UNIVERSALITY AND THE CIRCULAR LAW FOR SPARSE RANDOM MATRICES

BY PHILIP MATCHETT WOOD[1]

Stanford University

The universality phenomenon asserts that the distribution of the eigenvalues of random matrix with i.i.d. zero mean, unit variance entries does not depend on the underlying structure of the random entries. For example, a plot of the eigenvalues of a random sign matrix, where each entry is $+1$ or -1 with equal probability, looks the same as an analogous plot of the eigenvalues of a random matrix where each entry is complex Gaussian with zero mean and unit variance. In the current paper, we prove a universality result for sparse random n by n matrices where each entry is nonzero with probability $1/n^{1-\alpha}$ where $0 < \alpha \le 1$ is any constant. One consequence of the sparse universality principle is that the circular law holds for sparse random matrices so long as the entries have zero mean and unit variance, which is the most general result for sparse random matrices to date.

1. Introduction. Given an n by n complex matrix A, we define the *empirical spectral distribution* (which we will abbreviate *ESD*), to be the following discrete probability measure on $\mathbb{C}$:

$$\mu_A(z) := \frac{1}{n}|\{1 \le i \le n : \text{Re}(\lambda_i) \le \text{Re}(z) \text{ and } \text{Im}(\lambda_i) \le \text{Im}(z)\}|,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A with multiplicity. In this paper, we focus on the case where A is chosen from a probability distribution on $\mathbb{M}_n(\mathbb{C})$, the set of all n by n complex matrices, and thus μ_A is a randomly generated discrete probability measure on $\mathbb{C}$.

1.1. *Background*: *Universality and the circular law.* Suppose that A_n is an n by n matrix with i.i.d. random entries, each having zero mean and unit variance. The distribution of the eigenvalues of $(1/\sqrt{n})A_n$ approaches the uniform distribution on the unit disk as n goes to infinity, a phenomenon known as the circular law. The nonsparse circular law has been proven in many special cases by many authors, including Mehta [20] (Gaussian case), Girko [13, 14], Edelman [11] (real Gaussian case), Bai [2] and Bai and Silverstein [1] [continuous case with bounded $(2 + \delta)$th moment, for $\delta > 0$], Götze and Tikhomirov [15] (sub-Gaussian case) and

[16] [bounded $(2 + \delta)$th moment, for $\delta > 0$], Pan and Zhao [23] (bounded 4th moment) and Tao and Vu [32] [bounded $(2 + \delta)$th moment, for $\delta > 0$]. The following, due to Tao and Vu [34], Theorem 1.10, is the current best result, requiring only zero mean and unit variance (see also [33]).

THEOREM 1.1 (Nonsparse circular law ([34], Theorem 1.10)). *Let X_n be the n by 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.*

There has also been recent interest in generalizations of the circular law to random matrix ensembles where finite variance is relaxed (see [3]) and where some dependence among the entries is allowed (see [4, 6]).

Proving convergence in the almost sure sense is, in general, harder than proving convergence in probability, and in the current paper, we will focus exclusively on convergence in probability. See Section 1.4 toward the end of the Introduction for a description of convergence in probability and in the almost sure sense for the current context.

In [34], Tao and Vu ask the following natural question: what analog of Theorem 1.1 is possible in the case where the matrix is sparse, where entries become more likely to be zero as n increases, instead of entries having the same distribution for all n? One goal of the current paper is to provide an answer to this question in the form of Theorem 1.6 (see below), which proves the circular law for sparse random matrices with i.i.d. entries. In Figure 1, parts (b) and (d) give examples of the nonsparse circular law for Bernoulli and Gaussian random variables, and parts (a) and (c) give examples of the sparse circular law for Bernoulli and Gaussian random variables.

The mathematical literature studying the eigenvalues of sparse random matrices is distinctly smaller than that for nonsparse random matrices (there are, however, some nonrigorous approaches from a physics perspective, e.g., [12]). Most authors in mathematics and physics have focused on studying the eigenvalues in the symmetric case, including [9, 18, 21, 22, 24, 27–30]. There has been, however, some recent and notable progress for nonsymmetric sparse random matrices. Götze and Tikhomirov [15, 16] provide sparse versions for their proofs of the circular law with some extra conditions. In [15] they use the additional assumptions that the entries are sub-Gaussian and that each entry is zero with probability ρ_n where $\rho_n n^4 \to \infty$ as $n \to \infty$, and in [16] they use the additional assumption that the entries have bounded $(2 + \delta)$th moment. The strongest result in the literature for nonsymmetric sparse random matrices is due to Tao and Vu [32] who in 2008 proved a sparse version of the circular law with the assumption of bounded $(2 + \delta)$th moment (note that [32] proves almost sure convergence, rather than convergence in probability as shown by [15, 16]).
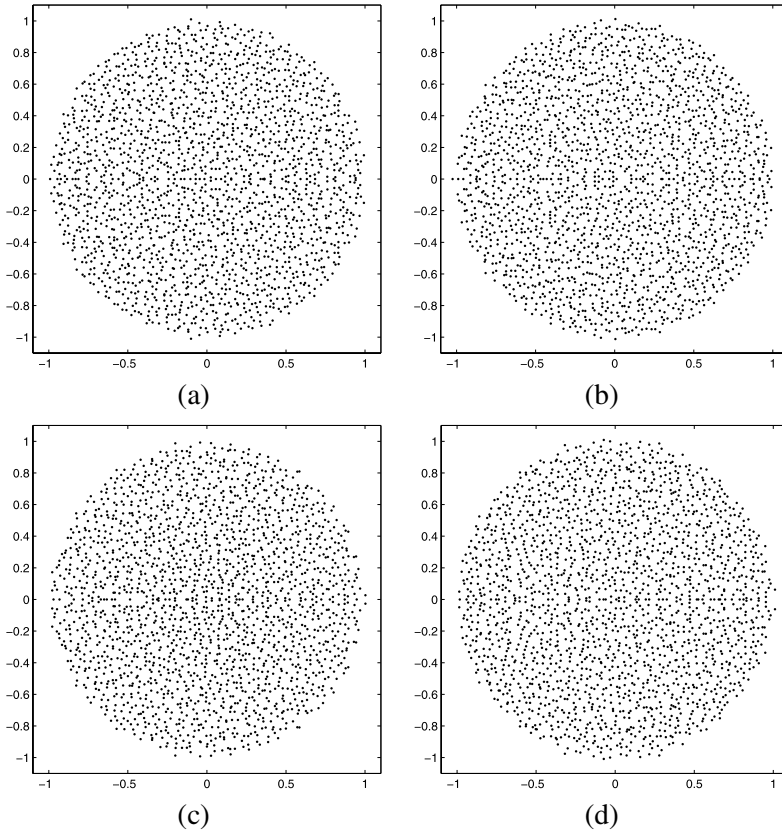
FIG. 1.   *The four figures above illustrate that the circular law holds for Bernoulli and Gaussian random matrix ensembles in both the sparse and nonsparse cases. Each plot is of the eigenvalues of a 2,000 by 2,000 random matrix with i.i.d. entries. In the first column [parts (a) and (c)] the matrices are sparse with parameter $\alpha = 0.4$, which means each entry is zero with probability $1 - \frac{1}{n^{0.6}}$, and in the second column [parts (b) and (d)] the matrices are not sparse (i.e., $\alpha = 1$). In the first row, both matrix ensembles are Bernoulli, so each nonzero entry is equally likely to be -1 or 1, and in the second row, the ensembles are Gaussian, so the nonzero entries are drawn from a Gaussian distribution with mean zero and variance one.*

THEOREM 1.2 ([32], Theorem 1.3).   *Let $\alpha > 0$ and $\delta > 0$ be arbitrary positive constants. Assume that x is a complex random variable with zero mean and finite $(2 + \delta)$th moment. Set $\rho = n^{-1+\alpha}$ and let A_n be the matrix with each entry an i.i.d. copy of $\frac{1}{\sqrt{\rho}}\mathbb{I}_\rho x$, where $\mathbb{I}_\rho$ is a random variable independent of x, and $\mathbb{I}_\rho$ takes the value 1 with probability ρ and the value 0 with probability $1 - \rho$. Let $\mu_{(1/(\sigma\sqrt{n}))A_n}$ be the ESD of $\frac{1}{\sigma\sqrt{n}}A_n$, where σ^2 is, as usual, the variance of x. Then $\mu_{(1/\sqrt{\sigma n})A_n}$ converges in the almost sure sense to the uniform distribution μ_∞ over the unit disk as n tends to infinity.*

In this paper, we prove a sparse circular law without the bounded $(2 + \delta)$th moment condition, with our work being motivated by the proof in [34] of the (nonsparse) circular law in the general zero mean, unit variance case.

There has been much recent interest in demonstrating universal behavior for the eigenvalues of various types of random matrices. The following theorem is a fundamental result from [34]. For a matrix $A = (a_{ij})_{1 \le i, j \le n}$, we will use $\|A\|_2$ to denote the Hilbert–Schmidt norm, which is defined by $\|A\|_2 = \text{trace } AA^* = (\sum_{1 \le i, j \le n} |a_{ij}|^2)^{1/2}$.

THEOREM 1.3 (Universality principle [34]). *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*

$$(1) \qquad \sup_n \frac{1}{n^2} \|M_n\|_2^2 < \infty.$$

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.

The universality principle as proven in [34], Theorem 1.5, also includes an additional hypothesis under which $\mu_{(1/\sqrt{n})A_n} - \mu_{(1/\sqrt{n})B_n}$ converges almost surely to zero (see [34] for details). In [34], Tao and Vu suggest the project of extending their universality principle for random matrices to the case of sparse random matrices. In this paper, we will follow the program developed in [34] and prove a universality principle for sparse random matrices.

1.2. *New results for sparse random matrices.* We begin by defining the type of sparse matrix ensemble that we will consider in this paper.

DEFINITION 1.4 (Sparse matrix ensemble). Let $0 < \alpha \le 1$ be a constant, and let $\mathbb{I}_\rho$ be the random variable taking the value 1 with probability $\rho := n^{-1+\alpha}$ and the value 0 with probability $1 - \rho$. Let x be a complex random variable that is independent of $\mathbb{I}_\rho$. The *n by n sparse matrix ensemble for x with parameter α* is defined to be the matrix X_n where each entry is an i.i.d. copy of $\frac{1}{\sqrt{\rho}}\mathbb{I}_\rho x$.

The main result of the current paper is the following:

THEOREM 1.5 (Sparse universality principle). *Let $0 < \alpha \le 1$ be a constant, and let x be a random variable with mean zero and variance one. Let X_n be the n by n sparse matrix ensemble for x with parameter α, and let Y_n be the n by n matrix having i.i.d. copies of x for each entry (in particular, Y_n is not sparse). For each n, let M_n be a deterministic n by n matrix such that*

$$(2) \qquad \sup_n \frac{1}{n^2} \|M_n\|_2^2 < \infty,$$

and 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.*

Figure 2 gives an illustration of Theorem 1.5 with nontrivial M_n for sparse and nonsparse Bernoulli and Gaussian ensembles. In [25], a method is given for
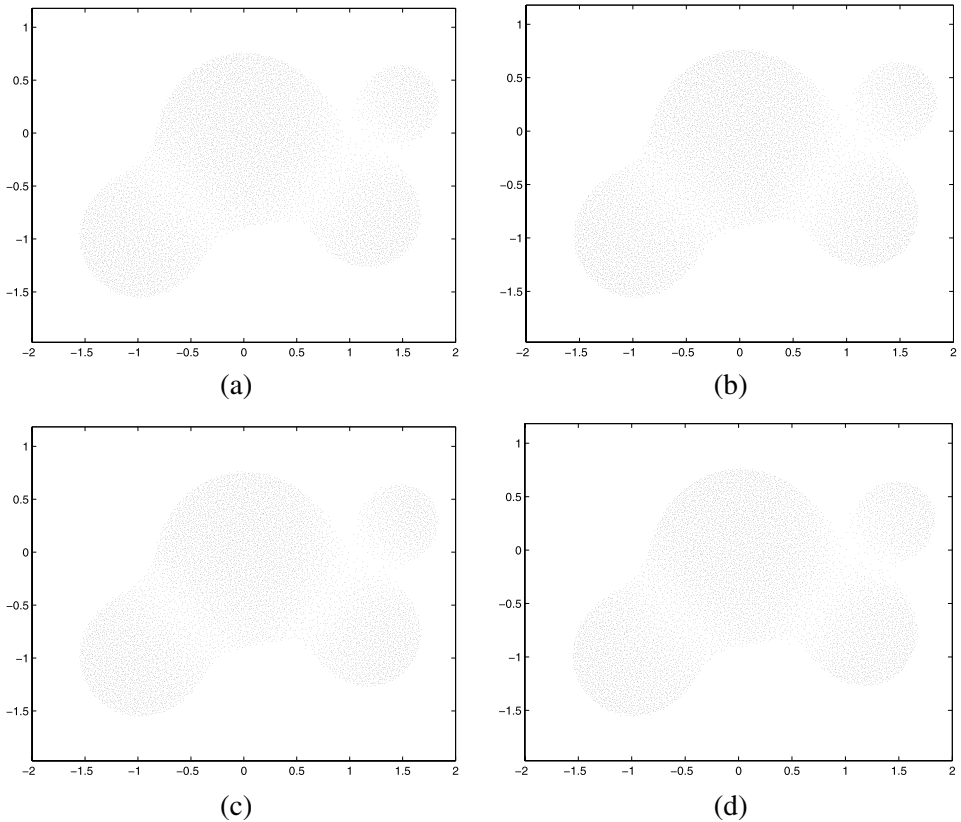

FIG. 2. *The four plots above illustrate that the universality principle holds for Bernoulli and Gaussian random matrix ensembles in both the sparse and nonsparse cases. Each plot is of the eigenvalues of a* 10,000 *by* 10,000 *random matrix with of the form* $M_n + X_n$, *where* M_n *is a fixed, nonrandom matrix, and* X_n *contains i.i.d. entries. For each of the four plots,* $\frac{1}{\sqrt{n}}M_n$ *is the diagonal matrix with the first* $\lfloor n/4 \rfloor$ *diagonal entries equal to* $-1 - \sqrt{-1}$, *the next* $\lfloor n/6 \rfloor$ *diagonal entries equal to* $1.2 - 0.8\sqrt{-1}$, *the next* $n/12$ *diagonal entries equal to* $1.5 + 0.3\sqrt{-1}$ *and the remaining entries equal to zero. In the first column [parts* (a) *and* (c)*] the matrices* X_n *are sparse with parameter* $\alpha = 0.5$, *which means each entry is zero with probability* $1 - \frac{1}{n^{0.5}}$, *and in the second column [parts* (b) *and* (d)*] the matrices* X_n *are not sparse (i.e.,* $\alpha = 1$*). In the first row, both matrix ensembles are Bernoulli, so each nonzero entry of* X_n *is equally likely to be* -1 *or* 1, *and in the second row, the ensembles are Gaussian, so the nonzero entries of* X_n *are drawn from a Gaussian distribution with mean zero and variance one.*

predicting the eigenvalue distributions of a random matrix plus a deterministic matrix and also of a random matrix multiplied by a deterministic matrix.

Relating the sparse case to the nonsparse case in the above theorem is quite useful, since many results are known for random matrices with nonsparse i.i.d. entries, including a number of results in [34]. One of the motivating consequences of Theorem 1.5 is the following result, which is a combination of Theorems 1.5 and 1.1, the nonsparse circular law proven in [34].

THEOREM 1.6 (Sparse circular law). *Let $0 < \alpha \le 1$ be a constant, and let x be a random complex variable with mean zero and variance one. Let X_n be the sparse matrix ensemble for x with parameter α. Then the ESD for $\frac{1}{\sqrt{n}} X_n$ converges in probability to the uniform distribution on the unit disk.*

An illustration of Theorem 1.6 appears in Figure 1. Note that the sparse circular law (Theorem 1.6) does not hold when $\alpha = 0$, since the probability of a row of all zeroes approaches a constant as $n \to \infty$, and thus with probability tending to 1 as $n \to \infty$, a constant fraction of the rows contain all zeroes. Reasoning in analogy with the Hermitian case, where Wigner's semicircle law holds so long as $n\rho \to \infty$ (see [36]), it seems possible that one might be able to prove the circular law in the case where $\rho = \frac{\log n}{n}$ (see [26] for further evidence). One might also consider analogs of other models of sparseness that have been used in the Hermitian case; for example, see [8, 35].

In the nonsparse case, Tao, Vu and Krishnapur [34] also give a number of extensions and generalizations, one of which is the circular law for shifted matrices, including the case where the entries of a random matrix have constant, nonzero mean.

THEOREM 1.7 (Nonsparse circular law for shifted matrices ([34], Corollary 1.12)). *Let X_n be the n by n random matrix whose entries are i.i.d. complex random variables with mean 0 and variance 1, and let M_n be a deterministic matrix with rank o(n) and obeying inequality (1). Let $A_n := M_n + X_n$. Then the ESD of $\frac{1}{\sqrt{n}} A_n$ converges (both in probability and in the almost sure sense) to the uniform distribution on the unit disk.*

Because Theorem 1.7 applies to nonsparse matrices of the form $M_n + X_n$, it can be directly combined with the sparse universality principle of Theorem 1.5 to yield the following result:

THEOREM 1.8 (Sparse circular law for shifted matrices). *Let $0 < \alpha \le 1$ be a constant, and let x be a complex random variable with mean 0 and variance 1. Let X_n be the n by n sparse random matrix ensemble with parameter α, let*

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

An example of Theorems 1.7 and 1.8 appears in Figure 3.
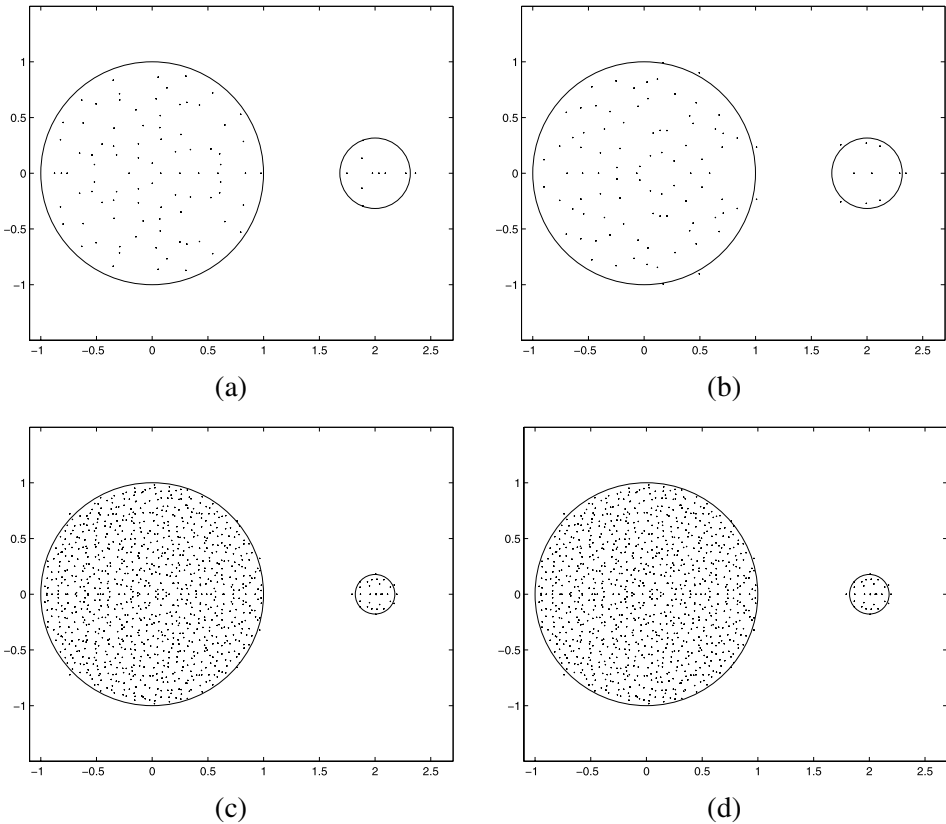

(a)

(b)

(c)

(d)

FIG. 3. *These six figures illustrate that the circular law holds for shifted sparse Bernoulli and shifted nonsparse Bernoulli random matrix ensembles. Each plot is of the eigenvalues of an n by n (with n as specified) random matrix of the form* $M_n + X_n$, *where* M_n *is a nonrandom diagonal matrix with the first* $\lfloor \sqrt{n} \rfloor$ *diagonal entries equal to* $2\sqrt{n}$ *and the remaining entries equal to zero, and* X_n *contains i.i.d. random entries. In the first column [parts* (a) $n = 100$, (c) $n = 1{,}000$ *and* (e) $n = 10{,}000$] *the matrices are sparse with parameter* $\alpha = 0.4$, *which means each entry is zero with probability* $1 - \frac{1}{n^{0.6}}$, *and in the second column [parts* (b) $n = 100$, (d) $n = 1{,}000$ *and* (f) $n = 10{,}000$] *the matrices are not sparse (i.e.,* $\alpha = 1$). *The matrix ensembles are Bernoulli, so each nonzero entry is equally likely to be* -1 *or* 1. *As n increases, the ESDs in both the sparse and nonsparse cases approach the uniform distribution on the unit disk. Empirically, the small circle on the right, which has roughly* $\sqrt{n}$ *eigenvalues in and near it, shrinks until its contribution to the ESD is negligible (as drawn, the small circle has radius* $n^{-1/4}$).*

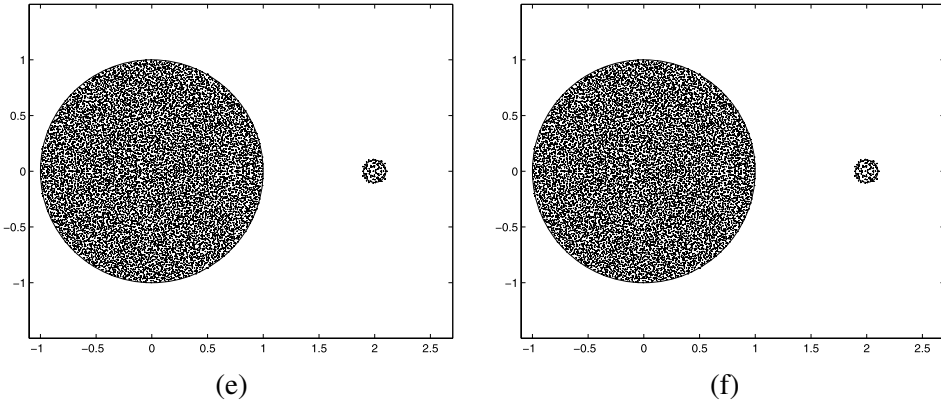(e)   (f)

FIG. 3.   (*Continued*).

The simple lemma below is an essential component for adapting arguments from [34] to the sparse case, and illustrates a critical transition that occurs when $\alpha = 0$.

LEMMA 1.9.   *Let ξ be a complex random variable such that $\mathbb{E}|\xi| < \infty$. Let X be a sparse version of ξ, namely $X := \mathbb{I}_\rho \xi / \rho$, where $\rho = n^{-1+\alpha}$, where $0 < \alpha \leq 1$ is a constant. Then*

$$\mathbb{E}\big(\big|\mathbb{1}_{\{X > n^{1-\alpha/2}\}} X\big|\big) \to 0$$

as $n \to \infty$.

PROOF.   The key steps to this proof are using independence of $\mathbb{I}_\rho$ and ξ, and applying monotone convergence. We compute

$$\mathbb{E}\big(\big|\mathbb{1}_{\{|X| > n^{1-\alpha/2}\}} X\big|\big) = \mathbb{E}\big(\big|\mathbb{1}_{\{|\mathbb{I}_\rho \xi| > n^{\alpha/2}\}} \mathbb{I}_\rho \xi / \rho\big|\big)$$

$$\leq \frac{1}{\rho} \mathbb{E}\big(\big|\mathbb{1}_{\{|\xi| > n^{\alpha/2}\}} \mathbb{I}_\rho \xi\big|\big)$$

$$= \mathbb{E}\big(\big|\mathbb{1}_{\{|\xi| > n^{\alpha/2}\}} \xi\big|\big).$$

Finally, $\mathbb{E}(|\mathbb{1}_{\{\xi > n^{\alpha/2}\}} \xi|) \to 0$ as $n \to \infty$ by monotone convergence. $\square$

REMARK 1.10.   The proof of Lemma 1.9 illustrates that $\rho = 1/n$ is a transition point for sparse random variables of the type $\mathbb{I}_\rho \xi$ where the arguments for universality break down. Notably, the proof of Lemma 1.9 also works for α depending on n so long as $\alpha \log n$ tends to infinity as $n \to \infty$; for example, $\alpha = \frac{1}{\log \log n}$ is suitable. It would be interesting to see if the universality principle extends to parameters α that tend slowly to zero as $n \to \infty$.

1.3. *Further directions*.   There are a number of natural further directions to consider with respect to the sparse universality principle Theorem 1.5. One natural question is whether Theorem 1.5 can be generalized to prove almost sure convergence in addition to proving convergence in probability. A result of Dozier and Silverstein [7] is one of the ingredients used in [34] to prove almost sure convergence; however, there does not seem to be a sparse analog of [7]. Proving a sparse analog of [7] would be a substantial step toward proving a universality principle with almost sure convergence (see Remark 2.4), though there may be other avenues as well. Finally, a general question of interest would be to study the rates of convergence for the universality principle. Convergence seems reasonably fast in the nonsparse case; however, empirical evidence indicates that convergence is slower in the sparse case and may in fact depend on the underlying type of random variables; see Figure 4 for an example. A bound on convergence rates in the nonsparse case where the $(2 + \delta)$th moment is bounded is given in [32], Section 14.

1.4. *Definitions of convergence and notation*.   Let X be a random variable taking values in a Hausdorff topological space. We say that X_n *converges in probability* to X if for every neighborhood N_X of X, we have

$$\lim_{n \to \infty} \Pr(X_n \in N_X) = 1.$$

Furthermore, we say that X_n *converges almost surely* to X if

$$\Pr\left(\lim_{n \to \infty} X_n = X\right) = 1.$$

If C_n is a sequence of random variables taking values in $\mathbb{R}$, we say that C_n is *bounded in probability* if

$$\lim_{K \to \infty} \liminf_{n \to \infty} \Pr(C_n \leq K) = 1.$$

In the current paper, we are interested in how a randomly generated sequence of ESDs μ_{A_n} converges as $n \to \infty$, and so we will put the standard vague topology on the space of probability measures on $\mathbb{C}$. In particular, if μ_n and μ'_n are randomly generated sequences of measures on $\mathbb{C}$, then $\mu_n - \mu'_n$ converges in probability to zero if for every smooth function with compact support f and for every $\varepsilon > 0$, we have

$$\lim_{n \to \infty} \Pr\left(\left|\int_{\mathbb{C}} f \, d\mu_n - \int_{\mathbb{C}} f \, d\mu'_n\right| \leq \varepsilon\right) = 1.$$

Furthermore, $\mu_n - \mu'_n$ converges to zero almost surely if for every smooth function with compact support f and for every $\varepsilon > 0$, the expression $|\int_{\mathbb{C}} f \, d\mu_n - \int_{\mathbb{C}} f \, d\mu'_n|$ converges to 0 with probability 1.

For functions f and g depending on n, we will make use of the asymptotic notation $f = O(g)$ to mean that there exists a positive constant c (independent of n) such that $f \leq cg$ for all sufficiently large n. Also, we will use the asymptotic notation $f = o(g)$ to mean that $f/g \to 0$ as $n \to \infty$.
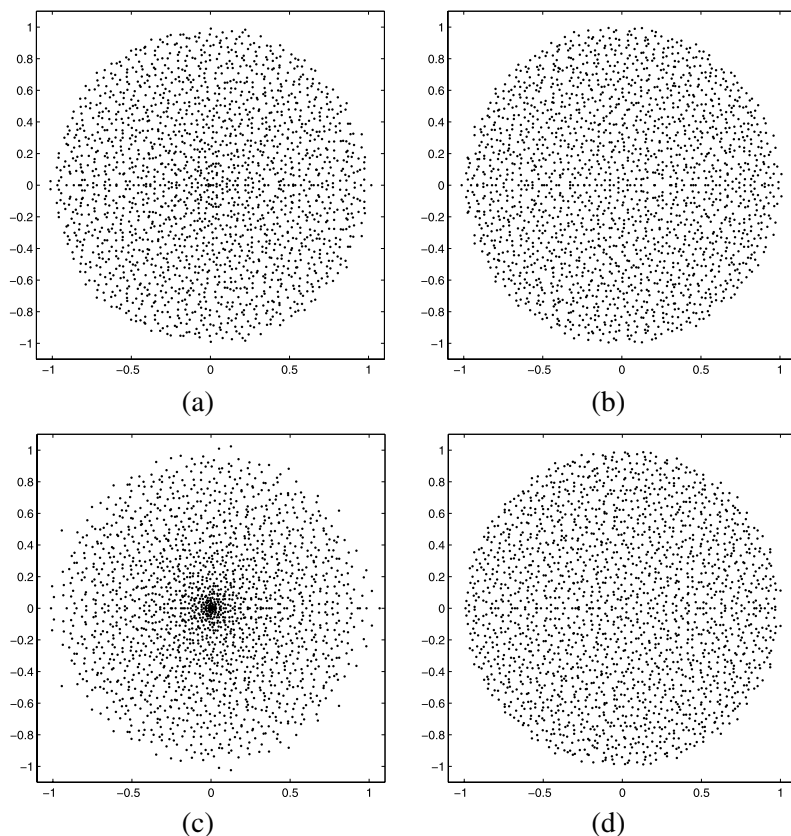
FIG. 4. *The four figures above indicate that the rates of convergence to the uniform distribution on the unit disk for sparse Bernoulli and sparse Gaussian random matrix ensembles are apparently not the same as each other, and that in particular the sparse Gaussian case converges more slowly that the nonsparse case. Each plot is of the eigenvalues of a* 2,000 *by* 2,000 *random matrix with i.i.d. entries. In the first column [parts* (a) *and* (c)] *the matrices are sparse with parameter* $\alpha = 0.2$, *which means each entry is zero with probability* $1 - \frac{1}{n^{0.8}}$, *and in the second column [parts* (b) *and* (d)] *the matrices are not sparse (i.e.,* $\alpha = 1$). *In the first row, both matrix ensembles are Bernoulli, so each nonzero entry is equally likely to be* −1 *or* 1, *and in the second row, the ensembles are Gaussian, so the nonzero entries are drawn from a Gaussian distribution with mean zero and variance one.*

1.5. *Paper outline.* Recall that the sparseness is determined by $\rho := n^{-1+\alpha}$. In the remaining sections, we will follow the approach used in [34] to prove a universality principle for sparse random matrices when $\alpha > 0$. In Section 2, we outline the main steps of the proof, highlighting a general result about convergence of ESDs from [34] that essentially reduces the question of convergences of ESDs to a question of convergence of the determinants of the corresponding matrices (one of which is sparse, and the other of which is not). Section 3 gives a proof of a sparse version of the necessary result on convergence of determinants based on a least singular value bound for sparse matrices in [32] and two lemmas, which are

proved in Sections 4 and 5, respectively. In Section 5, we make use of a complex version of a result of Chatterjee [5] (namely, Theorem 5.6) which requires adapting Krishnapur's ideas in [34], Appendix C, to a sparse context ([34], Appendix C, is dedicated to proving a universality principle for nonsparse random matrices where the entries are not necessarily i.i.d.).

2. Proof of Theorem 1.5. The following result was proven by Tao and Vu [34], Theorem 2.1, and can be applied directly in proving Theorem 1.5. All logarithms in this paper are natural unless otherwise noted.

THEOREM 2.1 ([34]). *Suppose for each n that $A_n, B_n \in \mathbb{M}_n(\mathbb{C})$ are ensembles of random matrices. Assume that*:

(i) *The expression*

$$(3) \qquad \frac{1}{n^2}\|A_n\|_2^2 + \frac{1}{n^2}\|B\|_2^2$$

is bounded in probability.

(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 to zero. In particular, for each fixed z, these determinants are nonzero with probability $1 - o(1)$.

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

Note that a stronger version of the above theorem appears in [34], Theorem 2.1, which additionally gives conditions under which $\mu_{(1/\sqrt{n})A_n} - \mu_{(1/\sqrt{n})B_n}$ converges almost surely to zero.

The lemma below is a sparse version of [34], Lemma 1.7.

LEMMA 2.2. *Let M_n, A_n and B_n be as in Theorem 1.5. Then $\frac{1}{n^2}\|A_n\|_2^2$ and $\int_{\mathbb{C}}|z|^2 d\mu_{(1/\sqrt{n})A_n}(z)$ are bounded in probability, and the same statement holds with B_n replacing A_n.*

PROOF. Our proof is the same as the proof [34], Lemma 1.7, except that we need to use a sparse version of the law of large numbers (which follows from, e.g., [10], Theorem 2.2.6). By the Weyl comparison inequality for second moment (see [34], Lemma A.2) it suffices to prove that $\frac{1}{n^2}\|A_n\|_2^2$ is bounded in probability, and by the triangle inequality along with inequality (2), it thus suffices to show that $\frac{1}{n^2}\|X_n\|_2^2$ is bounded in probability. By the sparse law of large numbers and the fact that $\mathbb{E}|x|^2 < \infty$, we see that $\frac{1}{n^2}\|X_n\|_2^2$ is bounded in probability. The statement with B_n replacing A_n is exactly [34], Lemma 1.7. □

The proof of Theorem 1.5 is completed by combining Theorem 2.1 and Lemma 2.2 with the following proposition:

PROPOSITION 2.3. *Let $0 < \alpha \leq 1$ be a constant, and let x be a random variable with mean zero and variance one. Let X_n be the sparse matrix ensemble for x with parameter α, and let Y_n be the n by n matrix having i.i.d. copies of x for each entry* (*in particular, Y_n is* not *sparse*). *For each n, let M_n be a deterministic n by n matrix satisfying inequality* (2), *and let $A_n := M_n + X_n$, and let $B_n := M_n + Y_n$. Then, for every fixed $z \in \mathbb{C}$, we have that*

$$(4) \qquad \frac{1}{n} \log \left| \det \left(\frac{1}{\sqrt{n}} A_n - zI \right) \right| - \frac{1}{n} \log \left| \det \left(\frac{1}{\sqrt{n}} B_n - zI \right) \right|$$

converges in probability to zero.

One useful property of the determinant is that it may be computed in a number of different ways. In particular, for a matrix M, we have

$$(5) \quad |\det(M)| = \prod_{i=1}^{n} |\lambda_i(M)| = \prod_{i=1}^{n} \sigma_i(M) = \prod_{i=1}^{n} \operatorname{dist}(R_i, \operatorname{Span}\{R_1, \ldots, R_{i-1}\}),$$

where $\lambda_i(M)$ and $\sigma_i(M)$ are the eigenvalues and singular values of M, respectively, and where R_i denotes the ith row of M.

In the remainder of the current section, we will outline the program for proving Proposition 2.3 and describe the differences between our proof and the proof of [34], Proposition 2.2. As in [34], we will prove Proposition 2.3 by writing the determinant as a product of distances between the ith row of a matrix and the span of the first $i - 1$ rows [thanks to (5)]. Proposition 2.3 can then be proven via three main steps:

(1) A bound on the least singular value due to Tao and Vu [32] for sparse and nonsparse random matrices is used to take care of terms very high-dimensional subspaces (i.e., span of more than $n - n^{1-\alpha/6}$ rows).

(2) Talagrand's inequality is used, along with other ideas from [34], to take care of terms with high dimension [i.e., span of more than $(1 - \delta)n$ rows] not already dealt with by the previous step. Some care must be taken in the sparse case with the constant α in the exponent in order to use Talagrand's inequality, which is where the $\alpha/6$ comes from in the previous step.

(3) A complex version of a result of Chatterjee [5] (namely Theorem 5.6) along with new ideas in [34] are used to take care of the remaining terms. Here, the sparse case differs substantially from the nonsparse case, in that we must use Theorem 5.6 in place of a result due to Dozier and Silverstein [7] used in [34]. This step, in general, follows Krishnapur [34], Appendix C, who investigates a universality principle for nonsparse random matrices with not necessarily i.i.d. entries, since there Dozier and Silverstein's result [7] cannot be applied.

REMARK 2.4. It would be natural to investigate a version of Theorem 1.5 where convergence in the almost sure sense is proved rather than convergence in probability. Typically, proving almost sure convergence is harder than proving convergence in probability; however, the universality principle in [34] is proven for both types of convergence, and so may provide a general approach to proving a universality principle for sparse random matrices with almost sure convergence. One of the steps in proving the universality principle of [34] in the almost sure sense uses a result due to Dozier and Silverstein [7]. In [7], a truncation argument is used that seems like it would need to be altered or replaced in order to prove a result for sparse random matrices. Another possible approach to proving a version of Theorem 1.5 for almost sure convergence would be to prove an analog of Chatterjee's [5], Theorem 1.1 (see Theorem 5.6) for almost sure convergence, though this might require a very different type of argument than the one used in [5]. A sparse version of the law of large numbers for almost sure convergence would also likely be necessary in any case.

3. Proof of Proposition 2.3. By shifting M_n by $zI\sqrt{n}$ [and noting that the new M_n still satisfies inequality (2)], it is sufficient to prove 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 to zero in probability.

Following the notation of [34], let $X_1, \ldots, X_n$ be the rows of A_n, and let $Y_1, \ldots, Y_n$ be the rows of B_n. Let $Z_1, \ldots, Z_n$ denote the rows of M_n, and note that by inequality (2) we have that

$$\sum_{j=1}^{n} \|Z_j\|_2^2 = O(n^2).$$

By re-ordering the rows of A_n, B_n and M_n if necessary, we may assume that the rows $Z_{\lceil n/2 \rceil}, \ldots, Z_n$ have the smallest norms, and so

(6) $\|Z_i\|_2 = O(\sqrt{n})$ for $n/2 \le i \le n$.

This fact will be used in part of the proof of Lemma 3.2.

For $1 \le i \le n$, let V_i be the $(i-1)$-dimensional space generated by $X_1, \ldots, X_{i-1}$, and let W_i be the $(i-1)$-dimensional space generated by $Y_1, \ldots, Y_{i-1}$. By standard formulas for the determinant [see (5)], we have that

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

and

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

It is thus sufficient to show that

$$(7) \qquad \frac{1}{n} \sum_{i=1}^{n} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) - \log \operatorname{dist}\left(\frac{1}{\sqrt{n}} Y_i, W_i\right)$$

converges in probability to zero. We will start by proving somewhat weak upper and lower bounds on $\operatorname{dist}(\frac{1}{\sqrt{n}} X_i, V_i)$ and $\operatorname{dist}(\frac{1}{\sqrt{n}} Y_i, W_i)$ that hold for all i. For the upper bound, note that by Chebyshev's inequality we have $\Pr(\|X_i\|_2 > n^2) \leq n^{-3}$, and thus by the Borel–Cantelli lemma, we have with probability 1 that $\|X_i\|_2 < n^2$ for all but finitely many n and for all i. This implies that, with probability 1,

$$\operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) \leq \|X_i\|_2 = n^{O(1)}$$

for all but finitely many n and for all i; and the same bound also holds for $\operatorname{dist}(\frac{1}{\sqrt{n}} Y_i, W_i)$. To show a lower bound, define $S_j^{(i)} := \operatorname{Span}(\{X_1, \ldots, X_i\} \setminus \{X_j\})$, and define $A_n^{(i)}$ to be the i by n matrix consisting of the first i rows of A_n. By [34], Lemma A.4, we have

$$\sum_{j=1}^{i} \operatorname{dist}(X_j, S_j^{(i)})^{-2} = \sum_{j=1}^{i} \sigma_j(A_n^{(i)})^{-2},$$

and since $V_i = S_i^{(i)}$, we thus have the crude bound

$$\operatorname{dist}(X_i, V_i)^{-2} \leq n \sigma_i(A_n^{(i)})^{-2}.$$

By Cauchy interlacing (see [34], Lemma A.1), we know that $\sigma_i(A_n^{(i)}) \geq \sigma_n(A_n)$, and thus we have

$$\frac{1}{n} \sigma_n(A_n) \leq \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right),$$

and by the same reasoning,

$$\frac{1}{n} \sigma_n(B_n) \leq \operatorname{dist}\left(\frac{1}{\sqrt{n}} Y_i, W_i\right).$$

Lower bounds on $\operatorname{dist}(\frac{1}{\sqrt{n}} X_i, V_i)$ and $\operatorname{dist}(\frac{1}{\sqrt{n}} Y_i, W_i)$ will now follow from lower bounds on the least singular values of A_n and B_n which were proven in [32].

LEMMA 3.1 (Least singular value bound for sparse random matrices [32]). *Let $0 < \alpha \leq 1$ be a constant, and let x be a random variable with mean zero and variance one. Let X_n be the sparse matrix ensemble for x with parameter α, and let Y_n be the n by n matrix having i.i.d. copies of x for each entry (in particular, Y_n is* not *sparse). For each n, let M_n be a deterministic n by n matrix satisfying*

inequality (2), *and let* $A_n := M_n + X_n$, *and let* $B_n := M_n + Y_n$. *Then with probability* 1 *we have*

$$\sigma_n(A_n), \sigma_n(B_n) \geq n^{-O(1)}$$

for all but finitely many n.

PROOF.   Paraphrasing [34], proof of Lemma 4.1, the proof follows by combining [32], Theorem 2.5 (for the nonsparse matrix) and [32], Theorem 2.9 (for the sparse matrix) each with the Borel–Cantelli lemma, noting that the hypotheses of [32], Theorem 2.5, and [32], Theorem 2.9, are satisfied due to [32], Lemma 2.4, and inequality (2).   □

Thus, with probability 1 we have

$$(8) \qquad \left| \log \text{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i \right) \right|, \left| \log \text{dist}\left(\frac{1}{\sqrt{n}} Y_i, W_i \right) \right| \leq O(\log n)$$

for all but finitely many n. In light of inequality (8), the following two lemmas suffice to prove that the quantity in display (7) converges in probability to zero.

Recall that α is the parameter used to determine the sparseness of the sparse matrix ensemble.

LEMMA 3.2 (High-dimensional contribution).   *For every* $\varepsilon > 0$, *there exists a constant* $0 < \delta_\varepsilon < 1/2$ *such that for every* $0 < \delta < \delta_\varepsilon$ *we have with probability* 1 *that*

$$\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n - n^{1-\alpha/6}} \left| \log \text{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i \right) \right| = O(\varepsilon)$$

for all but finitely many n.

Note that Lemma 3.2 with Y_i (which is not sparse) replacing X_i and with W_i replacing V_i was proven in [34], Lemma 4.2, with 0.99 replacing $1 - \alpha/6$. Alternatively, the nonsparse case follows from our proof of Lemma 3.2 if one sets $\alpha = 1$ (giving an exponent of 5/6 in place of the exponent 0.99 used in [34], Lemma 4.2). Also, note that for all sufficiently large n, we may assume that (6) holds for all i relevant to Lemma 3.2 above.

LEMMA 3.3 (Low-dimensional contribution).   *For every* $\varepsilon > 0$, *there exists* $0 < \delta < \varepsilon$ *such that with probability at least* $1 - O(\varepsilon)$ *we have*

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

for all but finitely many n.

To complete the proof of Proposition 2.3, one may combine Lemma 3.2 ([34], Lemma 4.2) (which is the nonsparse analog of Lemma 3.2) and Lemma 3.3. In particular, given $\varepsilon_{3.2} > 0$ in Lemma 3.2, we need the $\delta_{3.3}$ in Lemma 3.3 to be smaller than $\delta_{\varepsilon_{3.2}}$. This may be accomplished by choosing $\varepsilon_{3.3}$ from Lemma 3.3 to be smaller than $\delta_{\varepsilon_{3.2}}$ given by Lemma 3.2.

4. Proof of Lemma 3.2. Following [34], we will prove Lemma 3.2 in two parts, splitting the summands into cases where the log is positive and where the log is negative. The proof below follows the proof of [34], Lemma 4.2, closely, and we have included it in detail to make explicit the role of α, which determines the sparseness of the matrix A_n. One place where particular care must be taken with sparseness parameter α is in a truncation argument needed to apply Talagrand's inequality (see Section 4.3). There, we have made frequent use of the assumption that α is a positive constant, though it is possible that a very slowly decreasing α could also work; see Lemma 1.9 and Remark 1.10.

4.1. *Positive log component.* In this section, we will use the notation

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

By the Borel–Cantelli lemma, the desired bound on the positive log component may be proven by showing

$$\sum_{n=1}^{\infty} \Pr\left(\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n - n^{1-\alpha/6}} \log_+ \mathrm{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) \geq \varepsilon\right) < \infty.$$

We will use the crude bound $\log_+ \mathrm{dist}(\frac{1}{\sqrt{n}} X_i, V_i) \leq \log_+ (\frac{\|X_i\|_2}{\sqrt{n}})$. Note that if $2^{m_0} \leq \frac{\|X_i\|_2}{\sqrt{n}} < 2^{m_0+1}$, then $m_0 \leq \log_2(\frac{\|X_i\|_2}{\sqrt{n}}) < m_0 + 1$, and so

$$\sum_{m=0}^{\infty} \mathbb{1}_{\{\|X_i\|_2 \geq 2^m \sqrt{n}\}} = m_0 + 1 > \log_2\left(\frac{\|X_i\|_2}{\sqrt{n}}\right).$$

Thus,

(9)
$$\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n - n^{1-\alpha/6}} \log_+ \mathrm{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right)$$

$$\leq \sum_{m=0}^{\infty} \frac{1}{n} \sum_{(1-\delta)n \leq i \leq n - n^{1-\alpha/6}} \mathbb{1}_{\{\|X_i\|_2 \geq 2^m \sqrt{n}\}}.$$

If the left-hand side of inequality (9) is at least ε for a given n, then we must have for some $m \geq 0$ that

(10)
$$\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n - n^{1-\alpha/6}} \mathbb{1}_{\{\|X_i\|_2 \geq 2^m \sqrt{n}\}} \geq \frac{2\varepsilon}{(100 + m)^2}.$$

We now have two cases to consider. For the first case, assume that the small-est m satisfying inequality (10) satisfies $m \geq n^{1/5}$. Then for inequality (10) to be satisfied, there exists some $1 \leq i \leq n$ such that $\|X_i\|_2 \geq 2^{n^{1/5}}\sqrt{n}$. By Chebyshev's inequality and equation (6), we have that $\Pr(\|X_i\|_2 \geq 2^{n^{1/5}}\sqrt{n}) \leq O(\frac{1}{2^{2n^{1/5}}})$, and thus the probability of such an i existing is at most $f(n) := 1 - (1 - c2^{-2n^{1/5}})^n$, where c is some constant. It is not hard to show that $f(n)n^2 \to 0$ as $n \to \infty$, and thus, for all sufficiently large n, we have the probability that there exists an i such that $\|X_i\|_2 \geq 2^{n^{1/5}}\sqrt{n}$ is at most ε/n^2. Since this probability is summable in n, we have proved inequality (9) in the first case.

For the second case, assume that the smallest m satisfying inequality (10) satisfies $0 \leq m < n^{1/5}$. In this case we will use Hoeffding's inequality.

THEOREM 4.1 (Hoeffding's inequality [17]).   *Let* $\beta_1, \ldots, \beta_k$ *be independent random variables such that for* $1 \leq i \leq k$ *we have*

$$\Pr(\beta_i - \mathbb{E}(\beta_i) \in [0, 1]) = 1.$$

Let $S := \sum_{i=1}^k \beta_i$. *Then*

$$\Pr(S \geq kt + \mathbb{E}(S)) \leq \exp(-2kt^2).$$

The random variables β_i will be $\mathbb{1}_{\{\|X_i\|_2 \geq 2^m \sqrt{n}\}}$, and thus we need to control $\Pr(\|X_i\|_2 \geq 2^m \sqrt{n})$ in order to bound $\mathbb{E}(S)$. By (6) and Chebyshev's inequality, we have that

$$(11) \qquad\qquad \Pr(\|X_i\|_2 \geq 2^m \sqrt{n}) \leq O\left(\frac{1}{2^{2m}}\right).$$

We will take $k = n - n^{1-\alpha/6} - (1-\delta)n$, so we have that $\lim_{n\to\infty} \frac{k}{n} = \delta$. Also, δ_ε sufficiently small so that $\delta_\varepsilon < \frac{\varepsilon}{20,000C}$, where C is the implicit constant in inequality (11). If we take $t = \frac{n}{k}(\frac{\varepsilon}{(100+m)^2})$, we can compute that

$$\frac{kt}{n} + \frac{1}{n}\mathbb{E}(S) \leq \frac{\varepsilon}{(100+m)^2} + \frac{2\delta_\varepsilon C}{2^{2m}} \leq \frac{2\varepsilon}{(100+m)^2}$$

for all sufficiently large n (the second inequality follows by taking n sufficiently large so that $k/n \leq 2\delta < 2\delta_\varepsilon$). Thus, by Hoeffding's inequality and taking n sufficiently large, we have

$$\Pr\left(\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n-n^{1-\alpha/6}} \mathbb{1}_{\{\|X_i\|_2 \geq 2^m \sqrt{n}\}} \geq \frac{2\varepsilon}{(100+m)^2}\right)$$

$$\leq \exp\left(\frac{-n\varepsilon^2}{\delta(100+m)^4}\right)$$

$$\leq \max\left\{\exp\left(\frac{-n\varepsilon^2}{\delta(200)^4}\right), \exp\left(\frac{-n^{1/5}\varepsilon^2}{16\delta}\right)\right\},$$

where the last inequality follows from our assumption in this second case that $0 \leq m \leq n^{1/5}$. Thus, we have shown for all sufficiently large n and any $0 \leq \delta < \delta_\varepsilon$ that

$$\Pr\left(\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n - n^{1-\alpha/6}} \log_+ \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) \geq 0\right)$$

$$\leq \max\left\{\exp\left(\frac{-n\varepsilon^2}{\delta(200)^4}\right), \exp\left(\frac{-n^{1/5}\varepsilon^2}{16\delta}\right)\right\}.$$

Finally, we note that the bounds from the two cases sum to at most

$$\varepsilon/n^2 + \max\left\{\exp\left(\frac{-n\varepsilon^2}{\delta(200)^4}\right), \exp\left(\frac{-n^{1/5}\varepsilon^2}{16\delta}\right)\right\},$$

which is summable in n, thus completing the proof for the positive log component.

4.2. *Negative log component.* In this section, we will use the notation $\log_-(x) := \max\{-\log(x), 0\}$.

By the Borel–Cantelli lemma, it suffices to show that

$$(12) \qquad \sum_{n=1}^{\infty} \Pr\left(\frac{1}{n} \sum_{(1-\delta)n \leq i \leq n - n^{1-\alpha/6}} \log_- \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) \geq \varepsilon\right) < \infty.$$

Following the approach in [34], our main tool is the following lemma.

PROPOSITION 4.2. *Let $0 < \alpha \leq 1$ be a constant, let $1 \leq d \leq n - n^{1-\alpha/6}$, let $0 < c < 1$ be a constant and let W be a deterministic d-dimensional subspace of $\mathbb{C}^n$. Let X be a row of A_n. Then*

$$\Pr\left(\operatorname{dist}(X, W) \leq c\sqrt{n-d}\right) \leq 6\exp(-n^{\alpha/2})$$

for all n sufficiently large with respect to c and α.

We will give the proof of Proposition 4.2 in Section 4.3. The proof of the negative log component of Lemma 3.2 can be completed by using Proposition 4.2 and following the proof of Lemma 4.2 of [34], which we paraphrase below.

Taking $c = 1/2$ in Proposition 4.2 and conditioning on V_i, we have that for each $(1 - \delta)n \leq i \leq n - n^{\alpha/6}$ that

$$\Pr\left(\operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) > \frac{\sqrt{n-i+1}}{2\sqrt{n}}\right) \geq 1 - O(\exp(n^{-\alpha/2})).$$

Thus, the probability that

$$(13) \qquad \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) > \frac{\sqrt{n-i+1}}{2\sqrt{n}}$$

simultaneously for all $(1 - \delta)n \leq i \leq n - n^{\alpha/6}$ is at least $1 - O(n^{-10})$ (in fact, better bounds are possible, but this is sufficient).

Finally, choosing δ_ε sufficiently small so that $\frac{\delta_\varepsilon}{2} \log \frac{4}{\delta_\varepsilon} < \varepsilon$, we can take the log of inequality (13) and sum in i to get that the probability in the summand of inequality (12) in at most $O(n^{-10})$, and this is summable in n, completing the proof of inequality (12).

4.3. *Proof of Proposition* 4.2.   Recall that X has coordinates $a_i = \frac{\mathbb{I}_\rho x_i}{\sqrt{\rho}} + m_i$, where m_i is a fixed element (it comes from the matrix M_n), x_i is a fixed, mean zero, variance 1 random variable (it does not change with n) and $\rho = n^{-1+\alpha}$ where $0 < \alpha \leq 1$ is a constant. The proof of Proposition 4.2 closely follows the proof of Proposition 5.1 of [34], and we give the details below to highlight how the proof must be modified to accommodate sparseness with parameter α. In particular, care must be taken with the value of α in the following three steps: first, when reducing to the case where the sparse random variables are bounded (since sparseness requires scaling by $1/n^{-1+\alpha}$), second, when showing that the sparse random variables restricted to the bounded case still have variance tending to 1 as $n \to \infty$, and third, when applying Talagrand's inequality where one must keep track of α in the exponent on the upper bound.

PROOF OF PROPOSITION 4.2.   First we reduce to the case where X has mean 0. Let $v = \mathbb{E}(X)$. (Note that v is the row of M_n corresponding to X.)

Note that $\mathrm{dist}(X, W) \geq \mathrm{dist}(X - v, \mathrm{Span}(W, v))$. Thus, by changing constants slightly (while still preserving $0 < c < 1$) and replacing d by $d + 1$, it suffices to prove Proposition 4.2 in the mean zero case.

The second step is reducing to a case where the coordinates of X are bounded. In particular, we will show that, with probability at least $1 - 2\exp(-n^{\alpha/2})$, all but $n^{0.8}$ of the coordinates of X take values that are less than $n^{1/2-\alpha/4}$. Let $t_i := \mathbb{1}_{\{|a_i| \geq n^{(1-\alpha/2)/2}\}}$, and let $T := \sum_{i=1}^{n} t_i$. If $\mathbb{E}(T) = 0$, then with probability 1 we have that $|a_i| < n^{(1-\alpha/2)/2}$, and we are done with the reduction to the case where the coordinates are bounded. Thus, it is left to show this reduction in the case where $\mathbb{E}(T) > 0$.

By Chernoff (see [31], Corollary 1.9) we know that for every $\varepsilon > 0$ we have

$$\Pr(|T - \mathbb{E}(T)| \geq \varepsilon \mathbb{E}(T)) \leq 2\exp\left(-\min\left\{\frac{\varepsilon^2}{4}, \frac{\varepsilon}{2}\right\} \mathbb{E}(T)\right).$$

Since $\mathbb{E}(T) > 0$ by assumption, we may set $\varepsilon := \frac{n^{0.8}}{\mathbb{E}(T)} - 1$. By Chebyshev's inequality, we have $\Pr(|a_i| \geq n^{(1-\alpha/2)/2}) \leq n^{-1+\alpha/2}$ for all $1 \leq i \leq n$, and thus $\mathbb{E}(T) = n\mathbb{E}(t_i) \leq n^{\alpha/2}$, which implies that $\varepsilon \geq n^{0.8-\alpha/2} - 1 \geq 2$ for large n. Here

we used the fact that $0 < \alpha/2 \leq 0.5$. Using the Chernoff bound we have

$$\Pr\big(T \geq (1+\varepsilon)\mathbb{E}(T) = n^{0.8}\big) \leq 2\exp\left(-\frac{\varepsilon}{2}\mathbb{E}(T)\right)$$

$$\leq 2\exp\big(-n^{0.8}/2 + \mathbb{E}(T)/2\big)$$

$$\leq 2\exp\big(-n^{0.8}/2 + n^{\alpha/2}/2\big)$$

$$\leq 2\exp\big(-n^{0.8}/4\big)$$

$$\leq 2\exp\big(-n^{\alpha/2}\big).$$

Thus, with probability at least $1 - 2\exp(-n^{\alpha})$, there are at most $n^{0.8}$ indices for which $|a_i| \geq n^{(1-\alpha/2)/2}$. For a subset $I \subset \{1, 2, \ldots, n\}$, let E_I denote the event that $I = \{i : |a_i| \geq n^{1/2-\alpha/4}; 1 \leq i \leq n\}$.

By the law of total probability, we have

$$\Pr\big(\mathrm{dist}(X, W) \leq c\sqrt{n-d}\big)$$

$$\leq 2\exp(-n^{\alpha/2})$$

$$+ \sum_{\substack{I \subset \{1,\ldots,n\} \\ |I| \leq n^{0.8}}} \Pr\big(\mathrm{dist}(X, W) \leq c\sqrt{n-d} \,|\, E_I\big)\Pr(E_I).$$

Thus, it is sufficient to show that

$$\Pr\big(\mathrm{dist}(X, W) \leq c\sqrt{n-d} \,|\, E_I\big) \leq 4\exp(-n^{\alpha/2})$$

for each $I \subset \{1, \ldots, n\}$ such that $|I| \leq n^{0.8}$.

Fix such a set I. By renaming coordinates, we may assume that $I = \{n' + 1, \ldots, n\}$ where $n - n^{0.8} \leq n' \leq n$. The next step is projecting away the coordinates in I. In particular, let $\pi : \mathbb{C}^n \to \mathbb{C}^{n'}$ be the orthogonal projection onto the first n' coordinates, and note that

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

Thus, we can condition on $a_{n'+1}, \ldots, a_n$, adjust c slightly (without changing the fact that $0 < c < 1$) and (abusing notation to henceforth let n stand for n') see that it is sufficient to show

$$\Pr\big(\mathrm{dist}(X, W) \leq c\sqrt{n-d} \,\big|\, |a_i| < n^{1/2-\alpha/4}, \text{ for every } 1 \leq i \leq n\big)$$

$$\leq 4\exp(-n^{\alpha/2}).$$

LEMMA 4.3. *Let $\tilde{a}_i$ be the random variable a_i conditioned on $|a_i| < n^{1/2-\alpha/4}$. Then $\tilde{a}_i$ has variance $1 + o(1)$.*

PROOF.   By definition

$$\mathrm{Var}(\tilde{a}_i) = \mathbb{E}(|\tilde{a}_i|^2) - |\mathbb{E}(\tilde{a}_i)|^2$$

$$= \mathbb{E}(|a_i|^2 \,|\, |a_i| < n^{1/2-\alpha/4}) - |\mathbb{E}(a_i \,|\, |a_i| < n^{1/2-\alpha/4})|^2$$

$$= \frac{1}{\Pr(|a_i| < n^{1/2-\alpha/4})} \mathbb{E}\big(|a_i|^2 \mathbb{1}_{\{|a_i| < n^{1/2-\alpha/4}\}}\big)$$

$$- \frac{1}{\Pr(|a_i| < n^{1/2-\alpha/4})^2} \big|\mathbb{E}\big(a_i \mathbb{1}_{\{|a_i| < n^{1/2-\alpha/4}\}}\big)\big|^2.$$

Note that $a_i = \frac{\mathbb{I}_\rho x_i}{\sqrt{\rho}}$, and so $|a_i| < n^{1/2-\alpha/4}$ if and only if $|\mathbb{I}_\rho x_i| < n^{\alpha/4}$. Since x_i does not change with n, we see that $\Pr(|a_i| < n^{1/2-\alpha/4}) = \Pr(|\mathbb{I}_\rho x_i| < n^{\alpha/4}) \to 1$ as $n \to \infty$. Also, by Lemma 1.9, we know that $\mathbb{E}(|a_i|^2 \mathbb{1}_{\{|a_i| < n^{1/2-\alpha/4}\}}) \to \mathbb{E}(|a_i|^2) = 1$ and that $\mathbb{E}(a_i \mathbb{1}_{\{|a_i| < n^{1/2-\alpha/4}\}}) \to \mathbb{E}(a_i) = 0$. Thus, we have shown that $\tilde{a}_i$ has variance $1 + o(1)$. □

Next, we recenter $\tilde{a}_i$ by subtracting away its mean, and we call the result $\tilde{a}_i$. Note that this recentering does not change the variance. We will use the following version of Talagrand's inequality, quoted from [34], Theorem 5.2; see also [19], Corollary 4.10:

THEOREM 4.4 (Talagrand's inequality).   *Let* **D** *be the unit disk* $\{z \in \mathbb{C}, |z| \leq 1\}$. *For every product probability* μ *on* $\mathbf{D}^n$, *every convex* 1-*Lipschitz function* $F : \mathbb{C}^n \to \mathbb{R}$, *and every* $r \geq 0$,

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

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

Let $\tilde{X} = (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)$, and let μ be the distribution on $\mathbf{D}^n$ given by $\tilde{X}/2n^{1/2-\alpha/4}$. Let $F(u) := \frac{1}{2}\mathrm{dist}(u, W)^2$, and note that F is convex and 1-Lipschitz, which follows since $\mathrm{dist}(u, W)$ is both convex and 1-Lipschitz [and also using the fact that $\mathrm{dist}(u, W) \leq 1$, since $0 \in W$].

By Theorem 4.4 with $r = 3n^{\alpha/4}$, we have

$$\Pr\big(|\mathrm{dist}(\tilde{X}, W)^2 - M(\mathrm{dist}(\tilde{X}, W)^2)| \geq 12n^{\alpha/4}n^{1-\alpha/2}\big) \leq 4\exp(-n^{\alpha/2}),$$

which implies that

(14) $$\Pr\big(\mathrm{dist}(\tilde{X}, W)^2 \leq M(\mathrm{dist}(\tilde{X}, W)^2) - 12n^{1-\alpha/4}\big) \leq 4\exp(-n^{\alpha/2}).$$

Recall that $F = \frac{1}{2}\mathrm{dist}(\frac{\tilde{X}}{2n^{1/2-\alpha/4}}, W)^2$. Using Talagrand's inequality (Theorem 4.4) again, we will show that the mean of F is very close to the median of F.

We compute

$$|\mathbb{E}(F) - M(F)| \le \mathbb{E}|F - M(F)| = \int_0^\infty \Pr(|F - M(F)| \ge t)\,dt$$

$$\le \int_0^\infty 4\exp(-t^2/8)\,dt = 8\sqrt{2\pi}.$$

Thus, we have shown that

(15) $$|\mathbb{E}(\text{dist}(\tilde{X}, W)^2) - M(\text{dist}(\tilde{X}, W)^2)| \le (32\sqrt{2\pi})n^{1-\alpha/2}.$$

LEMMA 4.5. $\mathbb{E}(\text{dist}(\tilde{X}, W)^2) = (1 + o(1))(n - d)$.

PROOF. Let $\pi := (\pi_{ij})$ denote the orthogonal projection matrix to W. Note that $\text{dist}(\tilde{X}, W)^2 = \sum_{i=1}^n \sum_{j=1}^n \tilde{a}_i \pi_{ij} \tilde{a}_j$. Since $\tilde{a}_i$ are i.i.d., mean zero random variables, we have

$$\mathbb{E}(\text{dist}(\tilde{X}, W)^2) = \mathbb{E}(|\tilde{a}_i|^2) \sum_{i=1}^n \pi_{ii} = \mathbb{E}(|\tilde{a}_i|^2)\,\text{tr}(\pi).$$

The proof is completed by applying Lemma 4.3 and noting that the trace of π is $n - d$. $\square$

From inequality (14), we see that it is sufficient to show that

$$M(\text{dist}(\tilde{X}, W)^2) - 12n^{1-\alpha/4} \ge c^2(n - d).$$

Using inequality (15) and Lemma 4.5 we have for sufficiently large n that

$$M(\text{dist}(\tilde{X}, W)^2) - 12n^{1-\alpha/4}$$

$$\ge \mathbb{E}(\text{dist}(\tilde{X}, W)^2) - (32\sqrt{2\pi})n^{1-\alpha/2} - 12n^{1-\alpha/4}$$

$$\ge \left(c^2 + \frac{1-c^2}{2}\right)(n-d) - (32\sqrt{2\pi})n^{1-\alpha/2} - 12n^{1-\alpha/4}$$

$$\ge c^2(n-d) + \left(\frac{1-c^2}{2}\right)n^{1-\alpha/6} - (32\sqrt{2\pi})n^{1-\alpha/2} - 12n^{1-\alpha/4}$$

$$\ge c^2(n-d),$$

where the last inequality follows from the fact that

$$\left(\frac{1-c^2}{2}\right)n^{1-\alpha/6} - (32\sqrt{2\pi})n^{1-\alpha/2} - 12n^{1-\alpha/4}$$

is a positive quantity for sufficiently large n. Combining the above computation with inequality (14) completes the proof of Proposition 4.2. $\square$

5. Proof of Lemma 3.3. Lemma 3.3 follows directly from the slightly more detailed statement in Lemma 5.1 given below. In this section, we will prove Lemma 5.1 by adapting the proof of Lemma 4.3 of [34], with some changes. The biggest difference with the proof of Lemma 4.3 of [34] is in the proof of Lemma 5.3, where we must adapt the approach of Krishnapur from [34], Appendix C, to a sparse setting (see Lemma 5.5). This is one critical juncture where it seems like it would take a new idea to prove almost sure convergence in place of convergence in probability. One possible approach would be proving a sparse version of [7] (which is used in [34] in the proof of almost sure convergence in the nonsparse case). Other notable differences from the proof of Lemma 4.3 of [34] are that we must use Proposition 4.2 in place of Proposition 5.1 of [34], and that we kill keep track of a lower bound on δ, which simplifies some steps in the proof.

LEMMA 5.1.   *For every $\varepsilon_1 > 0$ and for all sufficiently small $\varepsilon_2 > 0$, where ε_2 depends on ε_1 and other constants, the following holds. For every $\delta > 0$ satisfying*

$$\varepsilon_2^2 < \delta \leq \frac{\varepsilon_2}{40 \log(1/\varepsilon_2)},$$

we have with probability $1 - O(\varepsilon_1)$

$$\left| \frac{1}{n} \sum_{1 \leq i \leq (1-\delta)n} \log \operatorname{dist}\left(\frac{1}{\sqrt{n}} X_i, V_i\right) - \log \operatorname{dist}\left(\frac{1}{\sqrt{n}} Y_i, W_i\right) \right| = O(\varepsilon_2)$$

for all but finitely many n.

As shown in [34], Section 6, it is sufficient to prove that with probability $1 - O(\varepsilon_1)$ we have

$$(16) \qquad \left| \frac{1}{n'} \sum_{i=1}^{n'} \log\left(\frac{1}{\sqrt{n}} \sigma_i(A_{n,n'})\right) - \log\left(\frac{1}{\sqrt{n}} \sigma_i(B_{n,n'})\right) \right| = O(\varepsilon_2)$$

for all but finitely many n, where $n' = \lfloor (1 - \delta)n \rfloor$, where $\sigma_i(A)$ denotes the ith largest singular value of a matrix A, and where $A_{n,n'}$ denotes the matrix consisting of the first n' rows of A_n and $B_{n,n'}$ denotes the matrix consisting of the first n' rows of B_n.

Proving (16) is equivalent to showing

$$(17) \qquad \left| \int_0^\infty \log t \, d\nu_{n,n'}(t) \right| = O(\varepsilon_2),$$

where $d\nu_{n,n'}$ is defined by the difference of the two relevant ESDs, namely

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

Following [34], we can prove (17) by dividing the range of t into a few parts, which follows from Lemma 5.2 (for large t), Lemma 5.3 (for intermediate-sized t) and Lemma 5.4 (for small t).

LEMMA 5.2 (Region of large t).   *For every $\varepsilon_1 > 0$, there exist constants $\varepsilon_2 > 0$ and R_{ε_2} such that with probability $1 - O(\varepsilon_1)$ we have*

$$\int_{R_{\varepsilon_2}}^{\infty} |\log t| |dv_{n,n'}(t)| \leq \varepsilon_2.$$

PROOF.   By Lemma 2.2 and Lemma A.2 of [34], we have that $\int_0^{\infty} t|dv_{n,n'}(t)|$ is bounded in probability. Thus, there exists a constant C_{ε_1} depending on ε_1 such that with probability $1 - O(\varepsilon_1)$ we have

$$\int_0^{\infty} t|dv_{n,n'}(t)| \leq C_{\varepsilon_1}.$$

Choose $\varepsilon_2 > 0$ sufficiently small with respect to ε_1 and C_{ε_1} so that

$$1 \geq 2C_{\varepsilon_1}\varepsilon_2 \log\left(\frac{1}{\varepsilon_2}\right).$$

Set $R_{\varepsilon_2} = (\frac{1}{\varepsilon_2})^2$, and assume without loss of generality that $R_{\varepsilon_2} > e$. Note that $\frac{t}{\log t}$ is increasing for $t \geq R_{\varepsilon_2} > e$, and thus by the definition of ε_2 we have

$$\frac{C_{\varepsilon_1}}{\varepsilon_2} \log(t) \leq t,$$

whenever $t \geq R_{\varepsilon_2}$. Thus, we have with probability $1 - O(\varepsilon_1)$ that

$$\int_{R_{\varepsilon_2}}^{\infty} |\log t| |dv_{n,n'}(t)| \leq \int_0^{\infty} \frac{\varepsilon_2}{C_{\varepsilon_1}} t|dv_{n,n'}(t)| \leq \varepsilon_2. \qquad \square$$

LEMMA 5.3 (Region of intermediate t, namely $\varepsilon_2^2 \leq t \leq R_{\varepsilon_2}$).   *Define a smooth function $\psi(t)$ which equals 1 on the interval $[\varepsilon_2^4, R_{\varepsilon_2}]$, equals zero outside the interval $(\varepsilon_2^4/2, 2R_{\varepsilon_2})$, is monotonically increasing on $(\varepsilon_2^4/2, \varepsilon_2^4)$ and is monotonically decreasing on $(R_{\varepsilon_2}, 2R_{\varepsilon_2})$.*
    Then with probability $1 - O(\varepsilon_1)$ we have

$$\left| \int_0^{\infty} \psi(t) \log(t) \, dv_{n,n'}(t) \right| = O(\varepsilon_2),$$

so long as $\delta \leq \frac{\varepsilon_2}{40\log(1/\varepsilon_2)}$.

The main step in this proof is applying Lemma 5.5, whereas in the analogous step in the nonsparse case, [34] uses a result of Dozier and Silverstein [7], which proves almost sure convergence of the relevant distributions (rather than convergence in probability, which is the limit of Lemma 5.5). It would be interesting to see if a sparse analog of [7] is possible, especially as it might be a step toward prov-

ing a universality result for sparse random matrices with almost sure convergence instead of convergence in probability.

PROOF OF LEMMA 5.3.   Using [34], Lemma A.1, and the upper bound on δ, it is possible to show that

$$\left| \int_0^\infty \psi(t) \log(t) \, dv_{n,n'}(t) \right| = \left| \int_0^\infty \psi(t) \log(t) \, dv_{n,n}(t) \right| + O(\varepsilon_2).$$

(A possible alternative to the step above would be proving an analog of Lemma 5.5 for rectangular n by n' matrices.)

By Lemma 5.5 (see Section 5.1), we know that $dv_{n,n}$ converges in probability to zero, and thus

$$\left| \int_0^\infty \psi(t) \log(t) \, dv_{n,n}(t) \right| = O(\varepsilon_2),$$

completing the proof.   □

The last step in proving (17) and thus completing the proof of Lemma 5.1 is the following lemma:

LEMMA 5.4 (Region of small t, namely $0 < t \le \varepsilon_2^4 < \delta^2$).   *With probability 1, we have*

$$\int_0^{\varepsilon_2^4} |\log t| |dv_{n,n'}(t)| = O(\varepsilon_2),$$

so long as $\delta \le \frac{1}{2} (\frac{\varepsilon_2}{\log(1/\varepsilon_2)})^{1/4}$.

PROOF.   The required upper bound on δ follows from the assumption that $\delta < \frac{\varepsilon_2}{40 \log(1/\varepsilon_2)}$. The proof is the same as the proof for Lemma 6.6 of [34], with the small change that one must use Proposition 4.2 in place of Proposition 5.1 of [34].
□

5.1. *Applying an approach of Chatterjee.*   In this subsection, we follow the ideas used by Krishnapur in [34], Appendix C, where a central-limit-type theorem due to Chatterjee [5] was used to prove a universality result for random matrices with independent but not necessarily identically distributed entries. Lemma 5.5 below is analog of Lemma C.3 of [34]. Recall that $\mathbb{I}_\rho$ is an i.i.d. copy of the random variable taking the value 1 with probability ρ and the value 0 with probability $1 - \rho$, where $\rho = n^{-1+\alpha}$ where $0 < \alpha \le 1$ is a positive constant.

LEMMA 5.5.   *Let x be a complex random variable with mean zero and variance one. Let $\mathbf{X} = (X_{1,1}^{(0)}, X_{1,1}^{(1)}, X_{2,1}^{(0)}, X_{2,1}^{(1)}, \ldots)$ be an array of $2n^2$ real random variables, where for each $1 \leq i, j \leq n$ we define $X_{i,j}^{(0)}$ and $X_{i,j}^{(1)}$ so that $X_{i,j}^{(0)} + \sqrt{-1}X_{i,j}^{(1)}$ is an i.i.d. copy of $x\mathbb{I}_\rho/\sqrt{\rho}$. Similarly, let $\mathbf{Y} = (Y_{1,1}^{(0)}, Y_{1,1}^{(1)}, Y_{2,1}^{(0)}, Y_{2,1}^{(1)}, \ldots)$ be another array of $2n^2$ real random variables, where for each $1 \leq i, j \leq n$ we define $Y_{i,j}^{(0)}$ and $Y_{i,j}^{(1)}$ so that $Y_{i,j}^{(0)} + \sqrt{-1}Y_{i,j}^{(1)}$ is an i.i.d. copy of x (thus, the $X_{i,j}^{(k)}$ are sparse versions of the $Y_{i,j}^{(k)}$, which are not sparse). Let $A_n(\mathbf{X})$ denote the n by n random matrix having $X_{i,j}^{(0)} + \sqrt{-1}X_{i,j}^{(1)}$ for the (i, j) entry, and similarly for $A_n(\mathbf{Y})$. Let $\mu_{(1/n)A_n(\mathbf{X})A_n(\mathbf{X})^*}$ and $\mu_{(1/n)A_n(\mathbf{Y})A_n(\mathbf{Y})^*}$ denote the ESDs of $\frac{1}{n}A_n(\mathbf{X})A_n(\mathbf{X})^*$ and $\frac{1}{n}A_n(\mathbf{Y})A_n(\mathbf{Y})^*$, respectively. Then $\mu_{(1/n)A_n(X)A_n(X)^*} - \mu_{(1/n)A_n(Y)A_n(Y)^*}$ converges in probability to zero as $n \to \infty$.*

PROOF.   Our approach will be applying [5], Theorem 1.1, in a similar way to [34], Lemma C.3.

Let $H_n(\mathbf{X}) := \begin{pmatrix} 0 & A_n(\mathbf{X})^*/\sqrt{n} \\ A_n(\mathbf{X})/\sqrt{n} & 0 \end{pmatrix}$.

Note that the eigenvalues of $H_n(\mathbf{X})$ with multiplicity are exactly the positive and negative square roots of the eigenvalues with multiplicity of $\frac{1}{n}A_n(\mathbf{X})A_n(\mathbf{X})^*$. Also, the same fact applies to $H_n(\mathbf{Y})$ and $\frac{1}{n}A_n(\mathbf{Y})A_n(\mathbf{Y})^*$. We will now follow the computation given in [5], Section 2.4. It is sufficient to show that $\mu_{H_n(\mathbf{X})} - \mu_{H_n(\mathbf{Y})}$ converges in probability to zero as $n \to \infty$.

Let $u, v \in \mathbb{R}$ with $v \neq 0$ and let $z = u + \sqrt{-1}v$. Define a function $f : \mathbb{R}^{2n^2} \to \mathbb{C}$ by

$$f(\mathbf{x}) = \frac{1}{2n}\operatorname{tr}\big((H_n(\mathbf{x}) - zI)^{-1}\big).$$

Here $\mathbf{x} = (x_{i,j}^{(k)})_{1 \leq i, j \leq n; k \in \{0,1\}}$, where $x_{i,j}^{(0)}$ corresponds to the real part (namely, $X_{i,j}^{(0)}$ or $Y_{i,j}^{(0)}$) and $x_{i,j}^{(1)}$ corresponds to the complex part (namely, $X_{i,j}^{(1)}$ or $Y_{i,j}^{(1)}$). We will show that for every fixed complex z with $\operatorname{Im}(z) = v \neq 0$, we have $\mathbb{E}(f(X)) - \mathbb{E}(f(F)) \to 0$ as $n \to \infty$, which implies that $\mu_{H_n(\mathbf{X})} - \mu_{H_n(\mathbf{Y})}$ converges in probability to zero as $n \to \infty$.

Define $G : \mathbb{R}^{2n^2} \to \mathbb{C}^{(2n)^2}$ by

$$G(\mathbf{x}) = \big(H_n(\mathbf{x}) - zI\big)^{-1}.$$

All eigenvalues of $H_n(\mathbf{x})$ are real, and thus all eigenvalues of $H(\mathbf{x}) - zI$ are nonzero (since $v \neq 0$). Thus, $G(\mathbf{x})$ is well defined. From the matrix inversion formula, each entry of $G(\mathbf{x})$ is a rational expression in $x_{i,j}^{(k)}$ for $1 \leq i, j \leq n$ and $k \in \{0, 1\}$. Thus G is infinitely differentiable in each coordinate $x_{i,j}^{(k)}$.

In the remainder of this section, we will use the shorthand G for $G(\mathbf{x})$ and the shorthand H for $H_n(\mathbf{x})$. Our goal is to apply the approach used by Chatterjee in [5], and we will first establish useful bounds on the partial derivatives of G.

Note that

$$(18) \qquad \frac{\partial G}{\partial x_{i,j}^{(k)}} = -G \frac{\partial H}{\partial x_{i,j}^{(k)}} G$$

[this can be seen by using the product rule and differentiating both sides of the equation $(H_n(\mathbf{x}) - zI)G = I$]. The following three formulas follow from (18) and the fact that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any two square matrices A and B, along with the fact that all higher partial derivatives of H are zero.

$$\frac{\partial f}{\partial x_{i,j}^{(k)}} = \frac{-1}{2n} \mathrm{tr}\left(\frac{\partial H}{\partial x_{i,j}^{(k)}} G^2 \right),$$

$$\frac{\partial^2 f}{\partial x_{i_1,j_1}^{(k_1)} \partial x_{i_2,j_2}^{(k_2)}} = \frac{1}{2n}\left(\mathrm{tr}\left(\frac{\partial H}{\partial x_{i_1,j_1}^{(k_1)}} G \frac{\partial H}{\partial x_{i_2,j_2}^{(k_2)}} G^2 \right) + \mathrm{tr}\left(\frac{\partial H}{\partial x_{i_2,j_2}^{(k_2)}} G \frac{\partial H}{\partial x_{i_1,j_1}^{(k_1)}} G^2 \right) \right),$$

$$\frac{\partial^3 f}{\partial x_{i_1,j_1}^{(k_1)} \partial x_{i_2,j_2}^{(k_2)} \partial x_{i_3,j_3}^{(k_3)}} = \frac{-1}{2n} \sum_{\sigma \in S_3} \mathrm{tr}\left(\frac{\partial H}{\partial x_{i_{\sigma(1)},j_{\sigma(1)}}^{(k_{\sigma(1)})}} G \frac{\partial H}{\partial x_{i_{\sigma(2)},j_{\sigma(2)}}^{(k_{\sigma(2)})}} G \frac{\partial H}{\partial x_{i_{\sigma(3)},j_{\sigma(3)}}^{(k_{\sigma(3)})}} G^2 \right),$$

where the last sum is over the six elements of S_3, the symmetric group on 3 letters.

As in [5], Section 2.4, we will use the following facts to bound the partial derivatives of f. Recall that for a matrix A, we define $\|A\|_2 := \mathrm{tr}(AA^*)$. Note that $|\mathrm{tr}(AB)| \le \|A\|_2 \|B\|_2$. Also, for A a k by k normal matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ and B any square matrix, we have $\max\{\|AB\|_2, \|BA\|_2\} \le (\max_{1 \le i \le k} \lambda_i)\|B\|_2$.

By the definition of G, it is clear that the absolute value of the largest eigenvalue of G is at most $|v|^{-1}$. Also, by the definition of H, it is clear that $\frac{\partial H}{\partial x_{i,j}^{(k)}}$ is the matrix having $(\sqrt{-1})^k n^{-1/2}$ for the $(n+i, j)$ entry, having $(-\sqrt{-1})^k n^{-1/2}$ for the $(j, n+i)$ entry and having zero for all other entries.

Thus, for all $1 \le i, j \le n$ and $k \in \{0, 1\}$, we have that

$$\left| \mathrm{tr}\left(\frac{\partial H}{\partial x_{i,j}^{(k)}} G^2 \right) \right| \le \left\| \frac{\partial H}{\partial x_{i,j}^{(k)}} \right\|_2 \|G^2\|_2$$

$$\le \sqrt{\frac{2}{n}} |v|^{-2} \|I\|_2$$

$$\le |v|^{-2}\sqrt{2},$$

and so $|\frac{\partial f}{\partial x_{i,j}^{(k)}}| < \frac{|v|^{-2}}{n}$.

By similar means, we can compute

$$\left| \operatorname{tr}\left(\frac{\partial H}{\partial x_{i_1,j_1}^{(k_1)}} G \frac{\partial H}{\partial x_{i_2,j_2}^{(k_2)}} G^2 \right) \right| \leq \left\| \frac{\partial H}{\partial x_{i_1,j_1}^{(k_1)}} \right\|_2 \left\| G \frac{\partial H}{\partial x_{i_2,j_2}^{(k_2)}} G^2 \right\|_2$$

$$\leq \sqrt{\frac{2}{n}} |v|^{-3} \sqrt{\frac{2}{n}}$$

$$\leq \frac{2|v|^{-3}}{n},$$

which shows that $|\frac{\partial^2 f}{\partial x_{i_1,j_1}^{(k_1)} \partial x_{i_2,j_2}^{(k_2)}}| \leq \frac{2|v|^{-3}}{n^2}$; and

$$\left| \operatorname{tr}\left(\frac{\partial H}{\partial x_{i_1,j_1}^{(k_1)}} G \frac{\partial H}{\partial x_{i_2,j_2}^{(k_2)}} G \frac{\partial H}{\partial x_{i_3,j_3}^{(k_3)}} G^2 \right) \right| \leq \left\| \frac{\partial H}{\partial x_{i_1,j_1}^{(k_1)}} \right\|_2 \left\| G \frac{\partial H}{\partial x_{i_2,j_2}^{(k_2)}} G \frac{\partial H}{\partial x_{i_3,j_3}^{(k_3)}} G^2 \right\|_2$$

$$\leq \sqrt{\frac{2}{n}} |v|^{-1} \frac{1}{\sqrt{n}} |v|^{-3} \sqrt{\frac{2}{n}}$$

$$\leq \frac{2|v|^{-4}}{n^{3/2}},$$

which shows that $|\frac{\partial^3 f}{\partial x_{i_1,j_1}^{(k_1)} \partial x_{i_2,j_2}^{(k_2)} \partial x_{i_3,j_3}^{(k_3)}}| \leq \frac{6|v|^{-4}}{n^{5/2}}$.

We will now apply a complex version of the main theorem from [5]. First, we need the following definitions for a function $h : \mathbb{R}^N \to \mathbb{C}$. We define the *derivative-product degree with respect to h* of a monomial of partial derivatives of h to be the sum of the number of partial derivatives taken in each factor when the monomial is written as a product of linear terms. We will use *derivative-product degree* when the function h is understood. For example, the derivative-product degree of $(\frac{\partial h}{\partial x_{i,j}^{(k)}})^3 = (\frac{\partial h}{\partial x_{i,j}^{(k)}})(\frac{\partial h}{\partial x_{i,j}^{(k)}})(\frac{\partial h}{\partial x_{i,j}^{(k)}})$ is 3, and the derivative product degree

of $(\frac{\partial^3 h}{\partial x_{i_1,j_1}^{(k_1)} \partial x_{i_2,j_2}^{(k_2)} \partial x_{i_3,j_3}^{(k_3)}})^2 (\frac{\partial^2 h}{\partial x_{i_3,j_3}^{(k_3)} \partial x_{i_4,j_4}^{(k_4)}}) = (\frac{\partial^3 h}{\partial x_{i_1,j_1}^{(k_1)} \partial x_{i_2,j_2}^{(k_2)} \partial x_{i_3,j_3}^{(k_3)}})(\frac{\partial^3 h}{\partial x_{i_1,j_1}^{(k_1)} \partial x_{i_2,j_2}^{(k_2)} \partial x_{i_3,j_3}^{(k_3)}}) \times$

$(\frac{\partial^2 h}{\partial x_{i_3,j_3}^{(k_3)} \partial x_{i_4,j_4}^{(k_4)}})$ is 8. Define two quantities as follows:

$$\lambda_2(h) := \sup_{x \in \mathbb{R}^N} \{ \gamma(x) : \gamma \text{ has derivative-product degree 2 with respect to } h \}$$

and let

$$\lambda_3(h) := \sup_{x \in \mathbb{R}^N} \{ \gamma(x) : \gamma \text{ has derivative-product degree 3 with respect to } h \}.$$

THEOREM 5.6.   *Let N be a positive even integer, let $\mathbf{X} = (X_1, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, \ldots, Y_N)$ be lists of real-valued random variables such that for $1 \le \ell \le N/2$, the random variables $X_{2\ell-1}$ and $X_{2\ell}$ are each independent of all X_j such that $1 \le j \le N$ and $j \notin \{2\ell - 1, 2\ell\}$, and similarly the random variables $Y_{2\ell-1}$ and $Y_{2\ell}$ are each independent of all Y_j such that $1 \le j \le N$ and $j \notin \{2\ell - 1, 2\ell\}$. Assume further that*

$$
\begin{aligned}
\mathbb{E}(X_j) &= \mathbb{E}(Y_j) && \text{for all } 1 \le j \le N, \\
\mathbb{E}(X_j^2) &= \mathbb{E}(Y_j^2) && \text{for all } 1 \le j \le N, \\
\mathbb{E}(X_{2\ell-1} X_{2\ell}) &= \mathbb{E}(Y_{2\ell-1} Y_{2\ell}) && \text{for all } 1 \le \ell \le N/2.
\end{aligned}
$$
(19)

Let $h : \mathbb{R}^N \to \mathbb{R}$ have continuous partial derivatives of order 1, 2 *and* 3, *including mixed partial derivatives. If we set $U = h(\mathbf{X})$ and $V = h(\mathbf{Y})$, then for any thrice differentiable $g : \mathbb{R} \to \mathbb{R}$ and any $K > 0$,*

$$
\begin{aligned}
&|\mathbb{E}g(U) - \mathbb{E}g(V)| \\
&\le C_1(g)\lambda_2(h) \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \big(\mathbb{E}(X_i^2 + |X_i X_{i+1}| + X_{i+1}^2; |X_i| + |X_{i+1}| > K) \\
&\qquad\qquad\qquad + \mathbb{E}(Y_i^2 + |Y_i Y_{i+1}| + Y_{i+1}^2; |Y_i| + |Y_{i+1}| > K) \big) \\
&\quad + C_2(g)\lambda_3(h) \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \big(\mathbb{E}(|X_i|^3 + X_i^2|X_{i+1}| + |X_i|X_{i+1}^2 + |X_{i+1}|^3; \\
&\qquad\qquad\qquad\qquad\qquad\qquad |X_i| + |X_{i+1}| \le K) \\
&\qquad\qquad\qquad + \mathbb{E}(|Y_i|^3 + Y_i^2|Y_{i+1}| + |Y_i|Y_{i+1}^2 + |Y_{i+1}|^3; \\
&\qquad\qquad\qquad\qquad\qquad\qquad |Y_i| + |Y_{i+1}| \le K) \big),
\end{aligned}
$$

where $C_1(g) = \|g'\|_\infty + \|g''\|_\infty$ and $C_2(g) = \|g'\|_\infty + 3\|g''\|_\infty + \|g'''\|_\infty$.

We prove Theorem 5.6 in the Appendix.

Theorem 5.6 requires h to be a real-valued function; thus we will apply Theorem 5.6 to $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ separately. We will give the application to $\mathrm{Re}(f)$ below, noting that the same argument applies with Im replacing Re.

Given $g : \mathbb{R} \to \mathbb{R}$ a thrice differentiable function, set $U = \mathrm{Re}(f(\mathbf{X}))$ and $V = \mathrm{Re}(f(\mathbf{Y}))$, where $\mathbf{X}$ and $\mathbf{Y}$ are as in the statement of Lemma 5.5 (notationally, set $N = 2n^2$ and define X_ℓ by $X_{1+2n(i-1)+2(j-1)+k} := X_{i,j}^{(k)}$). Note that from the assumption in Lemma 5.5 that the $X_{i,j}^{(k)}$ are sparse versions of the $Y_{i,j}^{(k)}$, the hypotheses in (19) are automatically satisfied. Also, the independence hypotheses in

Theorem 5.6 follow from the definitions of $X_{i,j}^{(k)}$ and $Y_{i,j}^{(k)}$ in Lemma 5.5. Finally, noting that $\lambda_r(\operatorname{Re} f) \leq \lambda_r(f)$, and noting that for our function f we have

$$\lambda_2(f) = \sup\left\{\frac{|v|^{-4}}{n^2}, \frac{2|v|^{-3}}{n^2}\right\}$$

and

$$\lambda_3(f) = \sup\left\{\frac{|v|^{-6}}{n^3}, \frac{2|v|^{-5}}{n^3}, \frac{6|v|^{-4}}{n^{5/2}}\right\},$$

we may apply Theorem 5.6 to get

$$|\mathbb{E}g(U) - \mathbb{E}g(V)|$$

$$\leq C_1(g)\lambda_2(h)$$

(20)

$$\times \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathbb{E}\left((X_{i,j}^{(0)})^2 + |X_{i,j}^{(0)}X_{i,j}^{(1)}| + (X_{i,j}^{(1)})^2; |X_{i,j}^{(0)}| + |X_{i,j}^{(1)}| > K\right)\right.$$

$$\left. + \mathbb{E}\left((Y_{i,j}^{(0)})^2 + |Y_{i,j}^{(0)}Y_{i,j}^{(1)}| + (Y_{i,j}^{(1)})^2; |Y_{i,j}^{(0)}| + |Y_{i,j}^{(1)}| > K\right)\right)$$

(21) $$+ C_2(g)\lambda_3(h)$$

$$\times \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathbb{E}\left(|X_{i,j}^{(0)}|^3 + (X_{i,j}^{(0)})^2|X_{i,j}^{(1)}| + |X_{i,j}^{(0)}|(X_{i,j}^{(1)})^2 + |X_{i,j}^{(1)}|^3;\right.\right.$$

$$\left. |X_{i,j}^{(0)}| + |X_{i,j}^{(1)}| \leq K\right)$$

$$+ \mathbb{E}\left(|Y_{i,j}^{(0)}|^3 + (Y_{i,j}^{(0)})^2|Y_{i,j}^{(1)}| + |Y_{i,j}^{(0)}|(Y_{i,j}^{(1)})^2 + |Y_{i,j}^{(1)}|^3;\right.$$

$$\left.\left. |Y_{i,j}^{(0)}| + |Y_{i,j}^{(1)}| \leq K\right)\right).$$

Choose $K = \varepsilon\sqrt{n}$, where $\varepsilon > 0$ is a small positive constant. The double-sum term in (21) is bounded by ε times a constant depending only on g and v [here, we used that $\mathbb{E}(|X|^3; |X| \leq K) \leq K\mathbb{E}(X^2)$ for any real random variable X]. Also, using the fact that $ab \leq \frac{1}{2}(a^2 + b^2)$ for any positive real numbers a and b, the double-sum term in (20) is bounded by another constant depending only on g and v times the quantity

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left((X_{i,j}^{(0)})^2 + (X_{i,j}^{(1)})^2; |X_{i,j}^{(0)}| + |X_{i,j}^{(1)}| > \varepsilon\sqrt{n}\right)$$

$$+ \mathbb{E}\left((Y_{i,j}^{(0)})^2 + (Y_{i,j}^{(1)})^2; |Y_{i,j}^{(0)}| + |Y_{i,j}^{(1)}| > \varepsilon\sqrt{n}\right).$$

Since the random variables $Y_{i,j}^{(k)}$ do not change with n, it is clear from monotone convergence that $\mathbb{E}((Y_{i,j}^{(0)})^2 + (Y_{i,j}^{(1)})^2; |Y_{i,j}^{(0)}| + |Y_{i,j}^{(1)}| > \varepsilon\sqrt{n}) \to 0$ as $n \to \infty$.

Thus, it is sufficient to show that $\mathbb{E}((X_{i,j}^{(k)})^2; |X_{i,j}^{(k)}| > \varepsilon\sqrt{n}) \to 0$ as $n \to \infty$. Recall that $X_{i,j}^{(0)} + \sqrt{-1}X_{i,j}^{(1)}$ is an i.i.d. copy of $X\mathbb{I}_\rho/\sqrt{\rho}$, where X is a complex random variable with mean zero and variance one, and note that $|X_{i,j}^{(0)}| + |X_{i,j}^{(1)}| > \varepsilon\sqrt{n}$ implies that $\sqrt{|X_{i,j}^{(0)}|^2 + |X_{i,j}^{(1)}|^2} > \varepsilon\sqrt{2n/3}$. We have that

$$\mathbb{E}\left(\left|\frac{X\mathbb{I}_\rho}{\sqrt{\rho}}\right|^2; \left|\frac{X\mathbb{I}_\rho}{\sqrt{\rho}}\right| > \varepsilon\sqrt{2n/3}\right)$$

$$\leq \mathbb{E}\left(\left|\frac{X\mathbb{I}_\rho}{\sqrt{\rho}}\right|^2; |X| > \varepsilon\sqrt{2\rho n/3}\right)$$

$$= \mathbb{E}(|X|^2; |X| > \varepsilon\sqrt{2\rho n/3}),$$

where the last equality follows by the independence of $\mathbb{I}_\rho$ and X. Finally, by monotone convergence again, we see that $\mathbb{E}(|X|^2; |X| > \varepsilon\sqrt{2\rho n/3}) \to 0$ as $n \to \infty$, completing the proof. $\square$

APPENDIX: A COMPLEX VERSION OF CHATTERJEE'S INVARIANCE THEOREM

In this Appendix, we prove Theorem 5.6, which is a version of [5], Theorem 1.1, for the complex numbers. In order to prove the result in the complex case, we treat the real and complex parts of each random variable as separate, possibly dependent real random variables. The fact that the real and complex parts of a random variable may depend on each other introduces some complications. Our approach is modeled on that in [5], with the main differences being that we use the Lindeberg argument on pairs of random variables, rather than on single random variables, and also we also use two-dimensional Taylor expansions.

PROOF OF THEOREM 5.6.   Let $\Psi := g \circ h$, which is a function from $\mathbb{R}^N \to \mathbb{R}$. Later in the proof we will apply the two-dimensional version of Taylor's theorem to Ψ, and so to start we will establish bounds on the partial derivatives of Ψ. We will use the notation $\partial_{(i_1,i_2,\ldots,i_k)}\Psi$ as shorthand for $\frac{\partial^k \Psi}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}}$. Note that the order of the coordinates $(i_1, i_2, \ldots, i_k)$ is unimportant for $1 \leq k \leq 3$ since Ψ has continuous partial derivatives (including mixed partials) by assumptions on h and g.

Note that $\partial_i \Psi(x) = g'(h(x))\,\partial_i h(x)$, and so taking further partial derivatives one can compute that

$$\partial_{ij}\Psi(x) = g''(h(x))\,\partial_i h(x)\,\partial_j h(x) + g'(h(x))\,\partial_{ij}h(x).$$

Thus, $\sup_{x \in \mathbb{R}^N 1 \leq i,j \leq N} |\partial_{ij}\Psi(x)| \leq C_1(g)\lambda_2(h)$.

Taking further partial derivatives, one can compute that

$$
\begin{aligned}
\partial_{i,j,k} \Psi(x) = {}& g'''(h(x))\, \partial_i h(x)\, \partial_j h(x)\, \partial_k h(x) \\
& + g''(h(x))\, \partial_{ik} h(x)\, \partial_j h(x) + g''(h(x))\, \partial_i h(x)\, \partial_{jk} h(x) \\
& + g''(h(x))\, \partial_k h(x)\, \partial_{ij} h(x) + g'(h(x))\, \partial_{ijk} h(x).
\end{aligned}
$$

Thus, $\sup_{x \in \mathbb{R}^N} {}_{1 \le i,j,k \le N} \le C_2(g)\lambda_3(h)$.

For $1 \le i \le N$ and i odd, define

$$
\begin{aligned}
Z_i &:= (X_1, \dots, X_{i-1}, X_i, X_{i+1}, Y_{i+2}, \dots, Y_N), \\
W_i &:= (X_1, \dots, X_{i-1}, 0, 0, Y_{i+2}, \dots, Y_N)
\end{aligned}
$$

with $Z_{-1} := (Y_1, \dots, Y_N)$ and $Z_{N-1} := (X_1, \dots, X_N)$. Also, for $1 \le i \le N$ and i odd, define

$$
\begin{aligned}
R_i := {}& \Psi(Z_i) - \Psi(W_i) - X_i\, \partial_i \Psi(W_i) - X_{i+1}\, \partial_{i+1} \Psi(W_i) \\
& - \frac{X_i^2}{2}\, \partial_{ii} \Psi(W_i) - \frac{X_{i+1}^2}{2}\, \partial_{i+1,i+1} \Psi(W_i) \\
& - X_i X_{i+1}\, \partial_{i,i+1} \Psi(W_i), \\
T_i := {}& \Psi(Z_{i-2}) - \Psi(W_i) - Y_i\, \partial_i \Psi(W_i) - Y_{i+1}\, \partial_{i+1} \Psi(W_i) \\
& - \frac{Y_i^2}{2}\, \partial_{ii} \Psi(W_i) - \frac{Y_{i+1}^2}{2}\, \partial_{i+1,i+1} \Psi(W_i) \\
& - Y_i Y_{i+1}\, \partial_{i,i+1} \Psi(W_i).
\end{aligned}
$$

Note that by Taylor's theorem in two dimensions and bounds on the partials of Ψ, we have for odd i that

$$
\begin{aligned}
|R_i| &\le C_1(g)\lambda_2(h)(X_i^2 + |X_i X_{i+1}| + X_{i+1}^2), \\
|R_i| &\le C_1(g)\lambda_2(h)(Y_i^2 + |Y_i Y_{i+1}| + Y_{i+1}^2)
\end{aligned}
$$

using second order bounds, and that

$$
\begin{aligned}
|R_i| &\le C_2(g)\lambda_3(h)(|X_i|^3 + X_i^2|X_{i+1}| + |X_i|X_{i+1}^2 + |X_{i+1}|^3), \\
|T_i| &\le C_2(g)\lambda_3(h)(|Y_i|^3 + Y_i^2|Y_{i+1}| + |Y_i|Y_{i+1}^2 + |Y_{i+1}|^3)
\end{aligned}
$$

using third order bounds.

We now make use of the Lindeberg principle, writing $|\mathbb{E}(g(U)) - \mathbb{E}(g(V))|$ in terms of a telescoping sum involving Z_i.

$$
\begin{aligned}
& |\mathbb{E}(g(U)) - \mathbb{E}(g(V))| \\
& = \left| \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \mathbb{E}\big(\Psi(Z_i) - \Psi(Z_{i-2})\big) \right|
\end{aligned}
$$

$$= \left| \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \mathbb{E}\Bigg(\Psi(W_i) + X_i\, \partial_i \Psi(W_i) + X_{i+1}\, \partial_{i+1} \Psi(W_i) \right.$$

$$+ \frac{X_i^2}{2}\, \partial_{ii} \Psi(W_i) + \frac{X_{i+1}^2}{2}\, \partial_{i+1,i+1} \Psi(W_i)$$

$$+ X_i X_{i+1}\, \partial_{i,i+1} \Psi(W_i) + R_i \Bigg)$$

$$- \mathbb{E}\Bigg(\Psi(W_i) + Y_i\, \partial_i \Psi(W_i) + Y_{i+1}\, \partial_{i+1} \Psi(W_i)$$

$$+ \frac{Y_i^2}{2}\, \partial_{ii} \Psi(W_i) + \frac{Y_{i+1}^2}{2}\, \partial_{i+1,i+1} \Psi(W_i)$$

$$\left. + Y_i Y_{i+1}\, \partial_{i,i+1} \Psi(W_i) + T_i \Bigg) \right|$$

$$= \left| \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \mathbb{E}(R_i) - \mathbb{E}(T_i) \right|.$$

Note that in the above there is lots of cancellation, for example,

$$\mathbb{E} X_i X_{i+1}\, \partial_{i,i+1} \Psi(W_i) - \mathbb{E} Y_i Y_{i+1}\, \partial_{i,i+1} \Psi(W_i) = 0$$

by the independence assumptions along with the assumption that $\mathbb{E} X_i X_{i+1} = \mathbb{E} Y_i Y_{i+1}$.

To complete the proof we bound R_i and T_i using second order bounds when they are small and using third order bounds when they are large, arriving at

$$\left| \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \mathbb{E}(R_i) - \mathbb{E}(T_i) \right|$$

$$\leq C_1(g)\lambda_2(h) \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \big(\mathbb{E}(X_i^2 + |X_i X_{i+1}| + X_{i+1}^2; |X_i| + |X_{i+1}| > K)$$

$$+ \mathbb{E}(Y_i^2 + |Y_i Y_{i+1}| + Y_{i+1}^2; |Y_i| + |Y_{i+1}| > K) \big)$$

$$+ C_2(g)\lambda_3(h) \sum_{\substack{i=1 \\ i \text{ odd}}}^{N-1} \big(\mathbb{E}(|X_i|^3 + X_i^2 |X_{i+1}| + |X_i| X_{i+1}^2 + |X_{i+1}|^3;$$

$$|X_i| + |X_{i+1}| \leq K)$$

$$+ \mathbb{E}(|Y_i|^3 + Y_i^2 |Y_{i+1}| + |Y_i| Y_{i+1}^2 + |Y_{i+1}|^3;$$

$$|Y_i| + |Y_{i+1}| \leq K) \big). \qquad \square$$

REFERENCES

[1] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. MR2567175

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

[3] BORDENAVE, C., CAPUTO, P. and CHAFAÏ, D. (2010). Spectrum of non-Hermitian heavy tailed random matrices. Available at arXiv:1006.1713v2.

[4] BORDENAVE, C., CAPUTO, P. and CHAFAÏ, D. (2010). Circular law theorem for random markov matrices. Available at arXiv:0808.1502v3.

[5] CHATTERJEE, S. (2005). A simple invariance theorem. Available at arXiv:math/0508213v1.

[6] CHATTERJEE, S., DIACONIS, P. and SLY, A. (2010). Properties of uniform doubly stochastic matrices. Available at arXiv:1010.6136v1.

[7] 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

[8] DUMITRIU, I. and PAL, S. (2010). Sparse regular random graphs: Spectral density and eigenvectors. Available at arXiv:0910.5306v3.

[9] DURAS, M. M., SOKALSKI, K. and SUŁKOWSKI, P. (1997). Statistical properties of sparse Gaussian random symmetrical ensemble. *Acta Phys. Polon. B* **28** 1023–1038. MR1456787

[10] DURRETT, R. (2010). *Probability*: *Theory and Examples*, 4th ed. Cambridge Univ. Press, Cambridge. MR2722836

[11] EDELMAN, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** 543–560. MR0964668

[12] FYODOROV, Y. V., SOMMERS, H.-J. and KHORUZHENKO, B. A. (1998). Universality in the random matrix spectra in the regime of weak non-Hermiticity: Classical and quantum chaos. *Ann. Inst. H. Poincaré Phys. Théor.* **68** 449–489. MR1634312

[13] GIRKO, V. L. (1984). The circular law. *Teor. Veroyatn. Primen.* **29** 669–679. MR0773436

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

[15] GÖTZE, F. and TIKHOMIROV, A. (2007). On the circular law. Available at arXiv:math/0702386v1.

[16] GÖTZE, F. and TIKHOMIROV, A. (2010). The circular law for random matrices. *Ann. Probab.* **38** 1444–1491. MR2663633

[17] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[18] KÜHN, R. (2008). Spectra of sparse random matrices. *J. Phys. A* **41** 295002, 21. MR2455271

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

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

[21] MIRLIN, A. D. and FYODOROV, Y. V. (1991). Universality of level correlation function of sparse random matrices. *J. Phys. A* **24** 2273–2286. MR1118532

[22] NAGAO, T. and TANAKA, T. (2007). Spectral density of sparse sample covariance matrices. *J. Phys. A* **40** 4973–4987. MR2341055

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

[24] RODGERS, G. J. and BRAY, A. J. (1988). Density of states of a sparse random matrix. *Phys. Rev. B* (3) **37** 3557–3562. MR0932406

[25] ROGERS, T. (2010). Universal sum and product rules for random matrices. *J. Math. Phys.* **51** 093304, 15. MR2742824

[26] ROGERS, T. and PÉREZ CASTILLO, I. (2009). Cavity approach to the spectral density of non-Hermitian sparse matrices. *Phys. Rev. E* (3) **79** 012101.

[27] ROGERS, T., PÉREZ CASTILLO, I., KÜHN, R. and TAKEDA, K. (2008). Cavity approach to the spectral density of sparse symmetric random matrices. *Phys. Rev. E* (3) **78** 031116, 6. MR2496088

[28] SATO, S. and KOBAYASHI, K. (1976/77). Asymptotic distribution of eigenvalues and degeneration of sparse random matrices. *Bull. Math. Statist.* **17** 83–99. MR0461616

[29] SEMERJIAN, G. and CUGLIANDOLO, L. F. (2002). Sparse random matrices: The eigenvalue spectrum revisited. *J. Phys. A* **35** 4837–4851. MR1916090

[30] SODIN, S. (2009). The Tracy–Widom law for some sparse random matrices. *J. Stat. Phys.* **136** 834–841. MR2545551

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

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

[33] TAO, T. and VU, V. (2009). From the Littlewood–Offord problem to the circular law: Universality of the spectral distribution of random matrices. *Bull. Amer. Math. Soc.* (*N.S.*) **46** 377–396. MR2507275

[34] TAO, T. and VU, V. (2010). Random matrices: Universality of ESDs and the circular law. With an appendix by Manjunath Krishnapur. *Ann. Probab.* **38** 2023–2065. MR2722794

[35] TRAN, L., VU, V. and WANG, K. (2010). Sparse random graphs: Eigenvalues and eigenvectors. Available at arXiv:1011.6646v1.

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

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
BUILDING 380
STANFORD, CALIFORNIA 94305
USA
E-MAIL: pmwood@math.stanford.edu