma_Y)} \det(T_N(z)[X^c;Y^c]) \cdot \det(\Delta_N[X;Y]),
\end{equation}
where $X^c := [N] \setminus X$, $Y^c := [N] \setminus Y$, and for $Z \in \{X,Y\}$ $\sigma_Z$ is the permutation on $[N]$ which places all the elements of $Z$ before all the elements of $Z^c$, but preserves the order of the elements within the two sets. Define  
\[ \mathcal{S}_{d_1,d_2} := \{(i,j) \in [N] \times [N] : i - j \notin \{-N-\ell; \ell = 1,\ldots,d_1\} \cup \{N-\ell; \ell = 1,\ldots,d_2\}\}. \]
To prove Theorem 1.6 we will choose $\Delta_N$ which satisfies the following band structure:  
\[ (\Delta_N)_{i,j} \neq 0 \quad \text{only if } i - j \in \mathcal{S}_{d_1,d_2}, \]
where $(\Delta_N)_{i,j}$ denotes the $(i,j)$-th entry of $\Delta_N$. That is, $\Delta_N$ has non-zero entries only in its lower left and upper right corners, and the widths of those corners are determined by $d_1$ and $d_2$, respectively. As indicated in (2.3) such a band structure is necessary (as we will see it is also sufficient) to have a non-zero contribution from the sub-matrices of $T_N(z)$ whose determinants are of larger magnitudes compared to that of the whole matrix, in the expansion of $\det(T_N(z) + \Delta_N)$. Recall from (2.2)-(2.3) that the dominant term depends on the number of roots of $P_{z,a}(\cdot)$ of (2.1), that are greater than one in modulus. Hence, we split the complex plane into regions according to the number of roots of $P_{z,a}(\cdot)$ with modulus greater than one, using the following notation.

Let $\{-\lambda_\ell(z)\}_{i=1}^{d}$, $d := d_1 + d_2^1$, be the roots of the equation $P_{z,a}(\cdot) = 0$, arranged so that $|\lambda_1(z)| \geq |\lambda_2(z)| \geq \cdots \geq |\lambda_d(z)|$. For $z \in \mathbb{C}$, let $d_0(z)$ denote the number of roots of the equation $P_{z,a}(\cdot) = 0$ that are greater than or equal to one in modulus. Fixing $R < \infty$, for $-d_2 \leq \vartheta \leq d_1$ we define  
\[ S_{\vartheta} := \{z \in B_{\mathbb{C}}(0,R) : d_1 - d_0(z) = \vartheta \text{ and } |\lambda_{d_0(z)}(z)| > 1 \geq |\lambda_{d_0(z)+1}(z)|\}. \]
Note that  
\[ B_{\mathbb{C}}(0,R) \setminus (\cup_{\ell=-d_2}^{d_1} S_{\vartheta}) \subset \{z \in B_{\mathbb{C}}(0,R) : P_{z,a}(\lambda) = 0 \text{ for some } \lambda \in S^1\}. \]
If $P_{z,a}(\lambda) = 0$ for some $\lambda \in S^1$ then we also have that  
\[ z = \sum_{\ell=-d_2}^{d_1} a_\ell \lambda^\ell. \]
Therefore $B_{\mathbb{C}}(0,R) \setminus (\cup_{\ell=-d_2}^{d_1} S_{\vartheta})$ is contained in a set of Lebesgue measure zero and hence it is enough to consider $z \in \cup_{\ell=-d_2}^{d_1} S_{\vartheta}$. Further let $N$ be the set of $z$’s for which $P_{z,a}(\cdot)$ admits a double root. It follows from [5, Lemma 11.4] that the cardinality of $N$ is at most finite.

The next lemma identifies the dominant term in the expansion of $\det(T_N(z) + \Delta_N)$.

**Lemma 4.1.** Fix $\vartheta$ such that $-d_2 \leq \vartheta \leq d_1$. Let $\Delta_N$ be such that  
\[ (\Delta_N)_{i,j} = N^{-\gamma} \delta_{i,j} 1_{\{(i,j) \in \mathcal{S}_{d_1,d_2}\}}, \quad i,j \in [N], \]
for some $\gamma > d$, where $\{\delta_{i,j}\}$ are uniformly bounded real valued independent random variables with uniformly bounded densities with respect to the Lebesgue measure. Then, for Lebesgue a.e. $z \in S_{\vartheta}$, and any $\varepsilon_0 > 0$,  
\begin{equation}
\lim_{N \to \infty} \mathbb{P} \left( \frac{|P_{z,a}(z)|}{|a_{d_1}|^{N^2} \prod_{i=1}^{d_1} |\lambda_i(z)|^N} \leq N^{-(\gamma + |\vartheta| + \varepsilon_0)} \right) = 0,
\end{equation}
where an empty product by convention is set to one.

\textsuperscript{1}Hereafter $\{-\lambda_\ell(z)\}$ will denote the roots of the equation $P_{z,a}(\lambda) = 0$. This change in notation is adopted to avoid the unnecessary appearance of signs in the determinant of the sub-matrices of $J_N - \lambda \text{Id}_N$. 

---

\[ \mathbb{P} \]
Lemma 4.1 yields a lower bound on the order of the magnitude of the predicted dominant term in the expansion of $\det(T_N(z) + \Delta_N)$. Next we need to show that the sum of the rest of the terms is of smaller order. To show this, we split it into two sums: $\sum_{k \neq |\delta|} P_k(z)$ and $\sum_{k \neq |\delta|} P_k(z)$. The second sum will be shown to be polynomially small compared to the leading term, whereas the first will be shown to be exponentially small. This is the content of the two following lemmas.

**Lemma 4.2.** Let $\delta, \Delta_N$, and $\gamma_\ast$ be as in Lemma 4.1. Then, for Lebesgue a.e. $z \in S_\delta$,

$$
\lim_{N \to \infty} \frac{N^{\gamma_\ast|\delta| + \frac{2d - d}{d}} \left| \sum_{k=0}^N P_k(z) \right|}{|a_{d_1}|^N \prod_{i=1}^{d_1} |\lambda_i(z)|^N} = 0.
$$

**Lemma 4.3.** Under the same set-up as in Lemma 4.2, for Lebesgue a.e. $z \in S_\delta$, we have

$$
\lim_{N \to \infty} \frac{\left| \sum_{k=0}^{d-1} P_k(z) \right|}{|a_{d_1}|^N \prod_{i=1}^{d_1} |\lambda_i(z)|^N (1 - \varepsilon)^N} = 0,
$$

for some small constant $\varepsilon := \varepsilon(z, a) \in (0, 1)$.

The proofs of Lemmas 4.2 and 4.3 are in Section 4.1, while the proof of Lemma 4.1 is postponed to Section 4.2. To complete the proof of Theorem 1.6, we will also need an upper bound on the dominant term, which is contained in the next lemma, whose proof is deferred to Section 4.2.

**Lemma 4.4.** Under the same set-up as in Lemma 4.1, for Lebesgue a.e. $z \in S_\delta$, there exists a constant $C_0$ depending on $z$ and $a$ only, so that

$$
\limsup_{N \to \infty} \frac{P_{|\delta|}(z)}{|a_{d_1}|^N \prod_{i=1}^{d_1} |\lambda_i(z)|^N} \leq C_0.
$$

Equipped with these four lemmas, we now complete the proof of Theorem 1.6.

**Proof of Theorem 1.6.** From the definition of $\Delta_N$ it follows that there are at most $d$ non-zero entries in each row of $\Delta_N \Delta_N^\ast$. Furthermore, each entry of $\Delta_N \Delta_N^\ast$ is at most $O(N^{-2\gamma_\ast})$. Therefore, by the Gershgorin circle theorem, it follows that $||\Delta_N|| = O(N^{-\gamma_\ast})$, establishing the desired property (1.1). Next, as in the proof of Theorem 1.2, the weak convergence of $L_{N+\Delta}^T$ to $\mu_a$ follows from the convergence, for Lebesgue a.e. $z \in B_C(0, R_0)$, of the log-potentials:

$$
\mathcal{L}_{L_{N+\Delta}^T}(z) = \frac{1}{N} \log |\det(T_N(z) + \Delta_N)| \rightarrow \mathcal{L}_{\mu_a}(z), \quad \text{in probability.}
$$

To this end, recalling the definition of $P_k(z)$ from (4.1) and applying Lemma A.1 we have that

$$
\frac{1}{N} \log |\det(T_N(z) + \Delta_N)| = \frac{1}{N} \log \left| \sum_{k=0}^N P_k(z) \right| = \frac{1}{N} \log |P_{|\delta|}(z)| + \frac{1}{N} \log \left| 1 + \frac{\sum_{k \neq |\delta|} P_k(z)}{P_{|\delta|}(z)} \right|,
$$

for any integer $d$ between $-N$ and $N$. Setting $\varepsilon_0 = \frac{2d - d}{d} > 0$ in Lemma 4.1 and combining Lemmas 4.1-4.3 we note that for Lebesgue a.e. $z \in S_\delta$, there exists an event of probability at least $1 - o(1)$ such that, on that event, we have

$$
\left| \sum_{k \neq |\delta|} P_k(z) \right| \leq N^{-\frac{2d - d}{d}},
$$

for all large $N$. This in turn implies that

$$
\frac{1}{N} \log \left| 1 + \frac{\sum_{k \neq |\delta|} P_k(z)}{P_{|\delta|}(z)} \right| \rightarrow 0, \quad \text{in probability,}
$$

where $a \geq 0$.
for Lebesgue a.e. every \( z \in S_0 \). Finally combining Lemmas 4.1 and 4.4 we obtain that for Lebesgue a.e. every \( z \in S_0 \),

\[
\frac{1}{N} \log P_{0}(z) \rightarrow \log |a_{d_{1}}| + \sum_{\ell=1}^{d_{0}(z)} \log |\lambda_{\ell}(z)| = \mathcal{L}_{\mathcal{P}_{a}}(z), \quad \text{in probability.}
\]

Combining (4.5)-(4.7) we now deduce (4.4) for Lebesgue a.e. every \( z \in S_0 \) and any integer \( d \) such that \(-d_{2} \leq d \leq d_{1}\).

This completes the proof. \( \square \)

4.1. Upper bound on non-dominant terms. Recall from Section 2 that to establish bounds on the predicted non-dominant terms, one uses the fact any upper triangular Toeplitz matrix with a finite symbol can be expressed as a product of bidiagonal matrices. To use the same representation for a non-triangular Toeplitz matrix we view it as a sub-matrix of an upper triangular Toeplitz matrix of a slightly larger dimension. Toward this end, we introduce the following definition.

**Definition 4.1** (Toeplitz with a shifted symbol). Let \( T_{N} \) be a Toeplitz matrix with finite symbol \( a(\lambda) = \sum_{\ell=-d}^{d_{1}} a_{\ell} \lambda^{\ell} \) and as before \( d = d_{1} + d_{2} \). For \( \bar{d}_{1}, \bar{d}_{2} \in \mathbb{N} \) such that \( \bar{d}_{1} + \bar{d}_{2} = d \) and \( z \in \mathbb{C} \), set \( T_{N}(z; \bar{d}_{1}, \bar{d}_{2}) := T_{N}(z; d_{1}, d_{2}) \) to be the \( N \times N \) Toeplitz matrix with first row and column

\[
(a'_{0-d_{2}}, a'_{0-d_{2}+1}, \ldots, a'_{0}, 0, \ldots, 0) \quad \text{and} \quad (a'_{d_{1}-d_{1}}, a'_{d_{1}-d_{1}-1}, \ldots, a'_{-d_{2}}, 0, \ldots, 0)^{T},
\]

respectively, where \( a'_{j} := a_{j} - z \cdot 1_{(j=0)}, j = -d_{2}, -d_{2} + 1, \ldots, d_{1} \). That is,

\[
T_{N}(z; \bar{d}_{1}, \bar{d}_{2}) :=
\begin{bmatrix}
0 & \cdots & 0 & a_{-d_{2}} & \cdots & \cdots & \cdots & \cdots & a_{d_{1}-d_{1}} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & a_{0} & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & a_{0} & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & a_{0} & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & a_{-d_{2}} & \cdots & \cdots & \cdots & \cdots & 0 \\
\end{bmatrix}
\]

From Definition 4.1, it follows that

\[
T_{N}(z) = T_{N}(z; d_{1}, d_{2}) = T_{N+d_{2}}(z; d, 0)[[N]; [N + d_{2}] \setminus [d_{2}]].
\]

Note that \( T_{N+d_{2}}(z; d, 0) \) is an upper triangular Toeplitz matrix. Since \( \{ -\lambda_{\ell}(z) \}_{\ell=1}^{d} \) are the roots of the equation \( P_{z,a}(\lambda) = 0 \) we obtain that

\[
T_{N+d_{2}}(z; d, 0) = \sum_{\ell=0}^{d} (a_{-d_{2}} - z\delta_{\ell, d_{2}}) J_{N+d_{2}}^{\ell} = a_{d_{1}} \prod_{\ell=1}^{d} (J_{N+d_{2}} + \lambda_{\ell}(z) \text{Id}_{N+d-2}),
\]

where we recall that \( J_{n} \) is the nilpotent matrix given by \((J_{n})_{i,j} = 1_{j=i+1}, \) for \( i, j \in [n] \).
Hence, recalling the definition of \( \{ P_k(z) \}_{k=1}^N \) from (4.1), applying the Cauchy-Binet theorem, and writing \( S + \ell := \{ x + \ell, x \in S \} \) for any set of integers \( S \) and an integer \( \ell \), we obtain that
\[
(4.8) \quad P_k(z) = \sum_{X,Y \subset [N]} (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} \det(T_{N+d_2}(z; d, 0)[X^c; Y^c + d_2]) \det(\Delta_N[X; Y])
\]
\[
= \sum_{X,Y \subset [N]} \sum_{|X|=|Y|=k} (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} a_{d_1}^{N-k} \prod_{i=1}^d \det((J_{N+d_2} + \lambda_i(z) \text{Id}_{N+d_2})[X_i; \tilde{X}_{i+1}]) \cdot \det(\Delta_N[X; Y]),
\]
where
\[
(4.9) \quad X_1 := X_1(X) := X \cup [N + d_2] \setminus [N], \quad X_{d+1} := X_{d+1}(Y) := (Y + d_2) \cup [d_2],
\]
and \( \tilde{Z} := [N + d_2] \setminus Z \) for any set \( Z \subset [N + d_2] \). Equipped with this preparatory decomposition of \( P_k(z) \), we are now ready to step into the proof of Lemma 4.2.

**Proof of Lemma 4.2.** From the definition of the noise matrix it follows that the number of non-zero rows (and also non-zero columns) in \( \Delta_N \) is at most \( d \). This means that \( P_k(z) = 0 \) for any \( k > d \). Therefore, it is enough to show that (4.3) holds with the sum in the numerator being replaced by \( P_k(z) \), where \( |b| < k \leq d \).

To achieve this, we need to simplify (4.8); this simplification, summarized in (4.12) and (4.13) below, will also be useful in the proof of Lemma 4.3. From (4.8)-(4.9), we see that each \( X_i \) is of cardinality \( k + d_2 \). Therefore, we write
\[
X_i = \{ x_{i,1} < x_{i,2} < \cdots < x_{i,k+d_2} \}
\]
and for brevity we also denote \( X_k := (X_1, X_2, \ldots, X_{d+1}) \). Applying Lemma A.3 we see that
\[
\prod_{i=1}^d \det((J_{N+d_2} + \lambda_i(z) \text{Id}_{N+d_2})[X_i; \tilde{X}_{i+1}]) \neq 0
\]
only when \( X_k \in L_{\ell,k} \) for some \( \ell := (\ell_1, \ell_2, \ldots, \ell_d) \) with \( 0 \leq \ell_i \leq N - k \leq N + d_2 \), \( i = 1, 2, \ldots, d \), where
\[
L_{\ell,k} := \{ X_k : 1 \leq x_{i+1,1} \leq x_{i,1} < x_{i+1,2} \leq x_{i,2} < \cdots < x_{i+1,k+d_2} \leq x_{i,k+d_2} \leq N + d_2
\]
and \( x_{i+1,1} + \sum_{j=2}^{k+d_2} (x_{i+1,j} - x_{i,j-1}) + (N + d_2 - x_{i,k+d_2}) = \ell_i + k + d_2 \), for all \( i = 1, 2, \ldots, d \). Since
\[
x_{i+1,1} + \sum_{j=2}^{k+d_2} (x_{i+1,j} - x_{i,j-1}) + \sum_{j=1}^{k+d_2} (x_{i,j} - x_{i+1,j}) + (N + d_2 - x_{i,k+d_2}) = N + d_2
\]
we have the following following equivalent representation of \( L_{\ell,k} \):
\[
(4.10) \quad L_{\ell,k} := \{ X_k : 1 \leq x_{i+1,1} \leq x_{i,1} < x_{i+1,2} \leq x_{i,2} < \cdots < x_{i+1,k+d_2} \leq x_{i,k+d_2} \leq N + d_2,
\]
\[
\sum_{j=1}^{k+d_2} (x_{i,j} - x_{i+1,j} + 1) = \hat{\ell}_i; \quad i = 1, 2, \ldots, d_0,
\]
and \( x_{i+1,1} + \sum_{j=2}^{k+d_2} (x_{i+1,j} - x_{i,j-1}) + (N + d_2 - x_{i,k+d_2}) = \hat{\ell}_i + k + d_2; \quad i = d_0 + 1, d_0 + 2, \ldots, d \},
\]
where
\[
\hat{\ell}_i := \begin{cases} \ell_i & \text{if } i > d_0 \\ N + d_2 - \ell_i & \text{if } i \leq d_0. \end{cases}
\]
We also note that in (4.8) the outer sum is over \( X, Y \subset [N] \) and due to the constraint (4.9) we only need to consider \( X_k \in S_{\ell,k} \), where

\[
Q_{\ell,k} := \{ X_k \in L_{\ell,k} : x_{1,k+j} = N + j; \ j \in [d_2] \quad \text{and} \quad x_{d_2+1,j} = j; \ j \in [d_2] \}.
\]

Thus applying Lemma A.3 again, from (4.8) we now deduce that

\[
P_k(z) = \frac{a_{-N-k}}{\prod_{i=1}^{d_0} \lambda_i(z)^{N+d_2}} \sum_{\ell} Q_{\ell,k}(z),
\]

where

\[
Q_{\ell,k}(z) := \prod_{i=1}^{d_0} \lambda_i(z)^{-\hat{\ell}_i} \prod_{i=d_0+1}^{d} \lambda_i(z)^{\hat{\ell}_i} \sum_{X_k \in S_{\ell,k}} (-1)^{\sigma(x)} \det(\Delta_N[X;Y]),
\]

\( X := X(X_1) := X_1 \cap [N] \), and \( Y := Y(X_{d+1}) := (X_{d+1} - d_2) \cap [N] \).

Returning to the proof of the lemma, it suffices to bound \( Q_{\ell,k} \). Turning to this task, we assume without loss of generality that \(|(\Delta_N)_{i,j}| \leq 1\). This implies that

\[
|\det(\Delta_N[X;Y])| \leq N^{-\gamma_*} k!,
\]

for every \( X, Y \subset [N] \) such that \(|X| = |Y| = k\). On the other hand, the definition of \( d_0 = d_0(z, a) \) and the fact that \( z \in S_k \) imply that there are no roots of \( P_{z,a}() \) on the unit circle, hence we deduce that there exists \( \varepsilon_* = \varepsilon_*(z, a) > 0 \), such that

\[
\max \left\{ \sum_{i=1}^{d_0} \lambda_i(z)^{-1}, \max_{i=d_0+1} \lambda_i(z) \right\} \leq 1 - \varepsilon_*,
\]

Hence,

\[
N^{\gamma_* |b|} \left| \sum_{\ell} Q_{\ell,k}(z) \right| \leq k! N^{\gamma_* (|b| - k)} \sum_{\ell} (1 - \varepsilon) \sum_{i=1}^{\tilde{\ell}_i} \lambda_i(z)^{-1} |Q_{\ell,k}| \leq d！N^{-\gamma_*} \sum_{\ell} (1 - \varepsilon)^{\tilde{\ell}_i} |Q_{\ell,k}|,
\]

where the last inequality follows from the fact that \(|b| < k \leq d\). To finish the proof it remains to find an upper bound on the cardinality of \( S_{\ell,k} \). We claim that

\[
\left| S_{\ell,k} \right| \leq \left( \frac{N}{d} \right) \prod_{i=1}^{d_0} \left( \frac{\hat{\ell}_i - 1}{k + d_2 - 1} \right) \cdot \prod_{i=d_0+1}^{d} \left( \frac{\hat{\ell}_i + k + d_2}{k + d_2} \right).
\]

Equipped with (4.17), it now follows from (4.16) that

\[
N^{\gamma_* |b|} \left| \sum_{\ell} Q_{\ell,k}(z) \right| = O(N^{d - \gamma_*}).
\]

Since \( a_{d_2} \neq 0 \) implies that \( \{ \lambda_i(z) \} \) are bounded away from zero, (4.18) together with (4.12) yield (4.3).

It remains to establish the bound (4.17). To this end, set

\[
\delta_{i,j} := \delta_{i,j}(X_k) := \begin{cases} x_{i,j} - x_{i+1,j} & \text{for } i \in [d_0] \text{ and } j \in [k + d_2] \\ x_{i+1,j} & \text{for } i \in [d] \setminus [d_0], \ j = 1 \\ x_{i+1,j} - x_{i,j-1} & \text{for } i \in [d] \setminus [d_0], \ j \in [k + d_2] \setminus \{1\} \\ N + d_2 - x_{i,k+d_2} & \text{for } i \in [d] \setminus [d_0], \ j = k + d_2 + 1. 
\end{cases}
\]

For the \( \{x_{i,j}\} \) to satisfy \( X_k \in S_{\ell,k} \), we observe that the \( \{\delta_{i,j}(X_k)\}'s \) can be chosen in at most

\[
\prod_{i=1}^{d_0} \left( \frac{\hat{\ell}_i - 1}{k + d_2 - 1} \right) \cdot \prod_{i=d_0+1}^{d} \left( \frac{\hat{\ell}_i + k + d_2}{k + d_2} \right)
\]

ways. Next, recall that \( X_k \in S_{\ell,k} \) implies that

\[
x_{1,k+j} = N + j, \ j = 1, 2, \ldots, d_2.
\]
Thus, \( \{x_{i,\ell}\}_{i=1}^{k} \) and \( \{\delta_{i,j}(X_k)\} \) automatically fix \( X_k \). Since the number of choices of \( \{x_{i,\ell}\}_{i=1}^{k} \) is at most \( \binom{d}{k} \), as \( k \leq d \), for all large \( N \), the claim \( (4.17) \) follows from \( (4.20) \). The proof of the lemma is now complete.

Next we show that for \( z \in S_0 \) the sum \( \sum_{k<|\mathfrak{d}|} P_k(z) \) is of smaller order compared to the dominant term \( P_{|\mathfrak{d}|}(z) \).

**Proof of Lemma 4.3.** We first claim that for any \( \mathfrak{d} \), the set \( \mathfrak{L}_k \) (see \( (4.11) \)) being nonempty forces either \( \sum_{i=0}^{d} \hat{\ell}_i \) or \( \sum_{i=1}^{d} \hat{\ell}_i \) to be close to \( N \), depending on whether \( \mathfrak{d} > 0 \) or \( \mathfrak{d} < 0 \). This observation will be then combined with the bounds \( (4.15) \) and \( (4.17) \) to complete the proof.

Consider first the case \( \mathfrak{d} = d_1 - d_0 > 0 \). For any \( \mathfrak{d} < \mathfrak{d} \) we have that \( d_0 + k + d_2 + 1 \leq d_1 + d_2 = d \) and hence for any \( X_k \in \mathfrak{L}_k \subset \mathcal{L}_{k,\mathfrak{d}} \),

\[
x_{d_0+2,1} + (x_{d_0+3,2} - x_{d_0+2,1}) + \cdots + \sum_{k+d_2+1} (x_{d_0+k+d_2+1,k+d_2} - x_{d_0+k+d_2,k+d_2}) = \sum_{\ell=1}^{k+d_2+1} \delta_{d_0+\ell}(X_k) = N + d_2.
\]

As \( \delta_{i,j}(X_k) \leq \hat{\ell}_i + k + d_2 \) for \( i \in [d] \setminus \{d_0\} \) and \( j \in [k + d_2 + 1] \), it further implies that if \( \mathfrak{L}_k \neq \emptyset \) then we must have

\[
\sum_{i=d_0+1}^{d} \hat{\ell}_i \geq N + d_2 - (k + d_2)(d - d_0).
\]

Next we consider the case \( \mathfrak{d} < 0 \). For any \( X_k \in \mathfrak{L}_k \) we have that \( x_{1,k+1} = N + 1 \). Therefore

\[
N + 1 - x_{d_0+1,k+1} = x_{1,k+1} - x_{d_0+1,k+1} = \sum_{\ell=1}^{d_0} \delta_{\ell,k+1}(X_k).
\]

On other hand, we have that \( x_{d+1,\ell} = \ell \) for \( \ell \in [d_2] \). Since \( x_{i+1,\ell} \leq x_{i,\ell} < x_{i+1,\ell+1} \) for any \( \ell \in [k + d_2 - 1] \), and \( \{x_{i,\ell}\} \) are integers, using induction, we further obtain that

\[
x_{d_0+1,\ell} = \ell, \quad \ell \in [d_2 - (d - d_0)]
\]

for any \( X_k \in \mathfrak{L}_k \). As \( k + 1 \leq |\mathfrak{d}| = d_0 - d_1 = d_2 - (d - d_0) \) we find that \( x_{d_0+1,k+1} = k + 1 \). Hence, from \( (4.23) \) we deduce that

\[
\sum_{\ell=1}^{d_0} \delta_{\ell,k+1}(X_k) = N - k,
\]

for any \( X_k \subset \mathfrak{L}_k \). Noting that \( \delta_{i,k+1}(X_k) \leq \hat{\ell}_i - 1 \) for all \( i \in [d_0] \), we obtain

\[
\sum_{i=1}^{d_0} \hat{\ell}_i \geq N - (k + d_0).
\]

Thus, \( (4.22) \) and \( (4.24) \) implies that, if \( k < |\mathfrak{d}| \) then

\[
\mathfrak{L}_k \neq \emptyset \implies \sum_{i=1}^{d} \hat{\ell}_i \geq N - d^2.
\]

To complete the proof of the lemma we now use \( (4.13)-(4.15) \) and \( (4.17) \) to conclude that for any \( k < |\mathfrak{d}| \),

\[
\left| \sum_{\ell: \mathfrak{L}_k \neq \emptyset} Q_{\mathfrak{L}_k}(z) \right| \leq d! (1 - \varepsilon)(N - d^2) \sum_{i=1}^{N+d_2} |\mathfrak{L}_{i,k}| \leq (1 - \varepsilon)^2 N
\]

for all large \( N \), for some sufficiently small \( \varepsilon > 0 \), depending only on \( z \) and \( a \). The proof finishes upon using \( (4.12) \). \( \square \)
4.2. Lower and upper bounds on the dominant term. We will first prove Lemma 4.1, which is a lower bound on the dominant term. The proof is based on the following elementary anti-concentration bound for homogeneous polynomials of independent random variables, which may be of independent interest.

**Proposition 4.5.** Fix $k,n \in \mathbb{N}$ and let $\{U_i\}_{i=1}^n$ be a sequence of independent real-valued random variables, whose law possesses a density with respect to the Lebesgue measure which is uniformly bounded by one. Let $Q_k(U_1, U_2, \ldots, U_n)$ be a homogenous polynomial of degree $k$ such that the degree of each variable is at most one. That is,

$$Q_k(U_1, U_2, \ldots, U_n) := \sum_{I \in \binom{[n]}{k}} a(I) \prod_{i \in I} U_i,$$

for some collection of complex valued coefficients $\{a(I) : I \in \binom{[n]}{k}\}$, where $\binom{[n]}{k}$ denotes the set of all $k$ distinct elements of $[n]$.

Assume that there exists an $\mathcal{I}_0 \subset \binom{[n]}{k}$ such that $|a(\mathcal{I}_0)| \geq c_\ast$ for some absolute constant $c_\ast > 0$. Then for any $\varepsilon \in (0, e^{-1}]$ we have

$$\Pr(|Q_k(U_1, U_2, \ldots, U_n)| \leq \varepsilon) \leq (8e)^k (c_\ast \wedge 1)^{-1} \varepsilon \left( \log \left( \frac{1}{\varepsilon} \right) \right)^{k-1}.$$  

**Proof.** As the densities of $\{U_i\}_{i \in [n]}$ are uniformly bounded by one, the desired anti-concentration property is immediate for $k = 1$. To prove the general case, we proceed by induction. To this end, we introduce some notation. Order the elements of $\mathcal{I}_0$ and denote them by $i_0^0, i_1^0, \ldots, i_k^0$. For $j \leq k$, define $I_j^0 := \{i_0^0, i_1^0, \ldots, i_j^0\}$. Set

$$Q_k^0 := Q_k(U_i; i \notin I_j^0) := \sum_{I : I \supset I_j^0} a(I) \prod_{\ell \in I, \ell \notin I_j^0} U_\ell$$

and

$$Q_k^1 := Q_k(U_i; i \notin I_k^0) := \sum_{I : I \cap I_k^0 = \emptyset} a(I) \prod_{\ell \in I, \ell \notin I_k^0} U_\ell.$$  

For $1 \leq j \leq k - 1$, we iteratively define

$$Q_j^0 := Q_j(U_i; i \notin I_j^0) := \sum_{I : I \supset I_j^0} a(I) \prod_{\ell \in I, \ell \notin I_j^0} U_\ell$$

and

$$Q_j^1 := Q_j(U_i; i \notin I_j^0) := \sum_{I : I \cap I_j^0 = \emptyset} a(I) \prod_{\ell \in I, \ell \notin I_j^0} U_\ell.$$  

Equipped with the above notations we see that

$$Q_k(U_1, U_2, \ldots, U_n) =: Q_{k+1}^0 = U_{i_0} \cdot Q_k^0 + Q_k^1, \quad Q_{j+1}^0 = U_{i_0^j} \cdot Q_j^0 + Q_j^1, \quad j = 1, 2, \ldots, k - 1,$$

and $Q_1^0 = a(\mathcal{I}_0)$. We will prove inductively that

$$\Pr\left(|Q_j^0| \leq \varepsilon\right) \leq (8e)^{j-1} (c_\ast \wedge 1)^{-1} \varepsilon \left( \log \left( \frac{1}{\varepsilon} \right) \right)^{j-2}, \quad j = 2, 3, \ldots, k + 1,$$

from which the desired anti-concentration bound follows by taking $j = k + 1$. Hence, it only remains to prove (4.25).

For $j = 2$, $Q_j^0$ is a homogeneous polynomial of degree 1 in the variables $U_i$, and (4.25) follows from the assumptions on $\{U_i\}_{i=1}^n$ and the fact that $|a(\mathcal{I}_0)| \geq c_\ast$. Assuming that (4.25) holds for $j = j_\ast$ and fixing $\delta \in (0, 1)$, we have that with $C_j := (8e)^{j-1} (c_\ast \wedge 1)^{-1}$,

$$\Pr\left(|Q_{j_\ast+1}^0| \leq \varepsilon\right) \leq \Pr\left(|Q_{j_\ast}^0| \leq \delta\right) + \Pr\left(\left| \frac{Q_{j_\ast}^1}{Q_{j_\ast}^0} \right| U_i, i \notin I_j^0 \leq \frac{\varepsilon}{|Q_{j_\ast}^0|} \right) \cdot 1\left(|Q_{j_\ast}^0| \geq \delta\right)$$

$$\leq C_j \delta \log \left( \frac{1}{\delta} \right)^{j_\ast-2} + 2\varepsilon \cdot \mathbb{E}\left[|Q_{j_\ast}^0|^{k-1} 1\left(|Q_{j_\ast}^0| \geq \delta\right)\right],$$

where we have used the fact that $Q_{j_\ast}^1$ and $Q_{j_\ast}^0$ are independent of $U_{i_0^j}$, and the bound on the density for the latter. Using integration by parts, for any probability measure $\mu$ supported on $[0, \infty)$ we have that

$$\int_{\delta}^{e^{-1}} x^{-1} \, d\mu(x) = \mu([\delta, 1]) + \int_{\delta}^{e^{-1}} \frac{\mu([\delta, t])}{t^2} \, dt.$$
Therefore, using the induction hypothesis,
\[
\mathbb{E} \left[ |Q_{j,1}^0|^{-1} 1 \left( |Q_{j,1}^0| \geq \delta \right) \right] \leq e + \mathbb{E} \left[ |Q_{j,1}^0|^{-1} 1 \left( |Q_{j,1}^0| \in [\delta, e^{-1}] \right) \right] \leq 2e + \int_{\delta}^{e^{-1}} \frac{P(|Q_{j,1}^0| \leq t)}{t^2} \, dt
\]
\[
\leq 2e + C_{j,1} \int_{\delta}^{e^{-1}} t^{-1} \left( \log \left( \frac{1}{t} \right) \right)^{j_0-2} \, dt
\]
\[
\leq 2e + \frac{C_{j,1}}{j_0-1} \left( \log \left( \frac{1}{\delta} \right) \right)^{j_0-1}.
\]

Since for \( \delta \leq e^{-1} \) we have that \( \log(1/\delta) \geq 1 \), combining the above with (4.26) and setting \( \delta = \varepsilon \) we establish (4.25) for \( j = j_0 + 1 \). This completes the proof. \( \square \)

Equipped with Proposition 4.5 we now begin the proof of Lemma 4.1.

**Proof of Lemma 4.4.** Recalling (4.1) we note that \( P_0(z) \) is a homogeneous polynomial of degree \( |\mathfrak{d}| \) in the entries of the noise matrix \( \Delta_N \) such that the degree of each entry of \( \Delta_N \) is one. Therefore, to apply Proposition 4.5 we only need to show that there exists \( X, Y \subset [N] \) with \( |X| = |Y| = |\mathfrak{d}| \) such det\( (T_N(z)[X^c;Y^c]) \) is bounded below. The choice of such subsets will depend on the sign of \( \mathfrak{d} \). Hence, the proof is split into two cases.

Considering the case \( \mathfrak{d} > 0 \) we set \( X = [N] \setminus [N - \mathfrak{d}] \) and \( Y = [\mathfrak{d}] \). Recalling Definition 4.1 we find that
\[
T_N(z)[X^c;Y^c] = T_{N-\mathfrak{d}}(z; d_1 - \mathfrak{d}, d_2 + \mathfrak{d}).
\]
We apply Widom’s result on the determinant of finitely banded Toeplitz matrices, in particular [5, Theorem 2.8] to deduce that for any \( z \in \mathbb{C} \setminus \mathcal{N} \), one has
\[
\det(T_N(z)[X^c;Y^c]) = \sum_{X \in \left( [d_1 - \mathfrak{d}] \right)} C_X \cdot a_X^N \prod_{\ell \in \mathcal{I}} \lambda_\ell(z)^N,
\]
for some collection of coefficients \( \{C_X\} \), where recall that \( \mathcal{N} \) is the collection of \( z \)’s such that \( P_{z,0}(\cdot) \) has double roots. Furthermore, the coefficients \( \{C_X\} \) are bounded both below and above, for any \( z \in B_C(0, R) \setminus \mathcal{N} \). As \( z \in \mathcal{S}_\mathfrak{d} \) and \( d_1 - \mathfrak{d} = d_0(z) \), using (4.15) we therefore deduce that there exists some small positive constant \( c_0 > 0 \) so that, for all large \( N \),
\[
(4.27) \quad |\det(T_N(z)[X^c;Y^c])| \geq c_0 \cdot |a_{d_1}|^N \prod_{\ell=1}^{d_0(z)} |\lambda_\ell(z)|^N.
\]
From the definition of \( \Delta_N \) it follows that for the above choices of \( X \) and \( Y \) the determinant of \( \Delta_N[X;Y] \), ignoring the factor \( N^{-\gamma_\mathfrak{d}} \), is a homogeneous polynomial of degree \( \mathfrak{d} \) of independent uniformly bounded random variables with uniformly bounded densities. Therefore, we are in a position to apply Proposition 4.5.

Without loss of generality, assuming that the densities of \( \{(\Delta_N)_{i,j}\}_{i,j=1}^N \) are uniformly bounded by one, we apply Proposition 4.5 for
\[
N^{-\gamma_\mathfrak{d}} \cdot P_\mathfrak{d}(z) \quad \frac{|a_{d-1}|^N \prod_{\ell=1}^{d_0(z)} |\lambda_\ell(z)|^N}{|a_{d-1}|^N \prod_{\ell=1}^{d_0(z)} |\lambda_\ell(z)|^N}
\]
with \( c_0 = c_0 \) and \( \varepsilon = N^{-c_0/2}c_0 \) to arrive at (4.2) for any \( z \in \mathcal{S}_0 \setminus \mathcal{N} \). As \( \mathcal{N} \) contains at most finitely many points this proves the lemma when \( \mathfrak{d} > 0 \).

Turning to prove the same for \( \mathfrak{d} < 0 \), we reverse the roles of \( X \) and \( Y \). That is, we now set \( X = [\mathfrak{d}] \) and \( Y = [N] \setminus [N + \mathfrak{d}] \) and follow the same steps as above.

For \( \mathfrak{d} = 0 \) the proof is straightforward. From its definition we have \( P_0(z) = \det(T_N(z)) \). Upon setting \( X = Y = \emptyset \) in (4.27) the result is immediate. Now the proof of the lemma is complete. \( \square \)

We end this section with the proof of Lemma 4.4. Its proof is very similar to that of Lemma 4.2. Hence, only an outline is provided.
Definition 5.1. The Stieltjes transform of a probability measure \( \mu \) on \( \mathbb{R} \) is defined as

\[
G_\mu(\xi) := \int \frac{1}{\xi - y} \mu(dy), \quad \xi \in \mathbb{C} \setminus \mathbb{R}.
\]

To obtain a bound on the probability of any interval under \( \mu \) from that of \( G_\mu(\cdot) \) we use the following two inequalities. These are a consequence of [9, Eqns. (6)-(8)]: for any \( \tau, \varrho > 0 \) such that \( b - a > \varrho \)

\[
\mu([a, b]) \leq \int_{a - \varrho}^{b + \varrho} |3G_\mu(x + i\tau)|dx + \frac{\tau}{\varrho},
\]

and

\[
\mu([a, b]) \geq \int_{a + \varrho}^{b - \varrho} |3G_\mu(x + i\tau)|dx - \frac{\tau}{\varrho}.
\]

Now to find a difference of the Stieltjes transforms of \( \nu_{\Lambda_N + \Delta_N} \) and \( \nu_{\Lambda_N + E_N} \) we also need the symmetrized form of the Stieltjes transform, as follows. For a \( N \times N \) matrix \( C_N \), define

\[
\overline{C}_N := \begin{bmatrix} 0 & C_N^* \\ C_N & 0 \end{bmatrix}
\]

and the Stieltjes transform

\[
G_{C_N}(\xi) := \frac{1}{2N} \text{tr} \left( \xi - \overline{C}_N \right)^{-1}, \quad \xi \in \mathbb{C} \setminus \mathbb{R}.
\]

\( G_{C_N}(\cdot) \) is the Stieltjes transform of the symmetrized version of the empirical measure of the singular values of \( \overline{C}_N \). Equipped with the above notation we have the following lemma.

Lemma 5.1. For \( C_N \) and \( D_N \) any \( N \times N \) matrices,

\[
|G_{C_N}(\xi) - G_{D_N}(\xi)| \leq \frac{1}{\sqrt{N}} \frac{\|C_N - D_N\|_{\text{HS}}}{(\Im(\xi))^2},
\]

where \( \| \cdot \|_{\text{HS}} \) denotes the Hilbert-Schmidt norm.

Proof. Using the resolvent identity we have that

\[
G_{C_N}(\xi) - G_{D_N}(\xi) = \frac{1}{2N} \text{tr} \left[ \left( \xi - \overline{D}_N \right)^{-1} \cdot \left( \xi - \overline{C}_N \right)^{-1} \cdot (\overline{C}_N - \overline{D}_N) \right].
\]

Recall the following version of the Cauchy-Schwarz inequality: for any two \( (2N) \times (2N) \) matrices \( A_N \) and \( B_N \)

\[
\frac{1}{2N} |\text{tr}(A_NB_N)| \leq \frac{1}{2N} \sqrt{\text{tr}(A_N^2)} \cdot \sqrt{\text{tr}(B_N^2)} \leq \|A_N\| \cdot \|B_N\|_{\text{HS}}.
\]
Since for any Hermitian matrix $H_N$ one has $\| (\xi - H_N)^{-1} \| \leq 1/|\Im(\xi)|$, the claim follows from (5.5) upon using (5.6) with $A_N = \left( \xi - \hat{D}_N \right)^{-1} \cdot \left( \xi - \hat{C}_N \right)^{-1}$ and $B_N = \hat{C}_N - \hat{D}_N$. \hfill $\square$

As a last preliminary step, we need the following easy lemma.

**Lemma 5.2.** For any probability measure $\mu$,  
\begin{equation}
(5.7) \quad \lim_{\varepsilon \downarrow 0} \int \log |x-z| 1_{\{|x-z| \leq \varepsilon\}} \, d\mu(x) = 0
\end{equation}
for Lebesgue almost every $z \in \mathbb{C}$.  

**Proof.** For $\varepsilon < 1$, set $F(z, \varepsilon) := \int \log(1/|x-z|) 1_{\{|x-z| \leq \varepsilon\}} \, d\mu(x)$. Fix $z_0 \in C$. Note that, by Fubini's theorem,  
\[ \int_{B_\varepsilon(z_0,1)} F(z, \varepsilon) \, dz < \pi \varepsilon^2 \log(1/\varepsilon) \to_{\varepsilon \to 0} 0. \]
In particular, for any $\delta > 0$, with $A_\varepsilon(\delta) := \{ z \in B_\varepsilon(z_0,1) : F(z, \varepsilon) > \delta \}$, we obtain that  
\[ \text{Leb}(A_\varepsilon(\delta)) \to_{\varepsilon \to 0} 0. \]
In particular, $\text{Leb}(\cap_{\varepsilon < 1} A_\varepsilon(\delta)) = 0$. Using the monotonicity of $F(z, \varepsilon)$, we conclude that for Lebesgue almost every $z \in B_\varepsilon(z_0,1)$, $\lim \sup_{\varepsilon \to 0} F(z, \varepsilon) \leq \delta$. Taking a sequence $\delta_n \to 0$ gives (5.7), first for Lebesgue almost every $z \in B_\varepsilon(z_0,1)$, and then for almost every $z$. \hfill $\square$

We are now ready to prove Theorem 1.8.

**Proof of Theorem 1.8.** To establish (1.3) we first claim that $\nu_{A_N + N^{-\gamma} E_N}^z \Rightarrow \mu_z$, in probability, for Lebesgue a.e. $z \in B_C(0,R_0)$, where $\mu_z$ is the law of $|X-z|$ and $X \sim \mu$. The argument is similar to that employed in the proof of Theorem 1.2. Write $B_N := A_N + N^{-\gamma} E_N$ and $B_N(z) := B_N - z \text{Id}_N$. We have that $\nu_{A_N}^z \Rightarrow \mu_z$ by assumption (b), while Assumption 1.1(i) and Markov’s inequality imply that  
\begin{equation}
(5.8) \quad \mathbb{P} \left( \| E_N \|_{\text{HS}} \geq N^{1 + \frac{\gamma - \frac{1}{2}}{2}} \right) \leq N^{-\gamma - \frac{1}{2}} \frac{\mathbb{E} \| E_N \|_{\text{HS}}^2}{N^2} = o(1),
\end{equation}
for any $\gamma > 1/2$. On the other hand, by the Hoffman-Wielandt inequality, see [1, Lemma 2.1.19], the map $D_N \mapsto L_N^D$, viewed as a map from the space of $N \times N$ Hermitian matrices equipped with the normalized Hilbert-Schmidt norm $N^{-1/2} : \| \cdot \|_{\text{HS}}$ to the space of probability measures equipped with the weak topology, is continuous. Note that for any matrix, the singular values of $A$ are the same as the modulus of the eigenvalues of the matrix  
\[ \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}, \]
up to double the multiplicity for each singular value. In particular,  
\[ N^{-1/2} \| (B_N(z)B_N(z^*)^{1/2} - (A_N(z)A_N(z^*)^{1/2})\|_{\text{HS}} \leq CN^{-\gamma + 1/2} \| E_N \|_{\text{HS}} \to_{N \to \infty} 0, \]
in probability, by (5.8). We conclude from that and the above mentioned continuity of the empirical measure in the (normalized) Hilbert-Schmidt norm that  
\begin{equation}
(5.9) \quad \nu_{B_N}^z \Rightarrow \mu_z, \quad \text{in probability},
\end{equation}
as claimed.

To complete the proof we need to extend the convergence of $\nu_{B_N}^z$ to the convergence of the integral of $\log(\cdot)$ against this measure. To this end, using (5.8) again and the fact that the operator norm of $A_N(z)$ is bounded, we see that there exists a compact set $K \subset \mathbb{R}$ such that for any $z \in B_C(0,R_0)$  
\[ \mathbb{P}(\nu_{B_N}(K^c) > 0) = o(1). \]
Hence, for any $\varepsilon > 0$,  
\[ \int_{\varepsilon}^\infty \log(x) \, d\nu_{B_N}^z(x) \to \int_{\varepsilon}^\infty \log(x) \, d\mu_z(x) = \int \log(|x-z|) 1_{\{|x-z| > \varepsilon\}} \, d\mu(x), \quad \text{in probability}, \]
for Lebesgue a.e. $z \in B_C(0, R_0)$. Note that

$$L_{N, z}^B(z) = \frac{1}{N} \log |\det(B_N(z))| = \int \log(x) d\nu_{B_N}(x).$$

Thus, (5.7) together with (5.9) imply that it only remains to show that given any $\delta > 0$, there exists $\varepsilon_0(\delta)$ such that for any $\varepsilon \leq \varepsilon_0(\delta)$

$$\limsup_{N \to \infty} \mathbb{P} \left( \int_0^x \log(x) d\nu_{B_N}(x) \right) \geq C_0 \delta = 0,$$

for Lebesgue a.e. $z \in B_C(0, R_0)$ and some large constant $C_0$. To prove this, we first show that an analogue of (5.10) holds for the empirical measure of the singular values of $A_N + \Delta_N$.

Turning to do this task, using (1.1) and arguing similarly to the steps leading to (5.9), we obtain that

$$\limsup_{N \to \infty} \mathbb{P} \left( \int_0^x \log(x) d\nu_{A_N + \Delta_N}(x) \right) \geq \delta = 0.$$

Having shown (5.12) we now proceed to the proof of (5.10). Using Assumption 1.1(ii) we see that there exists $\varepsilon_0(\delta)$ such that for all $\varepsilon \leq 2\varepsilon_0(\delta)$,

$$\limsup_{N \to \infty} \mathbb{P} \left( \int_0^x \log(x) d\nu_{A_N + \Delta_N}(x) \right) \geq \delta = 0.$$

This implies that the integral of $\log(\cdot)$ over $(N^{-\kappa}, \varepsilon)$ we apply (5.4) to deduce that

$$|G_{A_N + \Delta_N}(x + i\tau) - G_{B_N}(x + i\tau)| \leq \frac{N^{-\gamma} \|E_N\|_{HS} + \|\Delta_N\|_{HS}}{\sqrt{N}\tau^2} \leq N^{-\delta'/2},$$

on the event

$$\Omega_N := \left\{ \|E_N\|_{HS} \leq N^{1+\frac{\gamma}{2}} \right\} \cap \{\|\Delta_N\| \leq N^{-\gamma_0}\},$$

where $\tau = N^{-\delta'/4}$ and $\delta' = \min\{\frac{1}{2}(\gamma - \frac{1}{2}), \gamma_0\}$.

Let $\nu_{A_N + \Delta_N}$ and $\nu_{B_N}$ denote the symmetrized versions of the probability measures $\nu_{A_N + \Delta_N}$ and $\nu_{B_N}$, respectively. Setting $\varrho = N^{-\delta'/8}$, $\kappa = \delta'/16$, and using (5.1)-(5.2) and (5.14) in the second inequality, we have that

$$- \int_{N^{-\kappa}}^{N^{-\kappa}} \log(x) d\nu_{B_N}(x) \leq \kappa \log N \cdot \nu_{B_N}([N^{-\kappa}, N^{-\kappa}])$$

$$\leq \kappa \log N \left( 2N^{-\delta'/8} + 2 \nu_{A_N + \Delta_N}([N^{-\kappa} - 2\varrho, N^{-\kappa} + 2\varrho]) \right)$$

$$\leq \kappa \log N \left( 3N^{-\delta'/8} + 4 \nu_{A_N + \Delta_N}([0, 2N^{-\kappa}]) \right)$$

$$\leq 3 \kappa \log N \cdot N^{-\delta'/8} - \frac{8 \kappa \rho}{\kappa} \int_0^{2N^{-\kappa}} \log(x) d\nu_{A_N + \Delta_N}(x),$$

on $\Omega_N$, for all large $N$, where in the third inequality we used the symmetry of $\nu_{A_N + \Delta_N}$ and $\varrho = o(N^{-\kappa})$. 
It remains to bound the integral of $\log(\cdot)$ over $(N^{-\kappa}, \varepsilon)$. Toward this, using integration by parts we note that, for $0 \leq a_1 < a_2 < 1$ and any probability measure $\mu$ on $\mathbb{R}$,

\begin{equation}
(5.16) \quad - \int_{a_1}^{a_2} \log(x) d\mu(x) = - \log(a_2) \mu([a_1, a_2]) + \int_{a_1}^{a_2} \frac{\mu([a_1, t])}{t} dt.
\end{equation}

Arguing as in (5.15) we obtain

\begin{equation}
(5.17) \quad \int_{N^{-\kappa}}^{\varepsilon} \frac{\nu_B N([N^{-\kappa}, t])}{t} dt \leq 3N^{-\delta'/8} \int_{N^{-\kappa}}^{\varepsilon} \frac{1}{t} dt + \int_{N^{-\kappa}}^{\varepsilon} \frac{\nu_A N + \nu_N([N^{-\kappa}/2, t + N^{-\kappa}])}{t} dt \leq 3\kappa N^{-\delta'/8} \cdot \log N + \int_{N^{-\kappa}/2}^{2\varepsilon} \frac{\nu_A N + \nu_N([N^{-\kappa}/2])}{t} dt,
\end{equation}

where in the last step we have used the fact that $t + N^{-\kappa} \leq 2t$ for any $t \geq N^{-\kappa}$, and a change of variables. Similar reasoning yields that

\begin{equation}
(5.18) \quad - \log(\varepsilon) \nu_B N([N^{-\kappa}, \varepsilon]) \leq - \log(\varepsilon) \left(3N^{-\delta'/8} + \nu_A N + \nu_N([N^{-\kappa}/2, 2\varepsilon])\right).
\end{equation}

Thus combining (5.15), and (5.17)-(5.18), and using (5.16) we deduce that for $\varepsilon \leq \varepsilon_0(\delta)$ sufficiently small and all large $N$,

\begin{equation}
(5.19) \quad \int_{N^{-\kappa}}^{\varepsilon} \log(x) d\nu_B N(x) \leq C_0 \left[ \log N \cdot N^{-\delta'/8} - \int_0^{2\varepsilon} \log(x) d\nu_A N + \nu_N(dx) \right],
\end{equation}

on the event $\Omega_N$, where $C_0$ is some large constant. Finally, (5.8) and (1.1) imply that $\mathbb{P}(\Omega_N^c) = o(1)$. Therefore, combining (5.12) and (5.13), the claim in (5.10) now follows. This completes the proof of the theorem.

\section*{Appendix A. Some algebraic facts}

In this section we collect a couple of standard matrix results which have been used in the proofs appearing in Section 4.

The first result shows that the determinant of the sum of the two matrices can be expressed as a linear combination of products of the determinants of appropriate sub-matrices. The proof follows from the definition of the determinant, see e.g. [10]. We adopt the convention that the determinant of the matrix of size zero is one. For an $N \times N$ matrix $A$, and $X, Y \subseteq [N]$ we write $A[X; Y]$ for the sub-matrix of $A$ which consists of the rows in $X$ and the columns in $Y$.

\textbf{Lemma A.1.} For any $N \times N$ matrices $A$ and $B$ we have

\begin{equation}
(A.1) \quad \det(A + B) = \sum_{X, Y \subseteq [N] \setminus \{X = |Y|\}} (-1)^{\text{sgn}(\sigma_X) \text{sgn}(\sigma_Y)} \det(A[X^c; Y^c]) \det(B[X; Y]),
\end{equation}

where $X^c := [N] \setminus X$, $Y^c := [N] \setminus Y$ and $\sigma_Z$ for $Z \subseteq \{X, Y\}$ is the permutation on $[N]$ which places all the elements of $Z$ before all the elements of $Z^c$, but preserves the order of elements within the two sets.

The next lemma evaluates the determinant of any sub-matrix of a bidiagonal matrix.

\textbf{Lemma A.2} ([8, Lemma 2.2]). Let $A_N$ be an upper bi-diagonal matrix and $X, Y \subseteq [N]$ such that $|X| = |Y|$. Then $\det(A_N[X; Y])$ equals the product of the diagonal entries of $A_N[X; Y]$.

The next lemma, which follows readily from Lemma A.2, evaluates the determinant of any sub-matrix of a bidiagonal Toeplitz matrix.

\textbf{Lemma A.3} ([8, Lemma 2.3]). Let $A_N = J_N + \mathfrak{j} \text{Id}_N$, $\mathfrak{j} \in \mathbb{C}$, $X = \{x_1 < x_2 < \cdots < x_k\} \subseteq [N]$, and $Y = \{y_1 < y_2 < \cdots < y_k\} \subseteq [N]$. Then, with $y_{k+1} = \infty$,

\begin{equation}
\det(A_N[X^c; Y^c]) = \mathfrak{j}^{y_1 - 1} \cdot \prod_{i=2}^{k} \mathfrak{j}^{y_i - x_{i-1} - 1} \cdot \mathfrak{j}^{N-x_k} 1 \{y_i \leq x_i < y_{i+1}, i \in [k]\}.
\end{equation}
References

[1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. An introduction to random matrices. No. 118. Cambridge University Press, 2010.

[2] A. Basak, E. Paquette, and O. Zeitouni. Regularization of non-normal matrices by Gaussian noise - the banded Toeplitz and twisted Toeplitz cases. Forum of Mathematics, Sigma, 7, E3, 2019.

[3] A. Basak and O. Zeitouni. Outliers of random perturbations of Toeplitz matrices with finite symbols. ArXiv preprint, arXiv:1905.10244, 2019.

[4] C. Bordenave and D. Chafaï. Around the circular law. Probability Surveys, 9, 1–89, 2012.

[5] A. Böttcher and S. M. Grudsky. Spectral Properties of Banded Toeplitz Matrices. Vol. 96, Siam, 2005.

[6] N. Cook. Lower bounds for the smallest singular value of structured random matrices. The Annals of probability, 46(6), 3442–3500, 2018.

[7] R. B. Davies and M. Hager. Perturbations of Jordan matrices. Journal of Approximation Theory 156, 82–94, 2009.

[8] O. N. Feldheim, E. Paquette, and O. Zeitouni. Regularization of non-normal matrices by Gaussian noise. International Mathematics Research Notices 18, 8724–8751, 2015.

[9] A. Guionnet, P. M. Wood, and O. Zeitouni. Convergence of the spectral measure of non-normal matrices. Proceedings of the American Mathematical Society 142, 667–679, 2014.

[10] M. Marcus. Determinant of sums. The college mathematical journal 21, 130–135, 1990.

[11] P. Schmidt and F. Spitzer. The Toeplitz matrices of an arbitrary Laurent polynomial. Mathematica Scandinavica, 8, 15–38, 1960.

[12] J. Sjöstrand and M. Vogel. Toeplitz band matrices with small random perturbations. ArXiv preprint, arXiv:1901.08982, 2019.

[13] J. Sjöstrand and M. Vogel. General Toeplitz matrices subject to Gaussian perturbations. ArXiv preprint, arXiv:1905.10265, 2019.

[14] T. Tao. Topics in random matrix theory. Vol. 132. Providence, RI: American Mathematical Society, 2012.

[15] T. Tao and V. Vu. Random matrices: the circular law. Communications in Contemporary Mathematics, 10(2), 261–307, 2008.

[16] L. Reichel and L. N. Trefethen. Eigenvalues and pseudo–eigenvalues of Toeplitz matrices. Linear algebra and its applications 162, 153–185, 1992.

[17] L. N. Trefethen and M. Embree. Spectra and pseudospectra: the behavior of nonnormal matrices and operators. Princeton University Press, 2005.

[18] M. Rudelson and R. Vershynin. Invertibility of random matrices: unitary and orthogonal perturbations. Journal of the American Mathematical Society, 27(2), 293–338, 2014.

[19] H. Widom. On the eigenvalues of certain Hermitian operators. Transactions of the American Mathematical Society, 88, 491–522, 1958.

International Center for Theoretical Sciences
Tata Institute of Fundamental Research
Bangalore 560089, India

and

Department of Mathematics, Weizmann Institute of Science
POB 26, Rehovot 76100, Israel

‡Department of Mathematics, The Ohio State University
Tower 100, 231 W 18th Ave, Columbus, Ohio 43210, USA

§Department of Mathematics, Weizmann Institute of Science
POB 26, Rehovot 76100, Israel

and

Courant Institute, New York University
251 Mercer St, New York, NY 10012, USA