The Annals of Probability
2010, Vol. 38, No. 5, 2023–2065
DOI: 10.1214/10-AOP534

RANDOM MATRICES: UNIVERSALITY OF ESDs AND THE CIRCULAR LAW

BY TERENCE TAO[1] AND VAN VU[2]

(WITH AN APPENDIX BY MANJUNATH KRISHNAPUR)

University of California, Los Angeles, Rutgers University and Indian Institute of Science

Given an $n \times n$ complex matrix A, let

$$\mu_A(x, y) := \frac{1}{n}|\{1 \le i \le n, \operatorname{Re}\lambda_i \le x, \operatorname{Im}\lambda_i \le y\}|$$

be the empirical spectral distribution (ESD) of its eigenvalues $\lambda_i \in \mathbb{C}$, $i = 1, \ldots, n$.

We consider the limiting distribution (both in probability and in the almost sure convergence sense) of the normalized ESD $\mu_{1/\sqrt{n}A_n}$ of a random matrix $A_n = (a_{ij})_{1 \le i,j \le n}$, where the random variables $a_{ij} - \mathbf{E}(a_{ij})$ are i.i.d. copies of a fixed random variable x with unit variance. We prove a *universality principle* for such ensembles, namely, that the limit distribution in question is *independent* of the actual choice of x. In particular, in order to compute this distribution, one can assume that x is real or complex Gaussian. As a related result, we show how laws for this ESD follow from laws for the *singular* value distribution of $\frac{1}{\sqrt{n}}A_n - zI$ for complex z.

As a corollary, we establish the circular law conjecture (both almost surely and in probability), which asserts that $\mu_{1/\sqrt{n}A_n}$ converges to the uniform measure on the unit disc when the a_{ij} have zero mean.

1. Introduction.

1.1. *Empirical spectral distributions.* This paper is concerned with the convergence of empirical spectral distributions of random matrices, both in the sense of convergence in probability and in the almost sure sense.

DEFINITION 1.1 (Modes of convergence). For each n, let F_n be a random variable taking values in some Hausdorff topological space X and let F be another element of X.

Received February 2009; revised December 2009.

[1]Supported by a grant from the MacArthur Foundation and by NSF Grant DMS-06-49473.

[2]Supported by NSF Grant DMS-09-01216 and DOD Grant AFOSAR-FA-9550-09-1-0167.

AMS 2000 subject classifications. Primary 15A52, 60F17; secondary 60F15.

Key words and phrases. Circular law, eigenvalues, random matrices, universality.

- We say that F_n *converges in probability* to F if, for every neighbourhood V of F, we have $\lim_{n\to\infty} \mathbf{P}(F_n \in V) = 1$.
- We say that F_n *converges almost surely* to F if we have $\mathbf{P}(\lim_{n\to\infty} F_n = F) = 1$.

Similarly, if X_n is a scalar random variable, we say that X_n is *bounded in probability* if we have

$$\lim_{C\to\infty} \liminf_{n\to\infty} \mathbf{P}(|X_n| \le C) = 1$$

and *almost surely bounded* if we have

$$\mathbf{P}\Bigl(\limsup_{n} |X_n| < \infty\Bigr) = 1.$$

Let $M_n(\mathbb{C})$ denote the set of $n \times n$ complex matrices. For $A \in M_n(\mathbb{C})$, we let

$$\mu_A(s,t) := \frac{1}{n}|\{1 \le i \le n, \operatorname{Re}\lambda_i \le s, \operatorname{Im}\lambda_i \le t\}|$$

be the *empirical spectral distribution* (ESD) of its eigenvalues $\lambda_i \in \mathbb{C}$, $i = 1, \dots, n$. This is a discrete probability measure on $\mathbb{C}$.

Now, suppose that $A_n \in M_n(\mathbb{C})$ is a random matrix ensemble [i.e., a probability distribution on $M_n(\mathbb{C})$] and let μ_∞ be a probability measure on $\mathbb{C}$. We give the space of probability measures on $\mathbb{C}$ the usual *vague topology*. Thus, a sequence of deterministic measures μ_n converges to μ if $\int_{\mathbb{C}} f\, d\mu_n$ converges to $\int_{\mathbb{C}} f\, d\mu$ for every *test function* (i.e., continuous and compactly supported function) $f : \mathbb{C} \to \mathbb{R}$. Thus, by Definition 1.1, we see that $\mu_{1/\sqrt{n}A_n}$ converge in probability to μ_∞ if, for every continuous and compactly supported function $f : \mathbb{C} \to \mathbb{R}$, the expression

$$\int_{\mathbb{C}} f(z)\, d\mu_{1/\sqrt{n}A_n}(z) - \int_{\mathbb{C}} f(z)\, d\mu_\infty \tag{1.1}$$

converges to zero in probability, thus

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\int_{\mathbb{C}} f(z)\, d\mu_{1/\sqrt{n}A_n}(z) - \int_{\mathbb{C}} f(z)\, d\mu_\infty\right| \ge \varepsilon\right) = 0$$

for every $\varepsilon > 0$. Similarly, $\mu_{1/\sqrt{n}A_n}$ converges almost surely to μ_∞ if, with probability 1, the expression (1.1) converges to zero for all $f : \mathbb{C} \to \mathbb{R}$.

REMARK 1.2. In practice, our matrices A_n will have bounded entries on the average, which suggests (by the Weyl comparision inequality—see Lemma A.2) that their eigenvalues should be of size about $O(\sqrt{n})$. Thus, the normalization by $\frac{1}{\sqrt{n}}$ is natural.

1.2. *Universality.* A fundamental problem in the theory of random matrices is to determine the limiting distribution of the ESD of a random matrix ensemble (either in probability or in the almost sure sense) as the size of the random matrix tends to infinity.

The situation with this problem, thus far, is that the analysis depends very much on which ensemble one is dealing with. In some cases, such as when the entries have a Gaussian distribution, powerful group theoretic structure [e.g., invariance under the orthogonal group $O(n)$ or unitary group $U(n)$] plays an essential role, as one can use it to derive an explicit formula for the joint distribution of the eigenvalues. The limiting distribution can then be computed directly from this formula. In the majority of cases, however, there is little symmetry and such a formula is not available. Consequently, the problem becomes much harder and its analysis typically requires tools from various areas of mathematics.

On the other hand, there is a well-known intuition behind this problem (and many others concerning random matrices), the *universality* phenomenon, that asserts that the limiting distribution should not depend on the particular distribution of the entries. This phenomenon motivates many theorems and conjectures in this area. In the following, we mention two famous examples: Wigner's semicircle law and the circular law conjecture.

Wigner's semicircle law. In the 1950s, motivated by numerical experiments, Wigner [27] proved that the ESD of an $n \times n$ Hermitian matrix with (upper diagonal) entries being i.i.d. Gaussian random variables converge to the semicircle law F whose density is given by

$$\rho(x) = \begin{cases} \dfrac{1}{2\pi}\sqrt{4 - x^2}, & |x| \le 2, \\ 0, & |x| > 2. \end{cases}$$

Wigner's result (which holds for both modes of convergence) was later extended to many other ensembles. The most general form only requires the mean and variance of the entries [2, 15].

THEOREM 1.3. *Let A_n be the $n \times n$ Hermitian random matrix whose upper diagonal entries are i.i.d. complex random variables with mean* 0 *and variance* 1. *The ESD of $\frac{1}{\sqrt{n}} A_n$ then converges* (*both in probability and in the almost sure sense*) *to the semicircle distribution.*

Circular law conjecture. The well-known circular law conjecture deals with non-Hermitian matrices.

CONJECTURE 1.4. *Let A_n be the $n \times n$ random matrix whose entries are i.i.d. complex random variables with mean* 0 *and variance* 1. *The ESD of $\frac{1}{\sqrt{n}} A_n$ then converges* (*both in probability and in the almost sure sense*) *to the uniform distribution on the unit disk.*

Similarly to Wigner's law, this conjecture was posed, based on numerical evidence, in the 1950s. The case when the entries have a complex Gaussian distribution was verified by Mehta [13] in 1967, using Ginibre's formula for the joint density function of the eigenvalues of A_n (see, e.g., [2], Chapter 10):

$$p(\lambda_1, \dots, \lambda_n) = c_n \prod_{i<j} |\lambda_i - \lambda_j|^2 \exp\left(-n \sum_{i=1}^{n} |\lambda_i|^2\right). \tag{1.2}$$

Another case where such a formula is available is when the entries have a real Gaussian distribution and, for this case, the conjecture was confirmed by Edelman [6]. For the general case when there is no formula, the problem appears much harder. Important partial results were obtained by Girko [7, 8], Bai [1, 2] and, more recently, Götze and Tikhomirov [9, 10], Pan and Zhou [14] and the present authors [25]. These results establish the conjecture (in almost sure or in probability forms) under additional assumptions on the distribution x. The strongest result in the previous literature is from [10, 25], in which the almost sure and in-probability forms of the conjecture were shown under the extra assumption that the entries have finite $(2+\epsilon)$th moment for any positive constant ϵ. An attempt to remove this extra ϵ (and thus to prove Conjecture 1.4 in full generality) was a motivation for this paper.

A demonstration of the circular law for the Bernoulli and Gaussian cases appears in Figure 1.

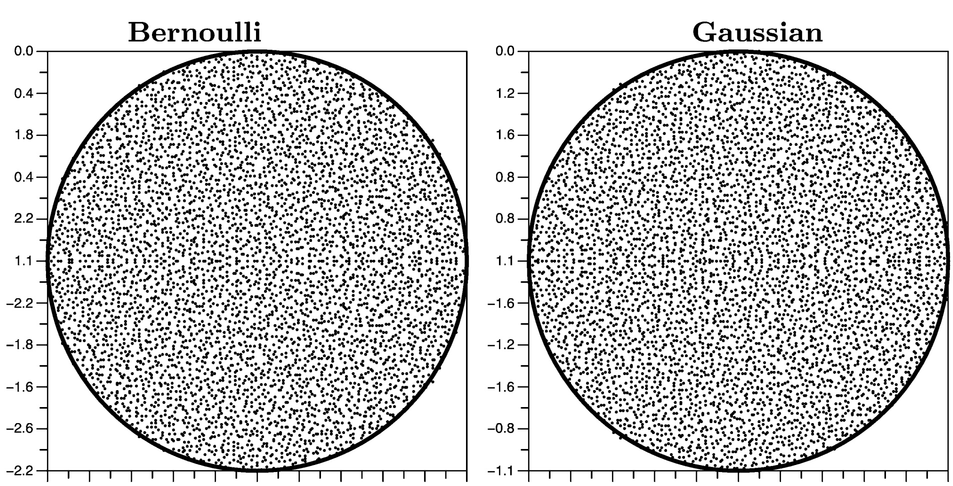

FIG. 1. *Eigenvalue plots of two randomly generated* 5000×5000 *matrices. On the left, each entry was an i.i.d. Bernoulli random variable, taking the values* $+1$ *and* -1 *each with probability* $1/2$. *On the right, each entry was an i.i.d. Gaussian normal random variable, with probability density function* $\frac{1}{\sqrt{2*\pi}} \exp(-x^2/2)$. [*These two distributions were shifted by adding the identity matrix, thus the circles are centered at* $(1, 0)$, *rather than at the origin.*]

In both the semicircular law and the circular law, we observe that only the mean and variance of the entries play a role in the limiting distribution. This is, in fact, a common situation for many other conjectures in random matrix theory, such as Dyson's conjecture [13], Chapter 1, and this phenomenon is sometimes referred to as *universality* in the literature.

In this paper, we rigorously prove the universality phenomenon for the ESD of random matrices. More precisely, we show that the limiting distribution of the ESD of a random matrix ensemble A_n depends only on the mean and variance of its entries, under a mild size condition on the mean $\mathbf{E}A_n$ and under the assumption that the matrix $A_n - \mathbf{E}A_n$ has i.i.d. entries.

For any matrix A, we define the *Hilbert–Schmidt norm* $\|A\|_2$ by the formula $\|A\| := \operatorname{trace}(AA^*)^{1/2} = \operatorname{trace}(A^*A)^{1/2}$.

THEOREM 1.5 (Universality principle). *Let x and y be complex random variables with zero mean and unit variance. Let $X_n = (x_{ij})_{1\le i,j\le n}$ and $Y_n := (y_{ij})_{1\le i,j\le n}$ be $n \times n$ random matrices whose entries x_{ij}, y_{ij} are i.i.d. copies of x and y, respectively. For each n, let M_n be a deterministic $n \times n$ matrix satisfying*

$$\sup_n \frac{1}{n^2}\|M_n\|_2^2 < \infty. \tag{1.3}$$

Let $A_n := M_n + X_n$ and $B_n := M_n + Y_n$. Then, $\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}$ converges in probability to zero. If, furthermore, we make the additional hypothesis that the ESDs

$$\mu_{(1/\sqrt{n}M_n - zI)(1/\sqrt{n}M_n - zI)^*} \tag{1.4}$$

converge to a limit for almost every z, then $\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}$ converges almost surely to zero.

REMARK 1.6. The theorem still holds if we restrict the size of the matrices to an infinite subsequence $n_1 < n_2 < \cdots$ of positive integers. This freedom to pass to a subsequence is useful for technical reasons involving compactness arguments.

The condition (1.3) has the following useful consequence, which we shall use repeatedly.

LEMMA 1.7 (Tightness of ESDs). *Let M_n and A_n be as in Theorem 1.5. Then, the quantities $\frac{1}{n^2}\|A_n\|_2^2$ and $\int_{\mathbb{C}} |z|^2 \, d\mu_{1/\sqrt{n}A_n}(z)$ are almost surely bounded (and hence also bounded in probability).*

PROOF. By the Weyl comparison inequality (Lemma A.2), it suffices to show that $\frac{1}{n^2}\|A_n\|_2^2$ is almost surely bounded. By (1.3) and the triangle inequality, it suffices to show that $\frac{1}{n^2}\|X_n\|_2^2$ is almost surely bounded. However, this follows from the finite second moment of x and the strong law of large numbers. □

As an immediate corollary of Theorem 1.5, we have the following result.

COROLLARY 1.8 (Universality principle). *Let x, y be complex random variables with zero mean and unit variance. Let X_n and Y_n be $n \times n$ random matrices whose entries are i.i.d. copies of x and y, respectively. For each n, let M_n be a deterministic $n \times n$ matrix satisfying* (1.3). *Let $A_n := M_n + X_n$ and $B_n := M_n + Y_n$. Then, if $\mu_{1/\sqrt{n}B_n}$ converges in probability to a limiting measure μ, then $\mu_{1/\sqrt{n}A_n}$ also converges in probability to μ. If, furthermore, we make the additional hypothesis that the ESDs* (1.4) *converge to a limit for almost every z, then we can replace "in probability" by "almost surely" in the previous sentence.*

A demonstration of this corollary appears in Figure 2.

REMARK 1.9. One consequence of Corollary 1.8 [in the case where (1.4) converges to a limit] is that the ESD $\mu_{1/\sqrt{n}A_n}$ behaves asymptotically deterministically,[3] in the sense that there exists a deterministic measure μ_n for each n such that $\mu_{1/\sqrt{n}A_n} - \mu_n$ converges almost surely to zero. Indeed, one can simply take μ_n to be an instance of $\mu_{1/\sqrt{n}B_n}$, where the B_n are selected independently of the A_n, and the claim will hold almost surely. The question remains as to whether μ_n itself

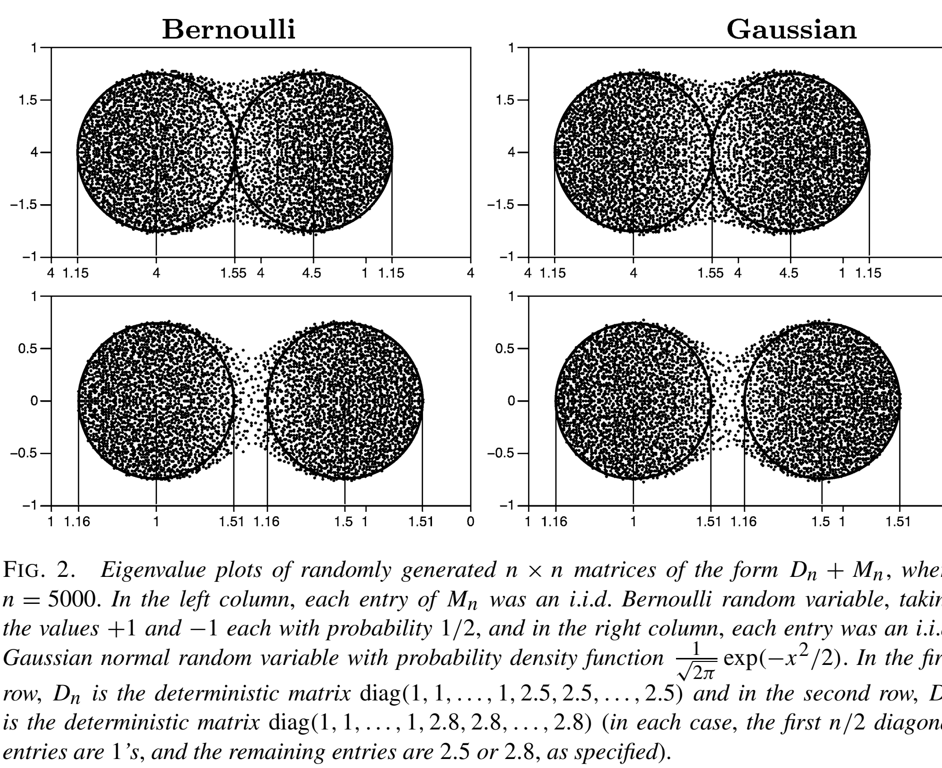

FIG. 2. *Eigenvalue plots of randomly generated $n \times n$ matrices of the form $D_n + M_n$, where $n = 5000$. In the left column, each entry of M_n was an i.i.d. Bernoulli random variable, taking the values $+1$ and -1 each with probability $1/2$, and in the right column, each entry was an i.i.d. Gaussian normal random variable with probability density function $\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. In the first row, D_n is the deterministic matrix $\mathrm{diag}(1, 1, \ldots, 1, 2.5, 2.5, \ldots, 2.5)$ and in the second row, D_n is the deterministic matrix $\mathrm{diag}(1, 1, \ldots, 1, 2.8, 2.8, \ldots, 2.8)$ (in each case, the first $n/2$ diagonal entries are 1's, and the remaining entries are 2.5 or 2.8, as specified).*

[3]The authors thank Oded Schramm for this observation.

converges to some limit as $n \to \infty$; we partially address this issue in Theorem 1.17 below.

1.3. *The circular law conjecture*. Thanks to Corollary 1.8, we can reduce the problem of computing the limiting distribution to the case where the entries are Gaussian[4] (or having any special distribution satisfying the variance bound). In particular, since the circular law is verified for random matrices with complex Gaussian entries (see [13]), it follows that this law (both in probability and in the almost sure sense) holds in full generality. In other words, we have shown the following theorem.

THEOREM 1.10 (Circular law). *Let X_n be the $n \times n$ random matrix whose entries are i.i.d. complex random variables with mean zero and variance one. Then, the ESD of $\frac{1}{\sqrt{n}} X_n$ converges* (*both in probability and in the almost sure sense*) *to the uniform distribution on the unit disk.*

REMARK 1.11. In [25] (see also [10] for an alternate proof for the in-probability sense), this theorem was proven with the extra assumption that the entries have finite $(2+\varepsilon)$th moment for any fixed $\varepsilon > 0$; earlier, related, results appear in [1, 2, 7–9].

Notice that in Theorem 1.10, we set M_n to be the all-zero matrix (for which the boundedness and convergence hypotheses are trivial). In [11], explicit distributions were computed for the case where M_n is an arbitrary diagonal matrix and X_n has i.i.d. Gaussian entries. The formula for the limiting distribution is somewhat technical, but its support is easy to describe: it is exactly the set of $z \in \mathbb{C}$ for which $\int |z-x|^{-2}\, d\mu(x) \geq 1$, where μ is the limiting distribution of the ESD of M_n. (In the case where M_n is all-zero, μ has all of its mass at the origin and so the set of z is the unit disk.)

The proof of Theorem 1.5 actually shows that if M_n and M_n' both obey (1.3) and have the property that the difference between the ESD (1.4) and the counterpart for M_n' converges to zero for almost every z, then Theorem 1.5 holds with $A_n := M_n + X_n$ and $B_n := M_n' + Y_n$ (see Remark B.3).

This has the following interesting consequence. Assume that M_n is a matrix with low rank, say $o(n)$. In this case, it is easy to see that the ESD (1.4) concentrates at $|z|^2$ since the matrix involved here is a self-adjoint low rank perturbation of $|z|^2 I$. Thus, we can replace M_n by the zero matrix and obtain the following.

COROLLARY 1.12 (Circular law for shifted matrices). *Let X_n be the $n \times n$ random matrix whose entries are i.i.d. complex random variables with mean zero*

[4]The idea of establishing a limiting law by first replacing a general random variable with a Gaussian one is sometimes referred to as the "Lindeberg trick" in the literature.

and variance one and let M_n *be a deterministic matrix with rank* $o(n)$ *and obeying* (1.3). *Let* $A_n := M_n + X_n$. *Then, the ESD of* $\frac{1}{\sqrt{n}}A_n$ *converges* (*in either sense*) *to the uniform distribution on the unit disk.*

In particular, this shows that Theorem 1.10 still holds if the entries have (the same) nonzero mean. This extends a result of Chafaï [5], which, in addition, assumed that the entries had finite fourth moment.

1.4. *Extensions.* We can extend Theorem 1.5 in several ways. First, by conditioning, we can obtain a theorem for M_n being a random matrix.

THEOREM 1.13 (Universality from a random base matrix). *Let* x *and* y *be complex random variables with zero mean and unit variance. Let* $X_n = (x_{ij})_{1\le i,j\le n}$ *and* $Y_n = (y_{ij})_{1\le i,j\le n}$ *be* $n\times n$ *random matrices whose entries are i.i.d. copies of* x *and* y, *respectively. For each* n, *let* M_n *be a random* $n\times n$ *matrix, independent of* X_n *or* Y_n, *such that* $\frac{1}{n^2}\|M_n\|_2^2$ *is bounded in probability* (*see Definition* 1.1). *Let* $A_n := M_n + X_n$ *and* $B_n := M_n + Y_n$. *Then,* $\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}$ *converges in probability to zero. If, furthermore, we assume that* $\frac{1}{n^2}\|M_n\|_2^2$ *is almost surely bounded and* (1.4) *converges almost surely to some limit for almost every* z, *then* $\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}$ *converges almost surely to zero.*

We can also address a more general form of random matrices (cf. [8]). Let K_n, L_n be two sequences of matrices. Define $A_n := M_n + K_nX_nL_n$ and $B_n := M_n + K_nY_nL_n$. We can show that under some mild assumptions on M_n, K_n, L_n, Theorem 1.5 still holds.

THEOREM 1.14. *Let* x *and* y *be complex random variables with zero mean and unit variance. Let* X_n *and* Y_n *be* $n\times n$ *random matrices whose entries are i.i.d. copies of* x *and* y, *respectively. Let* M_n, K_n, L_n *be random* $n\times n$ *matrices* (*independent of* X_n, Y_n) *and let* $A_n := M_n + K_nX_nL_n$ *and* $B_n := M_n + K_nY_nL_n$. *Assume that the expressions*

$$\frac{1}{n^2}\|A_n\|_2^2 + \frac{1}{n^2}\|B_n\|_2^2 + \frac{1}{n^2}\|K_n^{-1}M_nL_n^{-1}\|_2^2 + \frac{1}{n}\|K_n^{-1}L_n^{-1}\|_2^2 \tag{1.5}$$

are bounded in probability. If, furthermore, we assume that (1.5) *is almost surely bounded and that for almost every* z, *the ESDs*

$$\mu_{(1/\sqrt{n}K_n^{-1}M_nL_n^{-1} - zK_n^{-1}L_n^{-1})(1/\sqrt{n}K_n^{-1}M_nL_n^{-1} - zK_n^{-1}L_n^{-1})^*} \tag{1.6}$$

converge almost surely to a limit, then $\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}$ *converges almost surely to zero.*

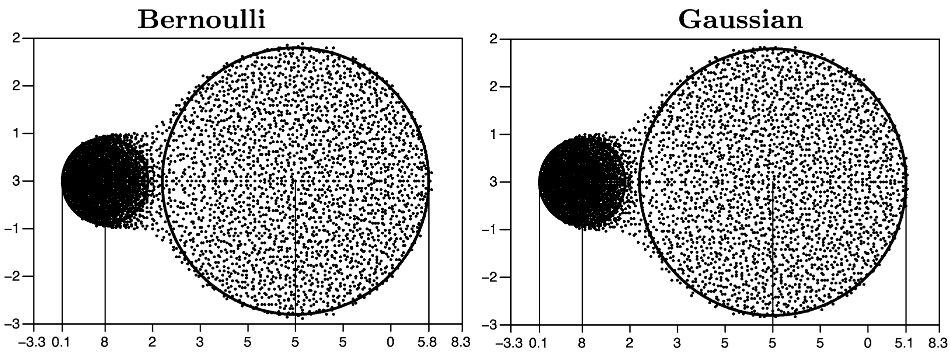

FIG. 3. *Eigenvalue plots of two randomly generated* 5000×5000 *matrices of the form* $A + BM_nB$, *where* A *and* B *are diagonal matrices having* $n/2$ *entries with the value* 1 *followed by* $n/2$ *entries with the value* 5 (*resp*., 2) *for* D (*resp*., X). *On the left, each entry of* M_n *was an i.i.d. Bernoulli random variable, taking the values* $+1$ *and* -1 *each with probability* $1/2$. *On the right, each entry of* M_n *was an i.i.d. Gaussian normal random variable with probability density function* $\frac{1}{\sqrt{2*\pi}} \exp(-x^2/2)$.

Note that Theorem 1.13 is the special case of Theorem 1.14 in which $K_n = L_n = I$. It seems of interest to see whether the hypotheses on (1.5) can be verified for various natural random or deterministic matrices M_n, K_n, L_n, normalized appropriately by a suitable power of n. We do not pursue this matter here.

A demonstration of the above theorem for the Bernoulli and Gaussian cases appears in Figure 3.

The proofs of these extensions are discussed in Section 7.

Another direction for generalization is to consider random matrices whose entries are independent, but not necessarily identically distributed. Most of the tools used in this paper (e.g., the law of large numbers, Talagrand's inequality and the least singular value bound from [25]) extend without difficulty to this setting. Furthermore, Krishnapur pointed out that one can also prove a "universal" version of Theorem B.1. This leads to a generalization in Appendix C (written by Krishnapur).

For similar reasons, one expects to be able to extend the above results to the case where X_n and Y_n are sparse i.i.d. random matrices; for instance, the least singular value bounds from [25] extend to this case and the circular law for sparse i.i.d. matrices is already known in several cases [9, 25]. We, however, will not pursue these matters here.

1.5. *Computing the ESD of a random non-Hermitian matrix via the ESD of a Hermitian one*. Theorem 1.5 provides one useful way to compute the (limiting distribution of) the ESD of a random non-Hermitian matrix, namely, that one can restrict to any particular distribution (such as complex Gaussian) of the entries. The proof of this theorem (with some modification) also provides another way to deal with this problem, namely, that one can reduce the problem of computing the ESD of $\frac{1}{\sqrt{n}}A_n$ to that of $(\frac{1}{\sqrt{n}}A_n - zI)(\frac{1}{\sqrt{n}}A_n - zI)^*$ for fixed $z \in \mathbb{C}$. More precisely, we have the following equivalences.

THEOREM 1.15 (Equivalences for convergence). *Let A_n be as in Theorem 1.5 and let μ be a probability measure on $\mathbb{C}$ with the second moment condition $\int |z|^2\, d\mu(z) < \infty$. The following are then equivalent*:

(i) *the ESD $\mu_{1/\sqrt{n}A_n}$ of $\frac{1}{\sqrt{n}}A_n$ converges in probability to μ*;

(ii) *for almost every complex number z, $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - zI)|$ converges in probability to $\int_{\mathbb{C}} \log|w - z|\, d\mu(w)$*;

(iii) *for almost every complex number z, there exists a sequence $\varepsilon_n > 0$ of positive numbers converging to zero such that $\frac{1}{n}\log\det(((\frac{1}{\sqrt{n}}A_n - zI) + \varepsilon_n I)(\frac{1}{\sqrt{n}}A_n - zI)^* + \varepsilon_n I)$ converges in probability to $2\int_{\mathbb{C}} \log|w - z|\, d\mu(w)$.*

If, furthermore, the ESDs (1.4) converge to a limit for almost every z, then we can replace convergence in probability by almost sure convegence in the above equivalences.

We prove this result in Section 8. As a corollary, we have the following criterion for when $\frac{1}{\sqrt{n}}A_n$ converges to a distribution μ.

COROLLARY 1.16. *Let A_n be as in Theorem 1.5 and let μ be a probability measure on $\mathbb{C}$ with the second moment condition $\int |z|^2\, d\mu(z) < \infty$. Suppose that for almost every complex number z, the ESD of $(\frac{1}{\sqrt{n}}A_n - zI)(\frac{1}{\sqrt{n}}A_n - zI)^*$ converges in probability to a limiting distribution η_z on $[0, +\infty)$ such that the integral $\int_{\mathbb{C}} \log t\, d\eta_z(t)$ is absolutely convergent and equal to $2\int_{\mathbb{C}} \log|w - z|\, d\mu(w)$. Then, the ESD of $\frac{1}{\sqrt{n}}A_n$ converges in probability to μ. If the ESDs (1.4) converge to a limit for almost every z, then we can replace convergence in probability by almost sure convergence in the above implication.*

PROOF. We verify the claim for almost sure convergence only; the proof for convergence in probability is similar and is left as an exercise to the reader.

By Lemma 1.7, we see that for fixed z, $|\frac{1}{n}\operatorname{trace}(\frac{1}{\sqrt{n}}A_n - zI)(\frac{1}{\sqrt{n}}A_n - zI)^*|$ is also almost surely bounded. Taking limits, we conclude that

$$\int_{\mathbb{C}} t\, d\eta_z(t) < \infty.$$

We then see from the dominated convergence theorem that for any $\varepsilon > 0$, $\frac{1}{n}\log\det(((\frac{1}{\sqrt{n}}A_n - zI) + \varepsilon I)(\frac{1}{\sqrt{n}}A_n - zI)^* + \varepsilon I)$ converges almost surely to $\int_{\mathbb{C}} \log(t + \varepsilon)\, d\eta_z(t)$. From this, we obtain hypothesis (iii) of Theorem 1.15 (if ε_n is chosen to decay to zero sufficiently slowly) and the claim follows. $\square$

Since the eigenvalues of $(\frac{1}{\sqrt{n}}A_n - zI)(\frac{1}{\sqrt{n}}A_n - zI)^*$ are the squares of the singular values of $\frac{1}{\sqrt{n}}A_n - zI$, we can also say that Theorem 1.15 reduces the problem of computing the limiting distribution of the eigenvalues of $\frac{1}{\sqrt{n}}A_n$ to that of the singular values of $\frac{1}{\sqrt{n}}A_n - zI$.

The big gain here is that the matrix $(\frac{1}{\sqrt{n}}A_n - zI)(\frac{1}{\sqrt{n}}A_n - zI)^*$ is Hermitian. (Random matrices of this type are often called *sample covariance matrices* in the literature.) This allows one to use standard tools such as truncation, Wigner's moment method and Stieljes transform (see, e.g., the proof of Theorem 1.3 in [2], Chapter 2), or results such as Theorem B.1; techniques from free probability are also very powerful for such problems. These methods cannot be applied to non-Hermitian matrices, for various reasons (see [2], Chapter 10 for a discussion) and their failure has been the main difficulty in attacking problems such as the circular law conjecture.

One can use Corollary 1.16 to give another proof of Theorem 1.10, without relying on explicit formulas such as (1.2). We omit the details.

1.6. *Existence of the limit.* The results in the previous chapters provide two different ways to compute (explicitly) the limiting measure of the ESD of random matrices. In fact, there is a simple compactness argument that guarantees the existence of the limit, assuming, of course, that the deterministic ESDs (1.4) already converge, although the argument does not provide too much information on what the limit actually is. More precisely, we have the following result.

THEOREM 1.17. *Let x be a complex random variable with zero mean and unit variance. Let X_n be the $n \times n$ random matrix whose entries are i.i.d. copies of x. For each n, let M_n be a deterministic $n \times n$ matrix satisfying*

$$\sup_n \frac{1}{n^2}\|M_n\|_2^2 < \infty. \tag{1.7}$$

Assume, furthermore, that the ESD (1.4) *converges for almost every $z \in \mathbb{C}$. Then, the ESD of $\frac{1}{\sqrt{n}}A_n$, where $A_n := M_n + X_n$, converges* (*in both senses*) *to a limiting measure μ.*

PROOF. We let $f_1, f_2, f_3, \ldots$ be an enumeration of a sequence of test functions which is dense in the uniform topology (such a sequence exists thanks to the Stone–Weierstrass theorem and the compact support of test functions). By applying the Bolzano–Weierstrass theorem once for each function in this sequence and then using the Arzelá–Ascoli diagonalization argument, we can refine the subsequence so that $\int_{\mathbb{C}} f_j(z)\, d\mu_{1/\sqrt{n}A_n}(z)$ converges in probability to some limit for each j and hence, by a limiting argument, $\int_{\mathbb{C}} g(z)\, d\mu_{1/\sqrt{n}A_n}(z)$ converges in probability to a limit for each test function g. By the Riesz representation function, we conclude that along this subsequence, $\mu_{1/\sqrt{n}A_n}$ converges in probability to some limit μ, which is also a probability measure, by the tightness bounds in Lemma 1.7.

Applying Theorem 1.15, we conclude that for almost every z, the expression

$$\frac{1}{n}\log\det\left(\left(\left(\frac{1}{\sqrt{n}}A_n - zI\right) + \varepsilon_n I\right)\left(\left(\frac{1}{\sqrt{n}}A_n - zI\right)^* + \varepsilon_n I\right)\right) \tag{1.8}$$

converges in probability to $2\int_{\mathbb{C}} \log|w-z|\,d\mu(w)$ along this sequence for some ε_n converging to zero. On the other hand, from the hypotheses and the theorem of Dozier and Silverstein (see Theorem B.1), we know that for almost every z, the expression (1.8) has an almost sure limit for the entire sequence of n. Combining the two facts, we see that for almost every z, (1.8) in fact converges almost surely to $2\int_{\mathbb{C}} \log|w-z|\,d\mu(w)$ for all n. The claim now follows from another application of Theorem 1.15. $\square$

1.7. *Notation*. Asymptotic notation is used under the assumption that $n \to \infty$, holding all other parameters fixed. Thus, for instance, if we say that a quantity $a_{z,n}$, depending on n and another parameter z, is equal to $o(1)$, this means that $a_{z,n}$ converges to zero as $n \to \infty$ for fixed z, but this convergence need not be uniform in z. As another example, the condition (1.3) is equivalent to asserting that $\|M_n\| = O(n)$ as $n \to \infty$.

2. The replacement principle.

The first step toward Theorem 1.5 is the following result that gives a general criterion for two random matrix ensembles $\frac{1}{\sqrt{n}}A_n$, $\frac{1}{\sqrt{n}}B_n$ to converge to the same limit.

THEOREM 2.1 (Replacement principle). *Suppose, for each n, that $A_n, B_n \in M_n(\mathbb{C})$ are ensembles of random matrices. Assume that*:

(i) *the expression*

$$\frac{1}{n^2}\|A_n\|_2^2 + \frac{1}{n^2}\|B_n\|_2^2 \tag{2.1}$$

is bounded in probability (*resp*., *almost surely*);

(ii) *for almost all complex numbers z*,

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right|$$

converges in probability (*resp*., *almost surely*) *to zero and*, *in particular*, *for each fixed z*, *these determinants are nonzero with probability $1 - o(1)$ for all n* (*resp*., *almost surely nonzero for all but finitely many n*).

Then, $\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}$ *converges in probability* (*resp*., *almost surely*) *to zero*.

We would like to remark here that we do not need to require independence among the entries of A_n and B_n. The proof of this theorem is rather "soft" in nature, relying primarily on the Stieltjes transform technique (following Girko [7]) that analyzes the ESD $\mu_{1/\sqrt{n}A_n}$ in terms of the log-determinants $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - zI)|$, combined with tools from classical real analysis, such as the dominated convergence theorem (see Lemma 3.1 for the precise version of this theorem that we need). The details are given in Section 3.

In view of Lemma 1.7, we see that Theorem 1.5 follows immediately from Theorem 2.1 and the following proposition.

PROPOSITION 2.2 (Converging determinant). *Let x and y be complex random variables with zero mean and unit variance. Let X_n and Y_n be $n \times n$ random matrices whose entries are i.i.d. copies of x and y, respectively. For each n, let M_n be a deterministic $n \times n$ matrix satisfying* (1.3). *Set $A_n := M_n + X_n$ and $B_n := M_n + Y_n$. Then, for every fixed $z \in \mathbb{C}$,*

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right| \tag{2.2}$$

converges in probability to zero. If, furthermore, we assume that (1.4) *converges to a limit for this value of z, then* (2.2) *converges almost surely to zero.*

For any square matrix A of size n, let $\lambda_i(A)$ and $s_i(A)$ be the eigenvalues and singular values of A. Furthermore, let $d_i(A)$ be the distance from the ith row vector of A to the subspace formed by the first $i-1$ row vectors. From linear algebra, we have the fundamental identity

$$|\det A| = \prod_{i=1}^{n}|\lambda_i(A)| = \prod_{i=1}^{n} s_i(A) = \prod_{i=1}^{n} d_i(A). \tag{2.3}$$

We will need to study the singular values and distances of $\frac{1}{\sqrt{n}}A_n - zI$ and $\frac{1}{\sqrt{n}}B_n - zI$ in order to estimate their determinants. The proof of Proposition 2.2, which occupies Sections 4, 5 and 6, is the heart of the paper. This proof relies on the following three ingredients:

- a result by Dozier and Silverstein [3] that compares the ESD of the singular values of the matrices $\frac{1}{\sqrt{n}}A_n - zI$ and $\frac{1}{\sqrt{n}}B_n - zI$—this will let us handle all the rows from 1 to $(1-\delta)n$ for some small $\delta > 0$;
- a lower tail estimate for the distance between a random vector and a fixed subspace of relatively large codimension, using a concentration inequality of Talagrand [12]—this will handle the contribution of the rows between $(1-\delta)n$ and (say) $n - n^{0.99}$;
- a polynomial lower bound for the least singular value of $\frac{1}{\sqrt{n}}A_n - zI$ and $\frac{1}{\sqrt{n}}B_n - zI$ from [25, 26]—this bound enables us to handle the contribution of the last $n^{0.99}$ rows.

3. The replacement principle. The purpose of this section is to establish Theorem 2.1. We begin with a version of the dominated convergence theorem.

LEMMA 3.1 (Dominated convergence). *Let (X, ν) be a finite measure space. For integers $n \geq 1$, let $f_n : X \to \mathbb{R}$ be random functions which are jointly measurable with respect to X and the underlying probability space. Assume that*:

(i) (*uniform integrability*) *there exists $\delta > 0$ such that $\int_X |f_n(x)|^{1+\delta}\, d\nu$ is bounded in probability* (*resp*., *almost surely*);

(ii) (*pointwise convergence in probability*) *for ν-almost every $x \in X$, $f_n(x)$ converges in probability* (*resp*., *almost surely*) *to zero*.

Then, $\int_X f_n(x)\, d\nu(x)$ converges in probability (*resp*., *almost surely*) *to zero*.

PROOF. We first prove the claim for convergence in probability. We can normalize ν to be a probability measure. Let $\varepsilon > 0$ be arbitrary. It suffices to show that

$$\int_X f_n(x)\, d\nu(x) = O(\varepsilon)$$

with probability $1 - O(\varepsilon) - o(1)$.

By hypothesis (i), we already know that with probability $1 - O(\varepsilon) - o(1)$,

$$\int_X |f_n(x)|^{1+\delta}\, d\nu(x) \leq C_\varepsilon$$

for some C_ε depending on ε. This implies that

$$\int_X f_n(x)\mathbf{I}\big(|f_n(x)| \geq M\big)\, d\nu(x) \leq C_\varepsilon / M^\delta$$

for any $M > 0$, where $\mathbf{I}(E)$ denotes the indicator of an event E. In particular, for M large enough, we have

$$\int_X f_n(x)\mathbf{I}\big(|f_n(x)| \geq M\big)\, d\nu(x) \leq \varepsilon$$

with probability $1 - O(\varepsilon) - o(1)$ and so it will suffice to show that

$$\int_X f_n(x)\mathbf{I}\big(|f_n(x)| \leq M\big)\, d\nu(x) = O(\varepsilon) \tag{3.1}$$

with probability $1 - o(1)$.

Fix M. By hypothesis, we have $\lim_{n\to\infty} \mathbf{P}(|f_n(x)| \geq \varepsilon) = 0$ for ν-almost every $x \in X$. By the dominated convergence theorem, we conclude that

$$\int_X \mathbf{P}\big(|f_n(x)| \geq \varepsilon\big)\, d\nu(x) = o(1).$$

By Fubini's theorem, we conclude that

$$\mathbf{E} \int_X \mathbf{I}\big(|f_n(x)| \geq \varepsilon\big)\, d\nu(x) = o(1)$$

and so by Markov's inequality, we have

$$\int_X \mathbf{I}(|f_n(x)| \geq \varepsilon)\, d\nu(x) = O(\varepsilon/M)$$

with probability $1 - o(1)$. The claim (3.1) easily follows.

We now prove the claim for almost sure convergence. Again, we let ν be a probability measure and $\varepsilon > 0$ be arbitrary. With probability $1 - O(\varepsilon)$, we have

$$\int_X |f_n(x)|^{1+\delta}\, d\nu(x) \leq C_\varepsilon$$

for all sufficiently large n, and some C_ε depending on n. Also, with probability 1, $f_n(x)$ converges to zero for almost every x. The claim now follows by invoking (the deterministic special case of) the convergence in probability version of the lemma that we have just proven. □

We now begin the proof of Theorem 2.1. We thus assume that A_n, B_n are as in that theorem. We shall first prove the claim for convergence in probability and later indicate how to modify the proof to obtain the principle for almost sure convergence.

From the boundedness in probability of (2.1) and Weyl's comparison inequality (Lemma A.2), we see that for every $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that for each n, the eigenvalues $\lambda_1, \dots, \lambda_n$ of A_n obey the bound

$$\sum_{j=1}^{n} \frac{1}{n^2} |\lambda_j|^2 \leq C_\varepsilon \tag{3.2}$$

or, equivalently, that

$$\int_{\mathbb{C}} |z|^2\, d\mu_{1/\sqrt{n}A_n}(z) \leq C_\varepsilon$$

with probability $1 - O(\varepsilon) - o(1)$. Similarly, we have

$$\int_{\mathbb{C}} |z|^2\, d\mu_{1/\sqrt{n}B_n}(z) \leq C_\varepsilon.$$

In particular, for each n, we see that, with probability $1 - O(\varepsilon) - o(1)$, we have the tightness bounds

$$\mu_{1/\sqrt{n}A_n}\{z \in \mathbb{C} : |z| \geq R\} \leq C_\varepsilon / R^2 \tag{3.3}$$

and

$$\mu_{1/\sqrt{n}B_n}\{z \in \mathbb{C} : |z| \geq R\} \leq C_\varepsilon / R^2 \tag{3.4}$$

for all $R > 0$.

We now take the standard step of passing from the ESDs $\mu_{1/\sqrt{n}A_n}, \mu_{1/\sqrt{n}B_n}$ to the characteristic functions $m_{1/\sqrt{n}A_n}, m_{1/\sqrt{n}B_n} : \mathbb{R}^2 \to \mathbb{C}$, which are defined by the formulae

$$m_{1/\sqrt{n}A_n}(u, v) := \int_{\mathbb{C}} e^{iu\operatorname{Re}(z)+iv\operatorname{Im}(z)}\, d\mu_{1/\sqrt{n}A_n}(z),$$
$$m_{1/\sqrt{n}B_n}(u, v) := \int_{\mathbb{C}} e^{iu\operatorname{Re}(z)+iv\operatorname{Im}(z)}\, d\mu_{1/\sqrt{n}B_n}(z).$$

Thus, the functions $m_{1/\sqrt{n}A_n}, m_{1/\sqrt{n}B_n}$ are continuous and are bounded uniformly in magnitude by 1.

Thanks to the tightness bounds (3.3) and (3.4), we can easily pass back and forth between convergence of ESDs and convergence of characteristic functions.

LEMMA 3.2. *Let the notation and assumptions be as above. The following are then equivalent*:

(i) $\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}$ *converges in probability*;
(ii) *for almost every* u, v, $m_{1/\sqrt{n}A_n}(u, v) - m_{1/\sqrt{n}B_n}(u, v)$ *converges in probability.*

PROOF. We first show that (i) implies (ii). Fix u, v and let $\varepsilon > 0$ be arbitrary. From (3.3), (3.4) we can find an R depending on C_ε and ε such that

$$\mu_{1/\sqrt{n}A_n}(\{z \in \mathbb{C} : |z| \geq R\}) + \mu_{1/\sqrt{n}B_n}(\{z \in \mathbb{C} : |z| \geq R\}) \leq \varepsilon$$

with probability $1 - O(\varepsilon) - o(1)$. In particular, with probability $1 - O(\varepsilon) - o(1)$, we have

$$\begin{aligned} &m_{1/\sqrt{n}B_n}(u, v) - m_{1/\sqrt{n}A_n}(u, v) \\ &\quad = \int \psi(z/R) e^{iu\operatorname{Re}(z)+iv\operatorname{Im}(z)} \big[d\mu_{1/\sqrt{n}B_n}(z) - d\mu_{1/\sqrt{n}A_n}(u, v)(z)\big] + O(\varepsilon), \end{aligned}$$

where ψ is any smooth compactly supported function which equals one on the unit ball. However, since $\mu_{1/\sqrt{n}B_n} - \mu_{1/\sqrt{n}A_n}$ converges in probability, the integral here converges to zero in probability. The claim follows.

We now prove that (ii) implies (i). Since continuous compactly supported functions are the uniform limit of smooth compactly supported functions, it suffices to show that $\int_{\mathbb{C}} f\, d\mu_{1/\sqrt{n}A_n} - \int_{\mathbb{C}} f\, d\mu_{1/\sqrt{n}B_n}$ converges in probability to zero for every smooth compactly supported function $f : \mathbb{C} \to \mathbb{C}$.

Now, fix a smooth compactly supported function $f : \mathbb{C} \to \mathbb{C}$. By Fourier analysis, we can write

$$\begin{aligned} &\int_{\mathbb{C}} f\, d\mu_{1/\sqrt{n}A_n} - \int_{\mathbb{C}} f\, d\mu_{1/\sqrt{n}B_n} \\ &\qquad = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}(u, v) \big(m_{1/\sqrt{n}A_n}(u, v) - m_{1/\sqrt{n}B_n}(u, v)\big)\, du\, dv \end{aligned} \tag{3.5}$$

for some smooth, rapidly decreasing function $\hat{f}$. In particular, the measure $d\nu = \hat{f}(u,v)\,du\,dv$ is finite. The claim now follows from dominated convergence (Lemma 3.1). Note that the function $m_{1/\sqrt{n}A_n} - m_{1/\sqrt{n}B_n}$ is bounded and so clearly obeys the moment condition required in that lemma. $\square$

In view of the above lemma, it suffices to show that $m_{1/\sqrt{n}A_n}(u,v) - m_{1/\sqrt{n}B_n}(u,v)$ converges in probability to zero for almost every $u, v \in \mathbb{R}$.

Fix u, v. Since we can exclude a set of measure zero, we can assume that u, v are nonzero. We allow all implied constants in the arguments below to depend on u, v.

Following Girko [7], we now proceed via the Stieltjes-like transform $g_{1/\sqrt{n}A_n} : \mathbb{C} \to \mathbb{R}$, defined almost everywhere by the formula

$$
\begin{aligned}
g_{1/\sqrt{n}A_n}(z) &:= 2\operatorname{Re}\int_{\mathbb{C}} \frac{z-w}{|z-w|^2}\,d\mu_{1/\sqrt{n}}(w) \\
&= \frac{2}{n}\operatorname{Re}\sum_{j=1}^{n} \frac{z - 1/\sqrt{n}\lambda_j}{|z - 1/\sqrt{n}\lambda_j|^2}.
\end{aligned}
\tag{3.6}
$$

Observe that this is a locally integrable function on $\mathbb{C}$ and that

$$
g_{1/\sqrt{n}A_n}(z) = \frac{\partial}{\partial \operatorname{Re}(z)} \frac{2}{n} \log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right|
\tag{3.7}
$$

for all but finitely many z.

We have the following fundamental identity.

LEMMA 3.3 (Girko's identity [7]). *For every nonzero u, v, we have*

$$
m_{1/\sqrt{n}A_n}(u,v) = \frac{u^2+v^2}{4\pi i u}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} g_{1/\sqrt{n}A_n}(s+it)e^{ius+ivt}\,dt\right)ds,
$$

where the inner integral is absolutely integrable for almost every s and the outer integral is absolutely convergent.

PROOF. We argue as in [2], Lemma 3.1. Since

$$
m_{1/\sqrt{n}A_n}(u,v) = \frac{1}{n}\sum_{j=1}^{n} e^{i(u\operatorname{Re}(1/\sqrt{n}\lambda_j)+v\operatorname{Im}(1/\sqrt{n}\lambda_j))},
$$

it suffices, from (3.6), to show that

$$
e^{i(u\operatorname{Re}(w)+v\operatorname{Im}(w))} = \frac{u^2+v^2}{2\pi i u}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} \frac{\operatorname{Re}(s+it-w)}{|s+it-w|^2}e^{ius+ivt}\,dt\right)ds
$$

for each complex number w, with an absolutely convergent inner integral and outer integral. However, standard contour integration shows that

$$\begin{aligned}\int_{\mathbb{R}} \frac{\operatorname{Re}(s+it-w)}{|s+it-w|^2} e^{ius+ivt}\, dt \\ = \pi \operatorname{sgn}(s-\operatorname{Re}(w)) e^{-v|s-\operatorname{Re}(w)|} e^{ius} e^{iv\operatorname{Im}(w)}\end{aligned} \tag{3.8}$$

for every $s \neq \operatorname{Re}(w)$ and the claim follows by an elementary integration. □

We can, of course, define $g_{1/\sqrt{n}B_n}$ similarly, with analogous identities. To conclude the proof of Theorem 2.1, it thus suffices to show that for any $\varepsilon > 0$ and any n, we have

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} (g_{1/\sqrt{n}A_n}(s+it) - g_{1/\sqrt{n}B_n}(s+it)) e^{ius+ivt}\, dt \right) ds = O(\varepsilon) \tag{3.9}$$

with probability $1 - O(\varepsilon) - o(1)$.

Fix $\varepsilon > 0$. By (3.3), (3.4), we can find an $R > 1$ large enough that, with probability $1 - O(\varepsilon)$,

$$\mu_{1/\sqrt{n}A_n}(\{z \in \mathbb{C} : |z| \geq R\}) + \mu_{1/\sqrt{n}B_n}(\{z \in \mathbb{C} : |z| \geq R\}) \leq \varepsilon. \tag{3.10}$$

We now condition on the event that (3.10) holds.

We now smoothly localize the z variable to a compact set, as follows. Let $\psi : \mathbb{R} \to \mathbb{R}^+$ be a smooth cutoff function which equals 1 on $[-1, 1]$ and is supported on $[-2, 2]$.

LEMMA 3.4 (Truncation in s, t). *Let $w \in \mathbb{C}$.*

(i) *The integral*

$$\int_{\mathbb{R}} \left| \int_{\mathbb{R}} \frac{\operatorname{Re}(w-(s+it))}{|w-(s+it)|^2} e^{ius+ivt}\, dt \right| (1 - \psi(s/R^2))\, ds$$

is of size $O(1)$ and (if R is large enough) is of size $O(\varepsilon)$ when $|w| \leq R$.

(ii) *The integral*

$$\int_{\mathbb{R}} \left| \int_{\mathbb{R}} \frac{\operatorname{Re}(w-(s+it))}{|w-(s+it)|^2} e^{ius+ivt} (1 - \psi(t/R^2))\, dt \right| \psi(s/R^2)\, ds \tag{3.11}$$

is of size $O(1)$ and (if R is large enough) is of size $O(\varepsilon)$ when $|w| \leq R$.

PROOF. Claim (i) follows easily from (3.8), so we turn to (ii). We first verify the claim that (3.11) is bounded. Replacing everything by absolute values, one sees that

$$\left| \int_{\mathbb{R}} \frac{\operatorname{Re}(w-(s+it))}{|w-(s+it)|^2} e^{ius+ivt} (1 - \psi(t/R^2))\, dt \right| = O(1)$$

(in fact, one can obtain an explicit upper bound of π), so we can dispose of the region of integration in which $s = \mathrm{Re}(w) + O(1)$. For the remaining values of s, we use repeated integration by parts, integrating the e^{ivt} term and differentiating the others. After two such integrations, we obtain the bound

$$\left|\int_{\mathbb{R}} \frac{\mathrm{Re}(w-(s+it))}{|w-(s+it)|^2} e^{ius+ivt}(1-\psi(t/R^2))\,dt\right| = O\big((R^{-2}+|s-\mathrm{Re}(w)|^{-1})^2\big).$$

The claim then follows.

Finally, if $|w| \le R$, one can easily verify (by repeated integration by parts) that

$$\int_{\mathbb{R}} \frac{\mathrm{Re}(w-(s+it))}{|w-(s+it)|^2} e^{ius+ivt}(1-\psi(t/R^2))\,dt = O(1/R^4)$$

(say) and so the final claim of (ii) follows. □

From this lemma, (3.6), the triangle inequality and (3.10), we conclude that

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} g_{1/\sqrt{n}A_n}(s+it) e^{ius+ivt}\,dt\right)(1-\psi(s/R^2))\,ds = O(\varepsilon) \tag{3.12}$$

and

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} g_{1/\sqrt{n}A_n}(s+it) e^{ius+ivt}(1-\psi(t/R^2))\,dt\right)\psi(s/R^2)\,ds = O(\varepsilon). \tag{3.13}$$

From (3.12), (3.13) (and their counterparts for $g_{1/\sqrt{n}B_n}$) and the triangle inequality, we thus see that to prove (3.9), it suffices to show that

$$\begin{aligned}\int_{\mathbb{R}}\int_{\mathbb{R}} &(g_{1/\sqrt{n}A_n}(s+it) - g_{1/\sqrt{n}B_n}(s+it))e^{ius+ivt} \\ &\times \psi(t/R^2)\psi(s/R^2)\,dt\,ds\end{aligned} \tag{3.14}$$

converges in probability to zero for every fixed $R \ge 1$. Note that the integrands here are now jointly absolutely integrable in t, s and so we may now freely interchange the order of integration.

Fix R. Using (3.7) and integration by parts in the s variable, we can rewrite (3.14) in the form

$$\int_{\mathbb{R}}\int_{\mathbb{R}} f_n(s,t)\phi_{u,v,R}(s,t)\,ds\,dt,$$

where

$$f_n(s,t) := \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right|$$

and

$$\phi_{u,v,R}(s,t) := -\frac{\partial}{\partial s}\big(e^{ius+ivt}\psi(t/R^2)\psi(s/R^2)\big).$$

(Note that there are finitely many values of t for which the integration by parts is not justified due to singularities in $g_{1/\sqrt{n}A_n}$ or $g_{1/\sqrt{n}B_n}$, but these values of t clearly give a zero contribution at the end of the day.) Thus, it will suffice to show that

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |f_n(s,t)||\phi_{u,v,R}(s,t)|\,ds\,dt$$

converges in probability to zero.

From (2.3), we have

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| = \frac{1}{n}\sum_{j=1}^{n}\log\left|\frac{1}{\sqrt{n}}\lambda_j - (s+it)\right| \tag{3.15}$$

and similarly for B_n. From the boundedness and compact support of $\phi_{u,v,R}$, we observe that

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \log\left|\frac{1}{\sqrt{n}}\lambda - (s+it)|^2\right|\phi_{u,v,R}(s,t)|\,ds\,dt \le O_{\phi_{u,v,R}}\left(1 + \frac{1}{n}|\lambda|^2\right)$$

for all $\lambda \in \mathbb{C}$; from this, (3.15), (3.2) and the triangle inequality, we see that

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |f_n(s,t)|^2|\phi_{u,v,R}(s,t)|\,ds\,dt \tag{3.16}$$

is bounded uniformly in n. Since, by hypothesis, $f_n(s,t)$ converges in probability to zero for almost every s, t, the claim now follows from dominated convergence (Lemma 3.1). The proof of Theorem 2.1 is now complete in the case of convergence in probability.

3.1. *The almost sure convergence case.* We now indicate how to adapt the above arguments to the case of almost sure convergence. First, since (2.1) is now almost surely bounded, instead of just bounded in probability, we can now say that for every $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that with probability $1 - O(\varepsilon)$, (3.3), (3.4) holds for all sufficiently large n [as opposed to these bounds holding with probability $1 - O(\varepsilon) - o(1)$ for each n separately].

Next, we observe the (well-known) fact that Lemma 3.2 continues to hold when convergence in probability is replaced by almost sure convergence throughout. Indeed, the implication of (ii) from (i) is nearly identical and is left as an exercise to the reader. To deduce (i) from (ii) in the almost sure case, observe, from the separability of the space of smooth compactly supported functions in the uniform topology, that it suffices to show that (3.5) converges almost surely to zero for each f. On the other hand, from (ii) and Fubini's theorem, we know that with probability 1, $m_{1/\sqrt{n}A_n}(u,v) - m(u,v)$ converges to zero for almost every u, v and the claim follows from the (ordinary) dominated convergence theorem.

Once again, we use Girko's identity, Lemma 3.3, and reduce to showing that for every $\varepsilon > 0$, one has, with probability $1 - O(\varepsilon)$, that (3.9) holds for all but finitely many n. From our bounds on (3.3), (3.4), we see that, with probability

$1 - O(\varepsilon)$, (3.10) holds for all but finitely many n. We apply Lemma 3.4 (which is deterministic) and reduce to showing that (3.14) converges almost surely to zero for each fixed $R \geq 1$. The rest of the argument proceeds as in the convergence in probability case.

3.2. *An alternate argument*. There is an alternate derivation[5] of Theorem 2.1 that avoids Fourier analysis and is instead based on the observation that, for any complex polynomial $P(z)$, the distributional Laplacian $\Delta \log|P(z)|$ of the logarithm of the magnitude of P is equal to the counting measure of the zeros of P (counting multiplicity). In particular, we see from Green's theorem that

$$\int_{\mathbb{C}} f \, d(\mu_{1/\sqrt{n}A_n} - \mu_{1/\sqrt{n}B_n}) = \frac{1}{2\pi n} \int_{\mathbb{C}} (\Delta f(z)) \log\left|\det\left(\frac{1}{\sqrt{n}} A_n - zI\right)\right| - \frac{1}{n} \log\left|\det\left(\frac{1}{\sqrt{n}} B_n - zI\right)\right|$$

for any smooth, compactly supported f. Applying Lemma 3.1, we can then get convergence of this integral (either in probability or in the almost sure sense, as appropriate); the uniform integrability required can be established by repeating the computations used to bound (3.16). One can then easily take limits to replace smooth compactly supported f to continuous compactly supported f; we omit the details.

4. Proof of Proposition 2.2. In this section, we present the proof of Proposition 2.2, modulo several key lemmas. Let x, y, M_n, A_n, B_n, z be as in that proposition. By shifting M_n by $\sqrt{n} z I$ if necessary, we can assume $z = 0$. Our task is now to show that

$$\frac{1}{n} \log\left|\det\left(\frac{1}{\sqrt{n}} A_n\right)\right| - \frac{1}{n} \log\left|\det\left(\frac{1}{\sqrt{n}} B_n\right)\right|$$

converges in probability to zero and also almost surely to zero if $\mu_{1/n M_n M_n^*}$ converges.

Let us first remark that the almost sure convergence claim implies the convergence in probability claim. Indeed, suppose that convergence in probability failed. There would then exist an $\varepsilon > 0$ such that

$$\mathbf{P}\left(\left|\frac{1}{n} \log\left|\det\left(\frac{1}{\sqrt{n}} A_n\right)\right| - \frac{1}{n} \log\left|\det\left(\frac{1}{\sqrt{n}} B_n\right)\right|\right| \geq \varepsilon\right) \geq \varepsilon \tag{4.1}$$

for a subsequence of n. By vague sequential compactness, one can pass to a further subsequence along which $\mu_{1/n M_n M_n^*}$ converges and, hence, by hypothesis, one has almost sure (and hence in-probability) convergence to zero along this sequence,

[5] We thank Manjunath Krishnapur for this simpler argument.

contradicting (4.1). Thus, it suffices to establish almost sure convergence assuming the convergence of $\mu_{1/nM_nM_n^*}$.

Let $Z_1, \dots, Z_n$ be the rows of M_n. By assumption (1.3), we have

$$\sum_{j=1}^{n} \|Z_i\|^2 = O(n^2).$$

In particular, at least half of the Z_i have norm $O(\sqrt{n})$. By permuting the rows of M_n, A_n, B_n if necessary, we may assume that the last half of the rows have this property, thus

$$\|Z_i\| = O(\sqrt{n}) \qquad \text{for all } n/2 \le i \le n. \tag{4.2}$$

Let $\sigma_1(A) \ge \dots \ge \sigma_n(A) \ge 0$ denote the singular values of a matrix A. We have the following fundamental lower bound.

LEMMA 4.1 (Least singular value bound). *With probability* 1, *we have*

$$\sigma_n(A_n), \sigma_n(B_n) \ge n^{-O(1)} \tag{4.3}$$

for all but finitely many n. In particular, with probability 1, *A_n and B_n are invertible for all but finitely many n.*

PROOF. This follows immediately from [25], Theorem 2.1 or [26], Theorem 4.1 and the Borel–Cantelli lemma, noting from (1.3) of Proposition 2.2 that the operator norm of M_n is of polynomial size $n^{O(1)}$. There are previous results in [16, 19, 23, 24] which handled special cases with more assumptions on M_n and the underlying distributions x, y (e.g., in some of the prior results, M_n was assumed to vanish or x, y were assumed to be integer-valued or to have finite higher moments). One can obtain explicit bounds on the tail probability and on the exponent $O(1)$; see [26]. However, for our applications, the above bounds will suffice. □

We also have, with probability 1, the crude upper bound

$$\sigma_1(A_n), \qquad \sigma_1(B_n) \le n^{O(1)} \tag{4.4}$$

for all but finitely many n, which follows easily from the polynomial size of M_n, the bounded second moment of x, y and the Borel–Cantelli lemma. Again, much sharper bounds are available, especially if x and y have finite fourth moment, but we will not need these bounds here.

Let $X_1, \dots, X_n$ be the rows of A_n and, for each $1 \le i \le n$, let V_i be the $(i-1)$-dimensional space generated by $X_1, \dots, X_{i-1}$. From (2.3), we have

$$\frac{1}{n} \log\left|\det\left(\frac{1}{\sqrt{n}} A_n\right)\right| = \frac{1}{n} \sum_{i=1}^{n} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right)$$

and, similarly,

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n\right)\right| = \frac{1}{n}\sum_{i=1}^{n}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right),$$

where $Y_1, \dots, Y_n$ are the rows of $\frac{1}{\sqrt{n}}B_n$ and W_i is spanned by $Y_1, \dots, Y_{i-1}$. Our task is then to show that

$$\frac{1}{n}\sum_{i=1}^{n}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right) - \log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right)$$

converges almost surely to zero.

From (4.3), (4.4) and Lemma A.4, we almost surely obtain the bound

$$\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right), \qquad \log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right) = O(\log n)$$

for all but finitely many n. It thus suffices to show that

$$\frac{1}{n}\sum_{1\le i\le n-n^{0.99}}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right) - \log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right)$$

(say) converges almost surely to zero. This follows immediately from the following two lemmas.

LEMMA 4.2 (High-dimensional contribution). *For every* $\varepsilon > 0$, *there exists* $0 < \delta < 1/2$ *such that*, *with probability* 1, *one has*

$$\frac{1}{n}\sum_{(1-\delta)n\le i\le n-n^{0.99}}\left|\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right)\right| = O(\varepsilon)$$

for all but finitely many n. *A similar result holds with* $\operatorname{dist}(\frac{1}{\sqrt{n}}X_i, V_i)$ *replaced by* $\operatorname{dist}(\frac{1}{\sqrt{n}}Y_i, W_i)$.

LEMMA 4.3 (Low-dimensional contribution). *For every* $\varepsilon > 0$, *there exists* $0 < \delta < 1/2$ *such that*, *with probability* $1 - O(\varepsilon)$, *one has*

$$\frac{1}{n}\sum_{1\le i\le (1-\delta)n}\log\operatorname{dist}\left(\frac{1}{\sqrt{n}}X_i, V_i\right) - \log\operatorname{dist}\left(\frac{1}{\sqrt{n}}Y_i, W_i\right) = O(\varepsilon)$$

for all but finitely many n.

The next two sections will be devoted to the proofs of these two lemmas.

5. Proof of Lemma 4.2. We now prove Lemma 4.2. We can, of course, take n to be large depending on all fixed parameters. Let $0 < \delta < 1/2$ be a small number depending on ε, to be chosen later.

Clearly, it suffices to prove this lemma for $\operatorname{dist}(\frac{1}{\sqrt{n}} X_i, V_i)$. We first prove the (much easier) bound for the positive component of the logarithm. By the Borel–Cantelli lemma, it suffices to show that

$$\sum_{n=1}^{\infty} \mathbf{P}\biggl(\frac{1}{n} \sum_{(1-\delta)n \le i \le n - n^{0.99}} \max\biggl(\log \operatorname{dist}\biggl(\frac{1}{\sqrt{n}} X_i, V_i\biggr), 0\biggr) \ge \varepsilon\biggr) < \infty.$$

To establish this, we use the crude bound

$$\max\biggl(\log \operatorname{dist}\biggl(\frac{1}{\sqrt{n}} X_i, V_i\biggr), 0\biggr) \le \max\biggl(\log \frac{1}{\sqrt{n}} \|X_i\|, 0\biggr)$$

and thus

$$\begin{aligned} &\frac{1}{n} \sum_{(1-\delta)n \le i \le n - n^{0.99}} \max\biggl(\log \operatorname{dist}\biggl(\frac{1}{\sqrt{n}} X_i, V_i\biggr), 0\biggr) \\ &\quad \le O\Biggl(\sum_{m=0}^{\infty} \frac{1}{n} \sum_{(1-\delta)n \le i \le n - n^{0.99}} \mathbf{I}(\|X_i\| \ge 2^m \sqrt{n})\Biggr). \end{aligned} \tag{5.1}$$

Thus if the left-hand side of (5.1) exceeds ε, we must have

$$\frac{1}{n} \sum_{(1-\delta)n \le i \le n - n^{0.99}} \mathbf{I}(\|X_i\| \ge 2^m \sqrt{n}) \ge \varepsilon/(100+m)^2$$

(say) for some $m \ge 0$. On the other hand, from (4.2) and the second moment method, we see that $\mathbf{P}(\|X_i\| \ge 2^m \sqrt{n}) = O(2^{-2m})$ and, thus, by Hoeffding's inequality, we have

$$\begin{aligned} &\mathbf{P}\biggl(\frac{1}{n} \sum_{(1-\delta)n \le i \le n - n^{0.99}} \mathbf{I}(\|X_i\| \ge 2^m \sqrt{n}) \varepsilon/(100+m)^2\biggr) \\ &\quad \le C \exp(-cn^{-0.01} - cm^{-0.01}) \end{aligned}$$

(say) for some constants $C, c > 0$ depending on ε, if δ is chosen sufficiently small, depending on ε. The claim follows.

It remains to establish the bound for the negative component of the logarithm. By the Borel–Cantelli lemma, it suffices to show that

$$\sum_{n=1}^{\infty} \mathbf{P}\biggl(\frac{1}{n} \sum_{(1-\delta)n \le i \le n - n^{0.99}} \max\biggl(-\log \operatorname{dist}\biggl(\frac{1}{\sqrt{n}} X_i, V_i\biggr), 0\biggr) \ge \varepsilon\biggr) < \infty.$$

This will follow from the union bound and the following estimate.

PROPOSITION 5.1 (Lower tail bound). *Let $1 \le d \le n - n^{0.99}$ and $0 < c < 1$, and let W be a (deterministic) d-dimensional subspace of $\mathbb{C}^n$. Let X be a row of A_n (the exact choice of row is not important). Then,*

$$\mathbf{P}\big(\mathrm{dist}(X, W) \le c\sqrt{n-d}\big) = O(\exp(-n^{0.01})).$$

(*The implied constant of course depends on c.*)

Indeed, since X_i and V_i are independent of each other, the proposition implies that

$$\mathrm{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) \ge \frac{1}{2\sqrt{n}}\sqrt{n-i+1}$$

(say) for each $(1-\delta)n \le i \le n - n^{0.99}$, with probability $1 - O(n^{-10})$ (say). Setting δ sufficiently small (compared to ϵ), taking logarithms and summing in i and n, one obtains the claim.

It remains to prove the proposition. Similar lower bounds concerning the distance of a random vector to a fixed subspace have appeared in [18, 19, 21]. Here, however, we have the complication that the coefficients of X have nonzero mean and no higher moment bounds than the second moment; in particular, they can be unbounded.

We first eliminate the problem that X has nonzero mean. Write $X = v + X'$, where $v := \mathbf{E}(X)$ is a deterministic vector (which could be quite large) and X' has mean zero. We then have $\mathrm{dist}(X, W) \ge \mathrm{dist}(X', \mathrm{span}(W, v))$. Thus, Proposition 5.1 follows from the mean zero case (after making the harmless change of incrementing d to $d+1$ and adjusting the parameters slightly to suit this).

Henceforth, we assume that X has mean zero, thus $X = (x_1, \dots, x_n)$ for some i.i.d. copies $x_1, \dots, x_n$ of x. We now deal with the problem that the $x_1, \dots, x_n$ can be unbounded. By Chebyshev's inequality, we have $\mathbf{P}(|x_i| \ge n^{0.1}) = O(n^{-0.2})$ for all $1 \le i \le n$. The event $|x_i| \ge n^{0.1}$ are jointly independent in i. By Chernoff's inequality (see, e.g., [22], Chapter 1), we can show that, with probability $1 - O(\exp(-n^{0.01}))$, there are at most $n^{0.9}$ indices i for which $|x_i| \ge n^{0.1}$. (One can also verify this directly using binomial coefficients and Sterling's formula.)

By conditioning on the various possible sets of indices for which $|x_i| \ge n^{0.1}$, we see that it suffices to show that

$$\mathbf{P}\big(\mathrm{dist}(X, W) \le c\sqrt{n-d}\,|E_I\big) = O(\exp(-n^{0.01}))$$

for each $I \subset \{1, \dots, n\}$ of cardinality at most $n^{0.9}$, where E_I is the event that $I = \{1 \le i \le n : |x_i| \ge n^{0.1}\}$.

Without loss of generality, we can take $I = \{n'+1, \dots, n\}$ for some $n - n^{0.9} \le n' \le n$. We then observe that

$$\mathrm{dist}(X, W) \ge \mathrm{dist}(\pi(X), \pi(W)),$$

where $\pi : \mathbb{C}^n \to \mathbb{C}^{n'}$ is the orthogonal projection. By conditioning on the coordinates $x_{n'+1}, \dots, x_n$ and making the minor change of replacing n with n' (and adjusting c slightly), we may thus reduce to the case where I is empty. It thus suffices to show that

$$\mathbf{P}\big(\operatorname{dist}(X, W) \le c\sqrt{n-d} \,|\, |x_i| < n^{0.1} \text{ for all } i\big) = O(\exp(-n^{0.01})).$$

Let $\tilde{x}$ be the random variable x conditioned on the event $|x| < n^{0.1}$ and let $\tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_n)$ be a vector consisting of i.i.d. copies of $\tilde{x}$. It then suffices to show that

$$\mathbf{P}\big(\operatorname{dist}(\tilde{X}, W) \le c\sqrt{n-d}\big) = O(\exp(-n^{0.01})). \tag{5.2}$$

Note that $\tilde{x}$ might have a nonzero mean, but this can be easily dealt with using the same trick as was used before: subtracting $\mathbf{E}\tilde{x}$ from $\tilde{x}$ to make X have zero mean. Since x had variance 1, we see from monotone convergence that $\tilde{x}$ has variance $1 - o(1)$.

To prove (5.2), we recall the following inequality of Talagrand.

THEOREM 5.2 (Talagrand's inequality). *Let $\mathbf{D}$ be the unit disk $\{z \in \mathbb{C}, |z| \le 1\}$. For every product probability μ on $\mathbf{D}^n$, every convex 1-Lipschitz function $F : \mathbb{C}^n \to \mathbb{R}$ and every $r \ge 0$,*

$$\mu\big(|F - M(F)| \ge r\big) \le 4\exp(-r^2/8),$$

where $M(F)$ denotes the median of F.

PROOF. This is the complex version of [12], Corollary 4.10, in which $\mathbf{D}$ was replaced by the unit interval [0, 1]. The proof is the same, with a slight modification that implies a worse constant (1/8 instead of 1/4) in the exponent. □

We apply this theorem with μ equal to the distribution of $\tilde{X}/n^{0.1}$ and $F : \mathbb{C}^n \to \mathbb{R}$ equal to the convex 1-Lipschitz function $F(v) := \operatorname{dist}(v, W)$, and conclude that

$$\mathbf{P}\big(|\operatorname{dist}(\tilde{X}, W) - M(\operatorname{dist}(\tilde{X}, W))| \ge n^{0.1} r\big) \le 4\exp(-r^2/8) \tag{5.3}$$

for every $r > 0$. On the other hand, we can easily compute the second moment (cf. [21], Lemma 2.5):

LEMMA 5.3. *We have*

$$\mathbf{E}(\operatorname{dist}(\tilde{X}, W)^2) = \big(1 - o(1)\big)(n - d).$$

PROOF. Let $\pi = (\pi_{ij})_{1 \le i, j \le n}$ be the orthogonal projection matrix to W. Observe that $\operatorname{dist}(\tilde{X}, W)^2 = \sum_{i=1}^n \sum_{j=1}^n \tilde{x}_i \pi_{ij} \overline{\tilde{x}}_j$. Since the $\tilde{x}_i$ are i.i.d. with mean zero, we thus have

$$\mathbf{E}(\operatorname{dist}(\tilde{X}, W)^2) = (\mathbf{E}\tilde{x}^2) \sum_{i=1}^n \pi_{ii}.$$

However, $\sum_{i=1}^{n} \pi_{ii} = \operatorname{trace}(\pi)$ is equal to $\tilde{n}$. Since $\tilde{x}$ had variance $1 - o(1)$, the claim follows. □

Since $n - d \geq n^{0.99}$ and $c < 1$, the claim (5.2) follows from (5.3) and the above lemma. The proof of Lemma 4.2 is now complete.

6. Proof of Lemma 4.3. We now begin the proof of Lemma 4.3. Fix ε and assume that δ is sufficiently small, depending on ε. Write $n' := \lfloor (1-\delta)n \rfloor$. Observe that $\prod_{i=1}^{n'} \operatorname{dist}(\frac{1}{\sqrt{n}} X_i, V_i)$ is the n'-dimensional volume of the parallelepiped spanned by $X_1, \dots, X_{n'}$, which is also equal to $\det(\frac{1}{n} A_{n,n'} A_{n,n'}^*)^{1/2}$, where $A_{n,n'}$ is the $n' \times n$ matrix with rows $X_1, \dots, X_{n'}$. Expressing this determinant as the product of singular values, we arrive at the identity

$$\frac{1}{n} \sum_{1 \leq i \leq (1-\delta)n} \log \operatorname{dist}\biggl(\frac{1}{\sqrt{n}} X_i, V_i\biggr) = \frac{1}{n} \sum_{i=1}^{n'} \log\biggl(\frac{1}{\sqrt{n}} \sigma_i(A_{n,n'})\biggr).$$

A similar result holds for Y_i, W_i and $B_{n,n'}$ (the matrix generated by $Y_1, \dots, Y_{n'}$). It thus suffices to show that, with probability $1 - O(\varepsilon)$, one has

$$\frac{1}{n'} \sum_{i=1}^{n'} \log\biggl(\frac{1}{\sqrt{n}} \sigma_i(A_{n,n'})\biggr) - \log\biggl(\frac{1}{\sqrt{n}} \sigma_i(B_{n,n'})\biggr) = O(\varepsilon) \tag{6.1}$$

for all but finitely many n. We rewrite (6.1) as

$$\int_0^\infty \log t \, d\nu_{n,n'}(t) = O(\varepsilon), \tag{6.2}$$

where $d\nu_{n,n'}$ is the difference of two ESDs:

$$d\nu_{n,n'} = \mu_{1/n' A_{n,n'} A_{n,n'}^*} - \mu_{1/n' B_{n,n'} B_{n,n'}^*}.$$

We control (6.1) by dividing the range of t into several parts.

6.1. *The region of very large t.* We now control the region where $t \geq R_\varepsilon$ for some large R_ε.

From Lemma A.2, we have that

$$\frac{1}{n} \sum_{i=1}^{n'} \biggl(\frac{1}{\sqrt{n}} \sigma_i(A_{n,n'})\biggr)^2, \qquad \frac{1}{n} \sum_{i=1}^{n'} \biggl(\frac{1}{\sqrt{n}} \sigma_i(B_{n,n'})\biggr)^2$$

is almost surely bounded and thus

$$\int_0^\infty t |d\nu_{n,n'}(t)|$$

is also almost surely bounded. Thus, with probability $1 - O(\varepsilon)$, we have

$$\int_0^\infty t |d\nu_{n,n'}(t)| \leq C_\varepsilon$$

for all but finitely many n, and some C_ε independent of n, which implies that

$$\int_{R_\varepsilon}^{\infty} |\log t||d\nu_{n,n'}(t)| \leq \varepsilon \tag{6.3}$$

for all but finitely many n and some R_ε depending only on ε.

6.2. *The region of intermediate t.* We now control the region $\varepsilon^4 \leq t \leq R_\varepsilon$.

LEMMA 6.1. *Let ψ be a smooth function which equals* 1 *on* $[\varepsilon^4, R_\varepsilon]$ *and is supported on* $[\varepsilon^4/2, 2R_\varepsilon]$. *Then*, *with probability* 1, *we have*

$$\int_0^{\infty} \psi(t) \log t \, d\nu_{n,n'}(t) = O(\varepsilon), \tag{6.4}$$

if δ is sufficiently small, *depending on ε and ψ.*

PROOF. From the interlacing property (Lemma A.1), we see that

$$\int_0^{\infty} \psi(t) \log t \, d\nu_{n,n'}(t) = \int_0^{\infty} \psi(t) \log t \, d\nu_{n,n}(t) + O(\varepsilon),$$

if δ is sufficiently small, depending on ε and ψ.

We now apply the recent result in [3], Theorem 1.1. For the reader's convenience, we restate this result in the Appendix; see Theorem B.1. This result asserts, under the above hypotheses, that the ESDs $d\mu_{1/nA_nA_n^*}$ and $d\mu_{1/nB_nB_n^*}$ converge almost surely to the same limit [in fact, this limit is given explicitly in terms of the limiting distribution of $\mu_{1/nM_nM_n^*}$ via the inverse Stieltjes transform of (B.1)]. In particular, $\nu_{n,n}$ converges almost surely to zero and the claim follows. □

REMARK 6.2. Note that, for the convergence in probability case of Proposition 2.2, we need to apply Theorem B.1 to a subsequence of n, rather than to all n, thanks to the subsequence extraction performed at the beginning of Section 4.

6.3. *The region of moderately small t.* We now control the region $\delta^2 \leq t \leq \varepsilon^4$. For this, we need some bounds on the low singular values of $A_{n,n'}$ and $B_{n,n'}$.

LEMMA 6.3. *With probability* 1, *we have*

$$\frac{1}{n} \sum_{i=1}^{n'} \left(\frac{1}{\sqrt{n}} \sigma_i(A_{n,n'}) \right)^{-2} = O(1) \tag{6.5}$$

for all but finitely many n, *and similarly with $A_{n,n'}$ replaced by $B_{n,n'}$.*

PROOF. Clearly, it suffices to establish the claim for $A_{n,n'}$. Using Proposition 5.1 and the Borel–Cantelli lemma, we see that, with probability 1, we have

$$\operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, \operatorname{span}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{n'}) \right) \geq \frac{1}{2} \sqrt{\delta n}$$

for all but finitely many n and all $1 \leq i \leq n'$. The claim then follows from Lemma A.4. □

Since the $\sigma_i(A_{n,n'})$ are decreasing in i and $n' = \lfloor (1-\delta)n \rfloor$, we see that the above lemma implies that, with probability 1, we have

$$\frac{1}{\sqrt{n}} \sigma_{\lfloor (1-2\delta)n \rfloor}(A_{n,n'}) \geq c\delta$$

for all but finitely many n and some absolute constant $c > 0$. We can generalize this lower bound to handle higher singular values as well, as follows.

LEMMA 6.4. *There exists an absolute constant $c > 0$ such that, with probability* 1, *we have*

$$\frac{1}{\sqrt{n}} \sigma_i(A_{n,n'}) \geq c \frac{n' - i}{n} \tag{6.6}$$

for all but finitely many n and all $1 \leq i \leq (1-2\delta)n$, and similarly with $A_{n,n'}$ replaced by $B_{n,n'}$.

PROOF. Clearly, it suffices to establish the claim for $A_{n,n'}$. Using Proposition 5.1 and the Borel–Cantelli lemma, we see that, with probability 1, we have

$$\operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, \operatorname{span}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{n''}) \right) \geq \frac{1}{2} \sqrt{n - n''}$$

for all but finitely many n, all $1 \leq i \leq n''$ and all $n/2 \leq n'' \leq n'$. Applying Lemma A.4, we conclude that we almost surely have

$$\frac{1}{n} \sum_{i=1}^{n''} \left(\frac{1}{\sqrt{n}} \sigma_i(A_{n,n''}) \right)^{-2} = O\left(\frac{n}{n - n''} \right)$$

for all but finitely many n and all $n/2 \leq n'' \leq n'$. Using the crude bound

$$\sum_{i=1}^{n''} \left(\frac{1}{\sqrt{n}} \sigma_i(A_{n,n''}) \right)^{-2} \geq (n - n'') \left(\frac{1}{\sqrt{n}} \sigma_{2n''-n}(A_{n,n''}) \right)^{-2},$$

we conclude that we almost surely have

$$\frac{1}{\sqrt{n}} \sigma_{2n''-n}(A_{n,n''}) \geq c' \frac{n - n''}{n}$$

for all but finitely many n, all $n/2 \leq n'' \leq n'$ and some absolute constant $c' > 0$. The claim now follows from the Cauchy interlacing property (Lemma A.1). □

REMARK 6.5. If one assumes stronger moment assumptions (e.g., sub-Gaussian) on x, then more precise bounds are known, especially in the $M_n = 0$ case; see [17, 18].

From this lemma, we can now bound the relevant contribution to (6.1), as follows.

LEMMA 6.6. *With probability* 1 *and if* δ *is sufficiently small*, *depending on* ε, *we have*

$$\int_{\delta^2}^{\varepsilon^4} |\log t||d\nu_{n,n'}(t)| = O(\varepsilon) \tag{6.7}$$

for all but finitely many n.

PROOF. By the triangle inequality and symmetry, it suffices to show that, with probability 1, we have

$$\int_{\delta^2}^{\varepsilon^4} |\log t|\, d\mu_{1/n' A_{n,n'} A^*_{n,n'}}(t) = O(\varepsilon)$$

for all but finitely many n. We rewrite the left-hand side as

$$\frac{1}{n}\sum_{i=1}^{n'} f\biggl(\frac{1}{\sqrt{n}}\sigma_i(A_{n,n'})\biggr),$$

where $f(t) := |\log t|\mathbf{I}(\delta^2 \le t^2 \le \varepsilon^4)$. Since f cannot exceed $|\log\delta|$, we see that the contribution of the case $i \ge (1-2\delta)n$ is acceptable if δ is small enough, so it suffices to show that we almost surely have

$$\frac{1}{n}\sum_{1\le i\le(1-2\delta)n} f\biggl(\frac{1}{\sqrt{n}}\sigma_i(A_{n,n'})\biggr) = O(\varepsilon)$$

for all but finitely many n.

By Lemma 6.4, we may assume that n is such that (6.6) holds. As a consequence, we see that the only terms in the above sum which are nonvanishing are those for which $i = (1 - O(\varepsilon^2))n$. However, if we then apply (6.6) and crudely estimate $f(t) \le -\log t$, we obtain the claim. $\square$

6.4. *The contribution of very small* t. Finally, we need to control the contribution when $t \le \delta$.

LEMMA 6.7. *With probability* 1 *and if* δ *is sufficiently small*, *depending on* ε, *we have*

$$\int_{0}^{\delta^2} |\log t||d\nu_{n,n'}(t)| = O(\varepsilon) \tag{6.8}$$

for all but finitely many n.

PROOF. By arguing as in the proof of Lemma 6.6, it suffices to show that we almost surely have

$$\frac{1}{n}\sum_{i=1}^{n'} g\left(\frac{1}{\sqrt{n}}\sigma_i(A_{n,n'})\right) = O(\varepsilon)$$

for all but finitely many n, where $g(t) := |\log t|\mathbf{I}(t^2 \leq \delta^2)$.

By Lemmas 6.3, we may assume n is such that (6.5) holds. On the other hand, if δ is small enough, we have the bound $g(t) \leq \varepsilon t^{-2}$. The claim now follows from (6.5). □

Putting together (6.3), (6.4), (6.7), (6.8), we see that, with probability $1 - O(\varepsilon)$, we have (6.2) for all but finitely many n and so the claim follows.

7. Extensions.

7.1. *Proof of Theorem* 1.13. The theorem in the case of almost sure convergence follows immediately from Theorem 1.5 by conditioning on M_n, so it remains to verify the theorem in the case of convergence in probability.

Let us fix a test function f [as in (1.1)] and a positive ε. By the boundedness in probability of $\frac{1}{n^2}\|M\|_2^2$, we can find a $C = C_\varepsilon$ such that $\mathbf{P}(M_n \in \Omega_n) \geq 1 - \varepsilon$, where

$$\Omega_n := \left\{M \in M_n(\mathbb{C}) : \frac{1}{n^2}\|M\|_2^2 \leq C\right\}.$$

Let M_n^f be the matrix in Ω_n which maximizes[6] the quantity

$$\mathbf{P}\left(\left|\int_{\mathbb{C}} f(z)\, d\mu_{1/\sqrt{n}(M_n^f+X_n)}(z) - \int_{\mathbb{C}} f(z)\, d\mu_{1/\sqrt{n}(M_n^f+Y_n)}(z)\right| \geq \varepsilon\right).$$

Applying Theorem 1.5 to the sequences $M_n^f + X_n$ and $M_n^f + Y_n$, we see that this quantity is $o(1)$.

Theorem 1.13 follows by integrating over all possible values of M_n using the definition of M_n^f, as well as the fact that $\mathbf{P}(\Omega_n) \geq 1 - \varepsilon$, and then letting $\varepsilon \to 0$.

7.2. *Proof of Theorem* 1.14. We first verify the claim for convergence in probability.

Condition (i) of Theorem 2.1 is satisfied thanks to the boundedness in probability of (1.5). In order to complete the proof, one needs to check (ii). Notice that

$$\det\left(\frac{1}{\sqrt{n}}A_n - zI\right) = \det\left(\frac{1}{\sqrt{n}}(K_n^{-1}M_nL_n^{-1} + X_n) - zK_n^{-1}L_n^{-1}\right)\det L_nK_n.$$

[6]If the maximum is not attained, one can instead choose M_n^f to be a matrix which maximizes this quantity to within a factor of two (say).

The term $\det L_n K_n$ also appears in $\det(\frac{1}{\sqrt{n}} B_n - zI)$ and becomes additive (and thus cancels) after taking the logarithm. Therefore, one only needs to show that

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}(K_n^{-1}M_nL_n^{-1}+X_n)-zK_n^{-1}L_n^{-1}\right)\right| \\ -\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}(K_n^{-1}M_nL_n^{-1}+Y_n)-zK_n^{-1}L_n^{-1}\right)\right|$$

converges in probability to zero.

One can obtain this by repeating the proof of Proposition 2.2. The slight change here is that zI is replaced by $zK_n^{-1}L_n^{-1}$, but this has no significant impact, except that we need to show that

$$F_n := \frac{1}{\sqrt{n}}(K_n^{-1}M_nL_n^{-1} - zK_n^{-1}L_n^{-1})$$

satisfies

$$\frac{1}{n^2}\operatorname{trace} F_nF_n^* = \frac{1}{n^2}\|F_n\|_2^2 = O(1)$$

almost surely [in order to guarantee (1.3)]. However, this is a consequence of the boundedness in probability of (1.5).

The proof of the almost sure convergence is established similarly, with the obvious changes (e.g., replacing boundedness in probability with almost sure boundedness). We omit the details.

8. Proof of Theorem 1.15.

We first prove that (ii) implies (i) for almost sure convergence. Let A_n and μ be as in Theorem 1.15. Construct a diagonal matrix B_n' whose diagonal entries are independent samples from μ and let $B_n := \sqrt{n}B_n'$. We wish to invoke Theorem 2.1. We first need to verify the almost sure boundedness of (2.1). The bound for A_n follows from Lemma 1.7 and the bound for B_n follows from the second moment hypothesis on μ and the (strong) law of large numbers. By Theorem 2.1, the problem now reduces to showing that for almost all complex numbers z,

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}A_n - zI\right)\right| - \frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right|$$

converges almost surely to zero. The right-hand side is easy to compute:

$$\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}B_n - zI\right)\right| = \frac{1}{n}\log|\det(B_n' - zI)| = \frac{\sum_{i=1}^n \log|\lambda_i - z|}{n},$$

where λ_i are i.i.d. samples from μ. On the other hand, from Fubini's theorem, we see that $\int_{\mathbb{C}} \log|w-z|\,d\mu(w)$ is locally integrable in z and thus

$$\int_{\mathbb{C}} \log|w-z|\,d\mu(w) < \infty \tag{8.1}$$

for almost every z. If z is such that (8.1) holds, then, by the strong law of large numbers, we see that $\frac{\sum_{i=1}^n \log|\lambda_i - z|}{n}$ converges almost surely to $\int_{\mathbb{C}} \log|w-z|\, d\mu(w)$. This shows that (ii) implies (i) for almost sure convergence. The proof for convergence in probability is identical and is left as an exercise to the reader.

We now show that (iii) implies (ii) for almost sure convergence. Let z be such that (8.1) and (iii) hold. To show (ii), it suffices from (2.3) to show that $\frac{1}{n}\sum_{i=1}^n \log\sigma_i$ converges almost surely to $\int_{\mathbb{C}} \log|w-z|\, d\mu(w)$, where $\sigma_i = \sigma_i(\frac{1}{\sqrt{n}}A_n - zI)$ are the singular values of $\frac{1}{\sqrt{n}}A_n - zI$. On the other hand, from (iii), we already know that $\frac{1}{n}\sum_{i=1}^n \log\sqrt{\sigma_i^2 + \varepsilon_n}$ converges almost surely to $\int_{\mathbb{C}} \log|w-z|\, d\mu(w)$. It thus suffices to show that

$$\frac{1}{n}\sum_{i=1}^n \log\sqrt{\sigma_i^2+\varepsilon_n} - \log\sigma_i \tag{8.2}$$

converges almost surely to zero.

From Lemma 1.7, we know that $\frac{1}{n^2}\|A_n\|_2^2$ is almost surely bounded and so, for each z,

$$\frac{1}{n}\sum_{i=1}^n \sigma_i^2 = \frac{1}{n}\left\|\frac{1}{\sqrt{n}}A_n - zI\right\|_2^2$$

is also almost surely bounded. From this, we easily see that

$$\frac{1}{n}\sum_{1\le i\le n:\,\sigma_i\ge\delta_n} \log\sqrt{\sigma_i^2+\varepsilon_n} - \log\sigma_i$$

converges almost surely to zero for some sequence δ_n (depending on ε_n) converging sufficiently slowly to zero. To conclude the almost sure convergence of (8.2) to zero, it thus suffices to show that

$$\frac{1}{n}\sum_{1\le i\le n:\,\sigma_i\le\delta_n} \log\frac{1}{\sigma_i}$$

converges almost surely to zero. Using Lemma 4.1, we almost surely have $\sup_i \log\frac{1}{\sigma_i} \le O(\log n)$ for all but finitely many n, so it suffices to show that

$$\frac{1}{n}\sum_{1\le i\le n-n^{0.99}:\,\sigma_i<\delta_n} \log\frac{1}{\sigma_i}$$

converges almost surely to zero. To do this, it suffices, by the union bound and the Borel–Cantelli lemma, to show that

$$\mathbf{P}\left(\sigma_{n-i} \le c\frac{i}{n}\right) = O(\exp(-n^{0.01})) \tag{8.3}$$

for all $1 \le i \le n - n^{0.99}$ and some $c > 0$ independent of n.

For this, we argue as in the proof of Lemma 6.4. Fix i. Let A'_n be the matrix formed by the first $n-k$ rows of $A_n - z\sqrt{n}I$ with $k := i/2$ and σ'_j, $1 \leq j \leq n-k$, be the singular values of A'_n (in decreasing order, as usual). By the interlacing law (Lemma A.1) and renormalizing, we have

$$\sigma_{n-i} \geq \frac{1}{\sqrt{n}}\sigma'_{n-i}. \tag{8.4}$$

By Lemma A.4, we have that

$$\sigma_1'^{-2} + \cdots + \sigma_{n-k}'^{-2} = \mathrm{dist}_1^{-2} + \cdots + \mathrm{dist}_{n-k}^{-2},$$

where dist_j is the distance from the jth row of A'_n to the subspace spanned by the remaining rows.

As shown in the proof of Lemma 4.2, with probability $1 - \exp(-n^{-0.01})$, dist_j is bounded from below by $\Omega(\sqrt{k}) = \Omega(\sqrt{i})$ for all j. Thus, with this probability, the right-hand side in the above identity is $O(n/i)$. On the other hand, as the σ'_j are ordered decreasingly, the left-hand side is at least

$$(i-k)\sigma_{n-i}'^{-2} = \frac{i}{2}\sigma_{n-i}'^{-2}.$$

It follows that, with probability $1 - \exp(-n^{-0.01})$,

$$\sigma'_{n-i} = \Omega\left(\frac{i}{\sqrt{n}}\right).$$

This and (8.4) complete the proof of (8.3) and so (8.2) converges almost surely to zero.

As previously observed, the convergence of (8.2) to zero shows that (ii) implies (iii) for almost sure convergence. An inspection of the argument shows the convergence of (8.2) to zero also lets us deduce (iii) from (ii). The claim for convergence in probability follows similarly. To conclude the proof of Theorem 1.15, it thus suffices to show that (i) implies (ii).

We again start with the almost sure convergence case. Assume that (i) holds and let z be such that (8.1) holds. By shifting A by $\sqrt{n}zI$ if necessary, we may take z to be zero. Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of $\frac{1}{\sqrt{n}}A_n$. By (2.3), it suffices to show that $\frac{1}{n}\sum_{j=1}^n \log|\lambda_j|$ converges almost surely to $\int_{\mathbb{C}} \log|w|\, d\mu(w)$. From (3.2), we know that $\frac{1}{n}\sum_{j=1}^n |\lambda_j|^2$ is almost surely bounded. From this and (i), we conclude that $\frac{1}{n}\sum_{j=1}^n \log(|\lambda_j| + \varepsilon)$ converges almost surely to $\int_{\mathbb{C}}(\log|w| + \varepsilon)\, d\mu(w)$ for any fixed $\varepsilon > 0$. Combining this with (8.1) and dominated convergence, we see that $\frac{1}{n}\sum_{j=1}^n \log(|\lambda_j| + \varepsilon_n)$ converges almost surely to $\int_{\mathbb{C}} \log|w|\, d\mu(w)$ for some sequence $\varepsilon_n > 0$ converging sufficiently slowly to zero. It thus suffices to show that

$$\frac{1}{n}\sum_{j=1}^n \log(|\lambda_j| + \varepsilon_n) - \log|\lambda_j|$$

converges almost surely to zero.

By repeating the arguments used to establish the almost sure convergence of (8.2) to zero, it suffices to show that

$$\frac{1}{n} \sum_{1 \le i \le n : |\lambda_i| \le \delta_n} \log \frac{1}{|\lambda_i|}$$

converges almost surely to zero.

Let us order the eigenvalues λ_i so that $|\lambda_1| \ge \cdots \ge |\lambda_n|$. From Lemma 4.1 and (8.3) (and the Borel–Cantelli lemma), we know that we almost surely have

$$\frac{1}{n} \sum_{(1-\kappa)n < i \le n} \log \frac{1}{\sigma_i} \le O\left(\kappa \log \frac{1}{\kappa}\right)$$

for all but finitely many n, for any fixed $0 < \kappa < 1/2$, and hence, by Weyl's comparison inequality (Lemma A.3), we also almost surely have

$$\frac{1}{n} \sum_{(1-\kappa)n < i \le n} \log \frac{1}{|\lambda_i|} \le O\left(\kappa \log \frac{1}{\kappa}\right)$$

for all but finitely many n. Since the left-hand side is bounded from below by $\kappa \log \frac{1}{|\lambda_{\lfloor (1-\kappa)n \rfloor}|}$, we almost surely conclude a lower bound of the form

$$\left|\lambda_{\lfloor (1-\kappa)n \rfloor}\right| \ge \kappa^{O(1)}$$

for all but finitely many n. In particular (by setting δ to be a suitable power of κ), this implies that almost surely

$$\frac{1}{n} \sum_{1 \le i \le n : |\lambda_i| \le \delta} \log \frac{1}{|\lambda_i|} \le O(\delta^c)$$

for all but finitely many n, for any fixed $0 < \delta \ll 1$ and some absolute constant $c > 0$, and the claim follows. The analogous implication for convergence in probability is similar. The proof of Theorem 1.15 is now complete.

APPENDIX A: LINEAR ALGEBRA INEQUALITIES

In this appendix, we record some elementary identities and inequalities regarding the eigenvalues and singular values of matrices.

LEMMA A.1 (Cauchy's interlacing law). *Let A be an $n \times n$ matrix with complex entries and A' be the submatrix formed by the first $m := n - k$ rows thereof. Let $\sigma_1(A) \ge \cdots \ge \sigma_n(A) \ge 0$ denote the singular values of A, and similarly for A'. We then have*

$$\sigma_i(A) \ge \sigma_i(A') \ge \sigma_{i+k}(A)$$

for every $1 \le i \le n - k$.

PROOF. The claim follows easily from the minimax characterization

$$\sigma_i(A) = \sup_{V_i \subset \mathbb{C}^n} \inf_{v \in V_i : \|v\|=1} \|Av_i\|$$

and

$$\sigma_i(A') = \sup_{V_i \subset \mathbb{C}^{n-k}} \inf_{v \in V_i : \|v\|=1} \|Av_i\|$$

of the singular values, where V_i range over i-dimensional complex subspaces. □

LEMMA A.2 (Weyl comparison inequality for second moment). *Let* $A = (a_{ij})_{1 \le i,j \le n} \in M_n(\mathbb{C})$ *have generalized eigenvalues* $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ *and singular values* $\sigma_1(A) \ge \dots \ge \sigma_n(A) \ge 0$. *Then,*

$$\sum_{j=1}^n |\lambda_j|^2 \le \sum_{j=1}^n \sigma_j(A)^2 = \|A\|_2^2 = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2.$$

PROOF. The two equalities here are clear, so it suffices to prove the inequality. By the Jordan normal form, we can write $A = BUB^{-1}$ for some upper triangular U and invertible B. By the QR factorization, we can write $B = QR$ for some orthogonal Q and upper triangular R. We conclude that $A = QVQ^{-1}$ for some upper triangular V. Conjugating by Q, we thus reduce to the case when A is an upper triangular matrix, in which case the eigenvalues are simply the diagonal entries $a_{11}, \dots, a_{nn}$, and the claim clearly follows. □

We also have the following (stronger) variant of the above inequality.

LEMMA A.3 (Weyl comparison inequality for products). *Let*

$$A = (a_{ij})_{1 \le i,j \le n} \in M_n(\mathbb{C})$$

have generalized eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{C}$, *ordered so that* $|\lambda_1| \le \dots \le |\lambda_n|$, *and singular values* $\sigma_1(A) \ge \dots \ge \sigma_n(A) \ge 0$. *We then have*

$$\prod_{j=1}^J |\lambda_j| \le \prod_{j=1}^J \sigma_j(A)$$

and

$$\prod_{j=J}^n \sigma_j(A) \le \prod_{j=J}^n |\lambda_j|$$

for all $0 \le J \le n$.

PROOF. It suffices to prove the former claim, as the latter then follows from (2.3). By arguing as in Lemma A.2, we may assume that A is upper triangular so that the diagonal entries are some permutation of $\lambda_1, \dots, \lambda_n$. Consider the symmetric minor A' of A formed by the rows and columns corresponding to the entries $\lambda_1, \dots, \lambda_J$. The determinant of this matrix is then $\lambda_1 \cdots \lambda_J$ and thus, by (2.3), we have

$$\prod_{j=1}^{J} \sigma_j(A') = \prod_{j=1}^{J} |\lambda_j|.$$

The claim then follows from the Cauchy interlacing inequality (Lemma A.1). $\square$

We now record a useful identity for the *negative* second moment of a rectangular matrix.

LEMMA A.4 (Negative second moment). *Let $1 \le n' \le n$ and let A be a full rank $n' \times n$ matrix with singular values $\sigma_1(A) \ge \cdots \ge \sigma_{n'}(A) > 0$ and rows $X_1, \dots, X_{n'} \in \mathbb{C}^n$. For each $1 \le i \le n'$, let W_i be the hyperplane generated by the $n' - 1$ rows $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{n'}$. Then,*

$$\sum_{j=1}^{n'} \sigma_j(A)^{-2} = \sum_{j=1}^{n'} \operatorname{dist}(X_j, W_j)^{-2}.$$

PROOF. Observe that the $n' \times n'$ matrix $(AA^*)^{-1}$ has eigenvalues

$$\sigma_1(A)^{-2}, \dots, \sigma_{n'}(A)^{-2}.$$

Taking traces, we conclude that

$$\sum_{j=1}^{n'} \sigma_j(A)^{-2} = \sum_{j=1}^{n'} (AA^*)^{-1} e_j \cdot e_j,$$

where $e_1, \dots, e_{n'}$ is the standard basis of $\mathbb{C}^{n'}$. However, if $v_j := (AA^*)^{-1} e_j = (v_{j,1}, \dots, v_{j,n'})$, then $A^* v_j = v_{j,1} X_1 + \cdots + v_{j,n'} X_{n'}$ is orthogonal to $A^* e_i = X_i$ for $i \ne j$ (and thus orthogonal to W_j) and has an inner product of 1 with $A^* e_j = X_j$. Taking inner products of $A^* v_j$ with the orthogonal projection of X_j to W_j, we conclude that

$$v_{j,j} \operatorname{dist}(X_j, W_j)^2 = 1.$$

Since $v_{j,j} = v_j \cdot e_j = (AA^*)^{-1} e_j \cdot e_j$, the claim follows. $\square$

APPENDIX B: A RESULT OF DOZIER AND SILVERSTEIN

Here, we reproduce Theorem 1.1 of [3], which we used in the end of Section 6.

THEOREM B.1 ([3], Theorem 1.1). *Let c be a positive constant and x be a random variable with variance one. Let X_n be an $n \times r$ random matrix whose entries are i.i.d. copies of x, where $r = (c + o(1))n$. Let M_n be a random $n \times r$ matrix independent from X_n such that the ESD of $M_n M_n^*$ converges to a limiting distribution H. Define $C_n := \frac{c}{n}(M_n + X_n)(M_n + X_n)^*$. The ESD of C_n then converges almost surely* (*and hence also in probability*) *to a limiting distribution F, whose Stieljes transform $m(z) := \int \frac{1}{\lambda - z}\, dF(\lambda)$ satisfies the integral equation*

$$m = \int \frac{dH(t)}{t/(1+cm) - (1+cm)z + (1-c)} \tag{B.1}$$

for any $z \in \mathbb{C}$.

REMARK B.2. The theorem still holds if we restrict the size n of the matrices to an infinite subsequence $n_1 < n_2 < \cdots$ of positive integers. One can show this by, for example, artificially filling in the missing indices or repeating the proof of Theorem B.1 under this restriction.

REMARK B.3. In (B.1), H appears, but the actual definition of M_n is irrelevant. Thus, one can conclude that if M_n and M'_n are such that the ESDs of $M_n M_n^*$ and $M'_n M'^*_n$ tend to the same limit, then the ESDs of $\frac{c}{n}(M_n + X_n)(M_n + X_n)^*$ and $\frac{c}{n}(M'_n + X_n)(M'_n + X_n)^*$ also tend to the same limit.

REMARK B.4. It was mentioned by Speicher [20] and also Krishnapur (private communication) that Theorem B.1 can be proven using free probability, which is different from the approach in [3].

APPENDIX C: USING A HERMITIAN INVARIANCE PRINCIPLE (BY MANJUNATH KRISHNAPUR)

The authors have shown invariance principles for ESDs of several non-Hermitian matrix models. As in earlier papers, the proof goes through Hermitian matrices, but does not need rates of convergence of the Hermitian ESDs, thanks to new ideas such as Lemma 4.2. However, because of the use of Theorem B.1, it may appear that a limiting result for the associated Hermitian matrices is necessary to carry the program through. In this appendix, we point out how one may obtain a weak invariance principle for ESDs of non-Hermitian matrices by using an invariance principle for Hermitian matrices due to Chatterjee [4], in cases where a convergence result such as Theorem B.1 is not available. As mentioned earlier, other parts of the proof do not require the entries to be i.i.d. Thus, as a consequence, we can obtain a weak invariance principle for a random matrix model with independent, but not identically distributed, entries.

We need the following definition from [25], Section 2.

DEFINITION C.1 (Controlled second moment). Let $\kappa \geq 1$. A complex random variable x is said to have a *κ-controlled second moment* if one has the upper bound

$$\mathbf{E}|x|^2 \leq \kappa$$

(in particular, $|\mathbf{E}x| \leq \kappa^{1/2}$) and the lower bound

$$\mathbf{E}\,\mathrm{Re}(zx - w)^2 \mathbf{I}(|x| \leq \kappa) \geq \frac{1}{\kappa} \mathrm{Re}(z)^2 \tag{C.1}$$

for all complex numbers z, w.

EXAMPLE. The Bernoulli random variable $[\mathbf{P}(x = +1) = \mathbf{P}(x = -1) = 1/2]$ has 1-controlled second moment. The condition (C.1) asserts, in particular, that x has variance of at least $\frac{1}{\kappa}$, but also asserts that a significant portion of this variance occurs inside the event $|x| \leq \kappa$, and also contains some more technical phase information about the covariance matrix of $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$.

THEOREM C.2. *Let $M_n = (\mu_{i,j}^{(n)})_{i,j \leq n}$ and $C_n = (\sigma_{i,j}^{(n)})_{i,j \leq n}$ be constant (i.e., deterministic) matrices satisfying*:

1. $\sup_n n^{-2} \|M_n\|_2^2 < \infty$;
2. $a \leq \sigma_{i,j}^{(n)} \leq b$ *for all* n, i, j, *for some* $0 < a < b < \infty$.

Given a matrix $\mathbf{X} = (x_{i,j})_{i,j \leq n}$, set

$$A_n(\mathbf{X}) = \frac{1}{\sqrt{n}}(M_n + C_n \cdot \mathbf{X}) = \frac{1}{\sqrt{n}}\big(\mu_{i,j}^{(n)} + \sigma_{i,j}^{(n)} x_{i,j}\big)_{i,j \leq n}$$

(*here*, "$\cdot$" *denotes Hadamard product*).

Now, suppose that $x_{i,j}^{(n)}$ are independent complex-valued random variables with $\mathbf{E}[x_{i,j}^{(n)}] = 0$ and $\mathbf{E}[|x_{i,j}^{(n)}|^2] = 1$, and that $y_{i,j}^{(n)}$ are independent random variables, also having zero mean and unit variance.

Assume, furthermore, that both $x_{ij}^{(n)}$ and $y_{ij}^{(n)}$ have κ-controlled second moment for some constant $\kappa > 0$.

Also, assume Pastur's condition,

$$\frac{1}{n^2} \sum_{i,j=1}^{n} \mathbf{E}\big[|x_{i,j}^{(n)}|^2 \mathbf{I}|x_{i,j}^{(n)}| \geq \epsilon\sqrt{n}\big] \longrightarrow 0 \qquad \textit{for all } \epsilon > 0, \tag{C.2}$$

and the same for $\mathbf{Y}$ *in place of* $\mathbf{X}$. *Then*,

$$\mu_{A_n(\mathbf{X})} - \mu_{A_n(\mathbf{Y})} \to 0$$

in the sense of probability.

Some remarks follow.

1. If we assume that $x_{i,j}^{(n)}$ are i.i.d. and $y_{i,j}^{(n)}$ are i.i.d., then Pastur's condition is obviously satisfied. Further, the condition of κ-controlled second moment is also not necessary (see the first step in the proof sketch).
2. Although the weak invariance principle in the paper uses only subsequential limits (see Remark 6.2), it does use Theorem B.1 to say that subsequential limits are the same for $\mathbf{X}$ as for $\mathbf{Y}$. Hence, we need some changes in the proof in order to establish Theorem C.2, which we achieve in this appendix.
3. This highlights the important new ideas of the paper, such as Lemma 4.2, which eliminate the need for rates of convergence of ESDs of the Hermitian matrices $(A_n - zI)^*(A_n - zI)$. This is unlike all earlier papers in the subject that followed Bai's approach and required such rates (e.g., [1, 9, 14, 25]). The need for rates made it impossible to use the invariance principle for Hermitian matrices as we shall do now.
4. Take $C_n = J$ (all-ones matrix) and $M_n = 0$. Pastur's condition (C.2) then implies almost sure convergence of the ESD of $A_n(\mathbf{X})^* A_n(\mathbf{X})$ (see [2], Theorem 3.9). For general C_n, since we use Chatterjee's invariance principle, which assumes Pastur's condition but only gives weak invariance, we are also able to assert only weak invariance for the non-Hermitian ESDs. Thus, there is some room for improvement here, namely, to strengthen the conclusion of Theorem C.2 to almost sure convergence.
5. Does the ESD of $A_n(\mathbf{X})$ converge? Perhaps so, provided the singular values of $C_n - zI$ have a limiting measure for every z. In [11], we have discussed some easy-to-check sufficient conditions on C_n which imply convergence.

The following lemma is a "Wishart" analogue of the computations in Section 2 of [4], which considers Wigner matrices. As in that paper, the idea is to consider the Stieltjes transform of the ESD of $A_n(\mathbf{X})^* A_n(\mathbf{X})$ as a function of $\mathbf{X}$. However, a slight twist is needed, as compared to Wigner matrices, because the entries of $A_n(\mathbf{X})^* A_n(\mathbf{X})$ are quadratic in $\mathbf{X}$, whereas the invariance principle we invoke requires bounds on the sup-norm of derivatives of the Stieltjes transform.

LEMMA C.3. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be as in Theorem* C.2. *Let* $\nu_n^{\mathbf{X}}$ *and* $\nu_n^{\mathbf{Y}}$ *be the ESDs of* $A_n(\mathbf{X})^* A_n(\mathbf{X})$ *and* $A_n(\mathbf{Y})^* A_n(\mathbf{Y})$, *respectively. Then*, $\nu_n^{\mathbf{X}} - \nu_n^{\mathbf{Y}} \to 0$ *weakly as* $n \to \infty$.

PROOF. Let

$$H_n(\mathbf{X}) = \begin{bmatrix} 0 & A_n(\mathbf{X}) \\ A_n(\mathbf{X})^* & 0 \end{bmatrix}$$

have ESD $\theta_n^{\mathbf{X}}$. The eigenvalues of $H_n(\mathbf{X})$ are exactly the positive and negative square roots of the eigenvalues of $A_n(\mathbf{X})^* A_n(\mathbf{X})$. Thus, we must show that $\theta_n^{\mathbf{X}} - \theta_n^{\mathbf{Y}} \to 0$ weakly, in probability. Fix any α in the upper half-plane and let $f(\mathbf{X}) := \frac{1}{2n}\operatorname{Tr}(H_n(\mathbf{X}) - \alpha I)^{-1}$. The proof is complete if we show that $\mathbf{E}[f(\mathbf{X})] - \mathbf{E}[f(\mathbf{Y})] \to 0$ for any α with $\operatorname{Im}\{\alpha\} > 0$. This can be done by following the same calculations as in [4]. It works because the entries of $H_n(\mathbf{X})$ are linear in $\mathbf{X}$ and hence the first partial derivative of H_n with respect to any $x_{i,j}$ is a constant matrix. One must also use the upper bound on $\sigma_{i,j}$ to bound the derivatives of f. □

REMARK. Obviously, the same conclusion holds for $A_n - zI$ by simply absorbing zI into M_n.

PROOF OF THEOREM C.2. The conditions on M_n and C_n show that the first condition of Theorem 2.1 is satisfied [where the two matrices A_n and B_n are now $A_n(\mathbf{X})$ and $A_n(\mathbf{Y})$, resp.].

Thus, we only need to show an analog of Proposition 2.2 (only the weak part). We sketch the modifications needed.

1. Lemma 4.1 can be proven under independence and κ-controlled second moment without the i.i.d. assumption (see [25], Theorem 2.5). If we make the i.i.d. assumption, then Lemma 4.1 is itself applicable, which explains the first remark after the statement of the theorem.

   The upper bounds on singular values in (4.4) are very general and hold in our setting for the same reasons. Hence, we reduce to Lemmas 4.2 and 4.3, as in the paper.
2. The high-dimensional contribution (analog of Lemma 4.2) is proved in almost the same way. In the proof of the lower tail bound (Proposition 5.1), we use the bounds on $\sigma_{i,j}^{(n)}$ appropriately. In particular, we get a lower bound of $a^2(n-d)$ for the second moment of $\operatorname{dist}(X, W)$ in Lemma 5.3 and, in applying Theorem 5.2, we get a Lipschitz constant of b for $F(X) = \operatorname{dist}(X, W)$.
3. In the low-dimensional contribution (Lemma 4.3), the calculations in Sections 6.1, 6.3 and 6.4 are exactly as before (in Section 6.3, we use the concentration result already outlined in the previous step).
4. That leaves Section 6.2, which is the only step that is differently handled. Here, we apply Lemma C.3 instead of quoting Theorem B.1. □

Acknowledgments. The authors would like to thank M. Krishnapur for useful discussions and his careful reading of an early draft, and Ken Miller, Ricky and Weiyu for further corrections. We would also like to thank P. Matchett Wood for providing the figures in the Introduction.

REFERENCES

[1] BAI, Z. D. (1997). Circular law. *Ann. Probab.* **25** 494–529. MR1428519

[2] BAI, Z. D. and SILVERSTEIN, J. (2006). *Spectral Analysis of Large Dimensional Random Matrices*. *Mathematics Monograph Series* **2**. Science Press, Beijing.

[3] DOZIER, R. B. and SILVERSTEIN, J. W. (2007). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.* **98** 678–694. MR2322123

[4] CHATTERJEE, S. (2005). A simple invariance principle. Available at arXiv:math/0508213.

[5] CHAFAI, D. (2008). Circular law for non-central random matrices. Preprint.

[6] EDELMAN, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** 543–560. MR964668

[7] GIRKO, V. L. (1984). The circular law. *Theory Probab. Appl.* **29** 694–706. MR0773436

[8] GIRKO, V. L. (2004). The strong circular law. Twenty years later. II. *Random Oper. Stochastic Equations* **12** 255–312. MR2085255

[9] GÖTZE, F. and TIKHOMIROV, A. N. (2007). On the circular law. Preprint.

[10] GÖTZE, F. and TIKHOMIROV, A. N. (2007). The circular law for random matrices. Preprint.

[11] KRISHNAPOUR, M. and VU, V. Manuscript in preparation.

[12] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[13] MEHTA, M. L. (1967). *Random Matrices and the Statistical Theory of Energy Levels*. Academic Press, New York. MR0220494

[14] PAN, G. and ZHOU, W. (2010). Circular law, extreme singular values and potential theory. *J. Multivariate Anal.* **101** 645–656. MR2575411

[15] PASTUR, L. A. (1972). The spectrum of random matrices. *Teoret. Mat. Fiz.* **10** 102–112. MR0475502

[16] RUDELSON, M. (2008). Invertibility of random matrices: Norm of the inverse. *Ann. of Math.* (2) **168** 575–600. MR2434885

[17] RUDELSON, M. and VERSHYNIN, R. (2008). The least singular value of a random square matrix is $O(n^{-1/2})$. *C. R. Math. Acad. Sci. Paris* **346** 893–896. MR2441928

[18] RUDELSON, M. and VERSHYNIN, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707–1739. MR2569075

[19] RUDELSON, M. and VERSHYNIN, R. (2010). The Littlewood-Offord problem and the condition number of random matrices. *Adv. Math.* **218** 600–633.

[20] SPEICHER, R. Survey in preparation.

[21] TAO, T. and VU, V. (2006). On random ± 1 matrices: Singularity and determinant. *Random Structures Algorithms* **28** 1–23. MR2187480

[22] TAO, T. and VU, V. (2006). *Additive Combinatorics*. *Cambridge Studies in Advanced Mathematics* **105**. Cambridge Univ. Press, Cambridge. MR2289012

[23] TAO, T. and VU, V. H. (2009). Inverse Littlewood–Offord theorems and the condition number of random discrete matrices. *Ann. of Math.* (2) **169** 595–632. MR2480613

[24] TAO, T. and VU, V. (2007). The condition number of a randomly perturbed matrix. In *STOC'*07*—Proceedings of the* 39*th Annual ACM Symposium on Theory of Computing* 248–255. ACM, New York. MR2402448

[25] TAO, T. and VU, V. (2008). Random matrices: The circular law. *Commun. Contemp. Math.* **10** 261–307. MR2409368

[26] TAO, T. and VU, V. (2010). Random matrices: The distribution of the smallest singular values. *Geom. Funct. Anal.* **20** 260–297.

[27] WIGNER, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* (2) **67** 325–327. MR0095527

T. TAO
DEPARTMENT OF MATHEMATICS
UCLA
LOS ANGELES, CALIFORNIA 90095-1555
USA
E-MAIL: tao@math.ucla.edu

V. VU
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
PISCATAWAY NEW JERSEY 08854-8019
USA
E-MAIL: vanvu@math.rutgers.edu

M. KRISHNAPUR
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE
BANGALORE 560012
INDIA
E-MAIL: manju@math.iisc.ernet.in