BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 46, Number 3, July 2009, Pages 377–396
S 0273-0979(09)01252-X
Article electronically published on February 24, 2009

FROM THE LITTLEWOOD-OFFORD PROBLEM TO THE CIRCULAR LAW: UNIVERSALITY OF THE SPECTRAL DISTRIBUTION OF RANDOM MATRICES

TERENCE TAO AND VAN VU

ABSTRACT. The famous *circular law* asserts that if M_n is an $n \times n$ matrix with iid complex entries of mean zero and unit variance, then the empirical spectral distribution of the normalized matrix $\frac{1}{\sqrt{n}}M_n$ converges both in probability and almost surely to the uniform distribution on the unit disk $\{z \in \mathbf{C} : |z| \leq 1\}$. After a long sequence of partial results that verified this law under additional assumptions on the distribution of the entries, the circular law is now known to be true for arbitrary distributions with mean zero and unit variance. In this survey we describe some of the key ingredients used in the establishment of the circular law at this level of generality, in particular recent advances in understanding the Littlewood-Offord problem and its inverse.

1. ESD OF RANDOM MATRICES

For an $n \times n$ matrix A_n with complex entries, let

$$\mu_{A_n} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i}$$

be the *empirical spectral distribution* (ESD) of its eigenvalues $\lambda_i \in \mathbf{C}, i = 1, \dots n$ (counting multiplicity); thus, for instance

$$\mu_{A_n}(\{z \in \mathbf{C} | \operatorname{Re} z \leq s; \operatorname{Im} z \leq t\}) = \frac{1}{n}|\{1 \leq i \leq n : \operatorname{Re}\lambda_i \leq s; \operatorname{Im}\lambda_i \leq t\}|$$

for any $s, t \in \mathbf{R}$ (we use $|A|$ to denote the cardinality of a finite set A), and

$$\int f \, d\mu_{A_n} = \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i)$$

for any continuous compactly supported f. Clearly, μ_{A_n} is a discrete probability measure on $\mathbf{C}$.

A fundamental problem in the theory of random matrices is to compute the limiting distribution of the ESD μ_{A_n} of a sequence of random matrices A_n with sizes tending to infinity [34, 4]. In what follows, we consider normalized random matrices of the form $A_n = \frac{1}{\sqrt{n}}M_n$, where $M_n = (\mathrm{x}_{ij})_{1 \leq i,j \leq n}$ has entries that are

Received by the editors October 16, 2008, and, in revised form, January 1, 2009.

2000 *Mathematics Subject Classification.* Primary 15A52, 60G50.

The first author is supported by NSF grant CCF-0649473 and a grant from the MacArthur Foundation.

The second author is supported by an NSF Career Grant 0635606.

©2009 American Mathematical Society
Reverts to public domain 28 years from publication

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

iid random variables $\mathrm{x}_{ij} \equiv \mathrm{x}$. Such matrices have been studied at least as far back as Wishart [58] (see [34, 4] for more discussion).

One of the first limiting distribution results is the famous semi-circle law of Wigner [57]. Motivated by research in nuclear physics, Wigner studied Hermitian random matrices with (upper triangular) entries being iid random variables with mean zero and variance one. In the Hermitian case, of course, the ESD is supported on the real line $\mathbf{R}$. He proved that the expected ESD of a normalized $n \times n$ Hermitian matrix $\frac{1}{\sqrt{n}} M_n$, where $M_n = (\mathrm{x}_{ij})_{1 \leq i,j \leq n}$ has iid Gaussian entries $\mathrm{x}_{ij} \equiv N(0,1)$, converges in the sense of probability measures[1] to the semi-circle distribution

$$\frac{1}{2\pi} 1_{[-2,2]}(x) \sqrt{4 - x^2}\; dx \tag{1}$$

on the real line, where 1_E denotes the indicator function of a set E.

Theorem 1.1 (Semi-circular law for the Gaussian ensemble [57])**.** *Let M_n be an $n \times n$ random Hermitian matrix whose entries are iid Gaussian variables with mean* 0 *and variance* 1*. Then, with probability one, the ESD of $\frac{1}{\sqrt{n}} M_n$ converges in the sense of probability measures to the semi-circle law* (1)*.*

Henceforth we shall say that a sequence μ_n of random probability measures converges *strongly* to a deterministic probability measure μ if, with probability one, μ_n converges in the sense of probability measures to μ. We also say that μ_n converges *weakly* to μ if for every continuous compactly supported f, $\int f\, d\mu_n$ converges in probability to $\int f\, d\mu$; thus, $\mathbf{P}(|\int f\, d\mu_n - \int f\, d\mu| > \varepsilon) \to 0$ as $n \to \infty$ for each $\varepsilon > 0$. Of course, strong convergence implies weak convergence; thus, for instance in Theorem 1.1, $\mu_{\frac{1}{\sqrt{n}} M_n}$ converges both weakly and strongly to the semi-circle law.

Wigner also proved similar results for various other distributions, such as the Bernoulli distribution (in which each x_{ij} equals $+1$ with probability $1/2$ and -1 with probability $1/2$). His work has been extended and strengthened in several aspects [1, 2, 36]. The most general form was proved by Pastur [36]:

Theorem 1.2 (Semi-circular law [36])**.** *Let M_n be an $n \times n$ random Hermitian matrix whose entries are iid complex random variables with mean* 0 *and variance* 1*. Then ESD of $\frac{1}{\sqrt{n}} M_n$ converges (in both the strong and weak senses) to the semi-circle law.*

The situation with non-Hermitian matrices is much more complicated due to the presence of *pseudospectrum*[2] that can potentially make the ESD quite unstable with respect to perturbations. The non-Hermitian variant of this theorem, the

[1] We say that a collection μ_n of probability measures converges to a limit μ if one has $\int f\, d\mu_n \to \int f\, d\mu$ for every continuous compactly supported function f, or equivalently if $\mu(\{z \in \mathbf{C} | \mathrm{Re}\, z \leq s; \mathrm{Im}\, z \leq t\})$ converges to $\mu(\{z \in \mathbf{C} | \mathrm{Re}\, z \leq s; \mathrm{Im}\, z \leq t\})$ for all s, t.

[2] Informally, we say that a complex number z lies in the pseudospectrum of a square matrix A if $(A - zI)^{-1}$ is large (or undefined). If z lies in the pseudospectrum, then small perturbations of A can potentially cause z to fall into the spectrum of A, even if it is initially far away from this spectrum. Thus, whenever one has pseudospectrum far away from the actual spectrum, the actual distribution of eigenvalues can depend very sensitively (in the worst case) on the coefficients of A. Of course, our matrices are random rather than worst case, and so we expect the most dangerous effects of pseudospectrum to be avoided; but this of course requires some analytical effort to establish, and deterministic techniques (e.g., truncation) should be used with extreme caution, since they are likely to break down in the worst case.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Circular Law Conjecture, has been raised since the 1950's (see Chapter 10 of [4] or the introduction of [3])

Conjecture 1.3 (Circular law)**.** *Let M_n be the $n \times n$ random matrix whose entries are iid complex random variables with mean* 0 *and variance* 1*. Then the ESD of $\frac{1}{\sqrt{n}} M_n$ converges (in both the strong and weak senses) to the uniform distribution $\mu := \frac{1}{\pi} 1_{|z| \leq 1} dz$ on the unit disk $\{z \in \mathbf{C} : |z| \leq 1\}$.*

The numerical evidence for this conjecture is extremely strong (see, e.g., Figure 1). However, there are significant difficulties in establishing this conjecture rigorously, not least of which is the fact that the main techniques used to handle Hermitian matrices (such as moment methods and truncation) cannot be applied to the non-Hermitian model (see [4, Chapter 10] for a detailed discussion). Nevertheless, the conjecture has been worked on intensively for many decades. The circular law was verified for the complex Gaussian distribution in [34] and the real Gaussian distribution in [12]. An approach to attack the general case was introduced in [18], leading to a resolution of the strong circular law for continuous distributions with bounded sixth moment in [3]. The sixth moment hypothesis in [3] was lowered to $(2+\eta)$-th moment for any $\eta > 0$ in [4]. The removal of the hypothesis of continuous distribution required some new ideas. In [21] the weak circular law for (possibly discrete) distributions with sub-Gaussian moment was established, with the sub-Gaussian condition relaxed to a fourth moment condition in [35] (see also [19] for an earlier result of similar nature), and then to $(2 + \eta)$-th moment in [22]. Shortly before this last result, the strong circular law assuming $(2 + \eta)$-th moment was established in [54]. Finally, in a recent paper [55], the authors proved this conjecture (in both strong and weak forms) in full generality. In fact, we obtained this result as a consequence of a more general theorem, presented in the next section.

2. Universality

An easy case of Conjecture 1.3 is when the entries x_{ij} of M_n are iid complex Gaussian. In this case there is the following precise formula for the joint density function of the eigenvalues, due to Ginibre [17] (see also [34], [25] for more discussion of this formula):

$$p(\lambda_1, \cdots, \lambda_n) = c_n \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{n} e^{-n|\lambda_i|^2}. \tag{2}$$

From here one can verify the conjecture in this case by a direct calculation. This was first done by Mehta and also Silverstein in the 1960's:

Theorem 2.1 (Circular law for Gaussian matrices [34])**.** *Let M_n be an $n \times n$ random matrix whose entries are iid complex Gaussian variables with mean* 0 *and variance* 1*. Then, with probability one, the ESD of $\frac{1}{\sqrt{n}} M_n$ tends to the circular law.*

A similar result for the real Gaussian ensemble was established in [12]. These methods rely heavily on the strong symmetry properties of such ensembles (in particular, the invariance of such ensembles with respect to large matrix groups such as $O(n)$ or $U(n)$) in order to perform explicit algebraic computations, and they do not extend directly to more combinatorial ensembles, such as the Bernoulli ensemble.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

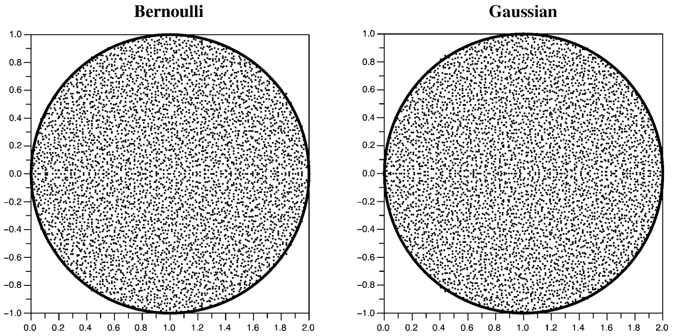

FIGURE 1. Eigenvalue plots of two randomly generated 5000×5000 matrices. On the left, each entry was an iid Bernoulli random variable, taking the values $+1$ and -1 each with probability $1/2$. On the right, each entry was an iid Gaussian normal random variable, with probability density function is $\frac{1}{\sqrt{2\pi}}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

The above-mentioned results and conjectures can be viewed as examples of a general phenomenon in probablity and mathematical physics, namely, that global information about a large random system (such as limiting distributions) does not depend on the particular distribution of the particles. This is often referred to as the *universality* phenomenon (see, e.g., [9]). The most famous example of this phenomenon is perhaps the central limit theorem.

In view of the universality phenomenon, one can see that Conjecture 1.3 generalizes Theorem 2.1 in the same way that Theorem 1.2 generalizes Theorem 1.1.

A demonstration of the circular law for the Bernoulli and the Gaussian case appears[3] in Figure 1.

The universality phenomenon seems to hold even for more general models of random matrices, as demonstrated by Figures 2 and 3.

This evidence suggests that the asymptotic shape of the ESD depends only on the mean and the variance of each entry in the matirx. As mentioend earlier, the main result of [55] (building on a large number of previous results) gives a rigorous proof of this phenomenon in full generality.

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

Theorem 2.2 (Universality principle)**.** *Let* x *and* y *be complex random variables with zero mean and unit variance. Let* $X_n = (\mathrm{x}_{ij})_{1 \le i,j \le n}$ *and* $Y_n := (\mathrm{y}_{ij})_{1 \le i,j \le n}$ *be* $n \times n$ *random matrices whose entries* x_{ij}, y_{ij} *are iid copies of* x *and* y*, respectively. For each* n*, let* M_n *be a deterministic* $n \times n$ *matrix satisfying*

$$\sup_n \frac{1}{n^2} \|M_n\|_F^2 < \infty. \tag{3}$$

Let $A_n := M_n + X_n$ *and* $B_n := M_n + Y_n$*. Then* $\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n}$ *converges weakly to zero. Furthermore, if we make the additional hypothesis that the ESDs*

$$\mu_{(\frac{1}{\sqrt{n}}M_n - zI)(\frac{1}{\sqrt{n}}M_n - zI)^*} \tag{4}$$

[3] We thank Phillip Wood for creating the figures in this paper.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

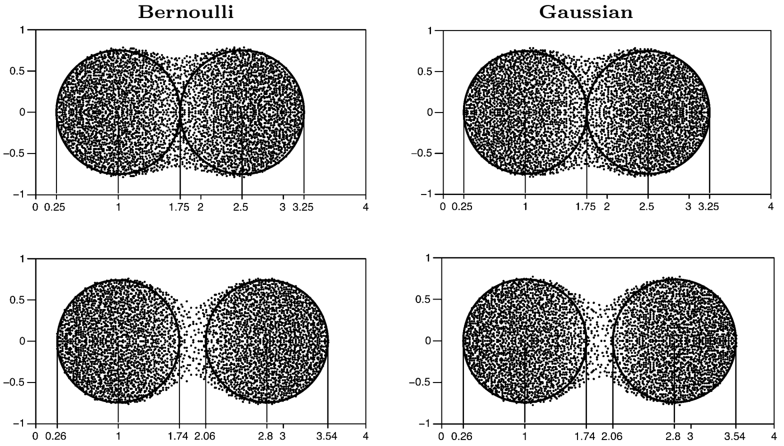

Figure 2. Eigenvalue plots of randomly generated $n \times n$ matrices of the form $D_n + M_n$, where $n = 5000$. In the left column, each entry of M_n was an iid Bernoulli random variable, taking the values $+1$ and -1 each with probability $1/2$, and in the right column, each entry was an iid Gaussian 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 $\operatorname{diag}(1, 1, \ldots, 1, 2.5, 2.5, \ldots, 2.5)$, and in the second row D_n is the deterministic matrix $\operatorname{diag}(1, 1, \ldots, 1, 2.8, 2.8, \ldots, 2.8)$ (in each case, the first $n/2$ diagonal entries are 1's, and the remaining entries are 2.5 or 2.8 as specified).

converge in the sense of probability measures to a limit for almost every z, then $\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n}$ converges strongly to zero.

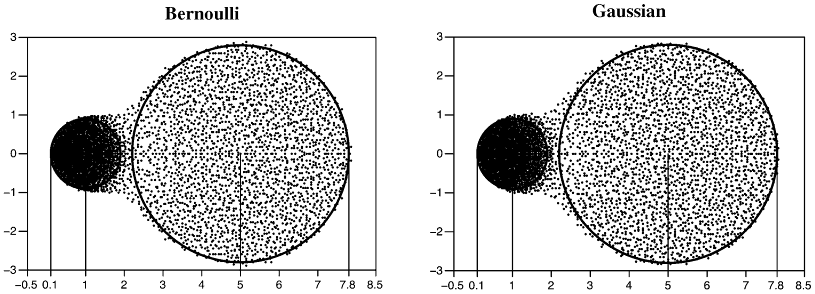

Figure 3. Eigenvalue plots of two randomly generated 5000×5000 matrices of the form $A + BM_nB$, where A and B are diagonal matrices having $n/2$ entries with the value 1 followed by $n/2$ entries with the value 5 (for D) and the value 2 (for X). On the left, each entry of M_n was an iid Bernoulli random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

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

This theorem reduces the computing of the limiting distribution to the case where one can assume[4] that the entries x have Gaussian (or any special) distribution. Combining this theorem (in the case $M_n = 0$) with Theorem 2.1, we conclude

Corollary 2.3. *The circular law* (*Conjecture* 1.3) *holds in both the weak and strong sense.*

It is useful to notice that Theorem 2.2 still holds even when the limiting distributions do not exist.

The proof of Theorem 2.2 relies on several surprising connections between seemingly remote areas of mathematics that have been discovered in the last few years. The goal of this article is to give the reader an overview of these connections and through them a sketch of the proof of Theorem 2.2. The first area we shall visit is combinatorics.

3. Combinatorics

As we shall discuss later, one of the primary difficulties in controlling the ESD of a non-Hermitian matrix $A_n = \frac{1}{\sqrt{n}} M_n$ is the presence of *pseudospectrum*—complex numbers z for which the resolvent $(A_n - zI)^{-1} = (\frac{1}{\sqrt{n}} M_n - zI)^{-1}$ exists but is extremely large. It is therefore of importance to obtain bounds on this resolvent, which leads one to understand for which vectors $v \in \mathbf{C}^n$ is $(A_n - zI)v$ likely to be small. Expanding out the vector $(A_n - zI)v$, one encounters expressions such as $\xi_1 v_1 + \cdots + \xi_n v_n$, where $v_1, \ldots, v_n \in \mathbf{C}$ are fixed and $\xi_1, \ldots, \xi_n$ are iid random variables. The problem of understanding ths distribution of such random sums is known as the *Littlewood-Offord problem*, and we now pause to discuss this problem further.

3.1. **The Littlewood-Offord problem.**

Let $\mathbf{v} = \{v_1, \ldots, v_n\}$ be a set of n integers and let $\xi_1, \ldots, \xi_n$ be iid random Bernoulli variables. Define $S := \sum_{i=1}^n \xi_i v_i$ and $p_{\mathbf{v}}(a) := \mathbf{P}(S = a)$ and $p_{\mathbf{v}} := \sup_{a \in \mathbf{Z}} p_{\mathbf{v}}(a)$.

In their study of random polynomials, Littlewood and Offord [32] raised the question of bounding $p_{\mathbf{v}}$. They showed that if the v_i are non-zero, then $p_{\mathbf{v}} = O(\frac{\log n}{\sqrt{n}})$. Very soon after, Erdős [13], using Sperner's lemma, gave a beautiful combinatorial proof for the following refinement.

Theorem 3.2. *Let $v_1, \ldots, v_n$ be non-zero numbers and ξ_i be iid Bernoulli random variables. Then*[5]

$$p_{\mathbf{v}} \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n} = O(\frac{1}{\sqrt{n}}).$$

Notice that the bound is sharp, as can be seen from the example $\mathbf{v} := \{1, \ldots, 1\}$, in which case S has a binomial distribution. Many mathematicians realized that while the classical bound in Theorem 3.2 is sharp as stated, it can be improved

[4] Some related ideas also appear in [19]. In the context of the central limit theorem, the idea of replacing arbitrary iid ensembles by Gaussian ones goes back to Lindeberg [31], and is sometimes known as the *Lindeberg invariance principle*; see [11] for further discussion, and a formulation of this principle for Hermitian random matrices.

[5] We use the usual asymptotic notation in this paper, thus $X = O(Y)$, $Y = \Omega(X)$, $X \ll Y$, or $Y \gg X$ denotes an estimate of the form $|X| \leq CY$ where C does not depend on n (but may depend on other parameters). We also let $X = o(Y)$ denote the bound $|X| \leq c(n)Y$, where $c(n) \to 0$ as $n \to \infty$.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

significantly under additional assumptions on $\mathbf{v}$. For instance, Erdős and Moser [14] showed that if the v_i are distinct, then

$$p_{\mathbf{v}} = O(n^{-3/2} \ln n).$$

They conjectured that the logarithmic term is not necessary, and this was confirmed by Sárközy and Szemerédi [42]. Again, the bound is sharp (up to a constant factor), as can be seen by taking $v_1, \dots, v_n$ to be a proper arithmetic progression such as $1, \dots, n$. Stanley [41] gave a different proof that also classified the extremal cases.

A general picture was given by Halász, who showed, among other things, that if one forbids more and more additive structure[6] in the v_i, then one gets better and better bounds on $p_{\mathbf{v}}$. One corollary of his results (see [24] or [48, Chapter 9]) is the following.

Theorem 3.3. *Consider* $\mathbf{v} = \{v_1, \dots, v_n\}$*. Let* R_k *be the number of solutions to the equation*

$$\varepsilon_1 v_{i_1} + \dots + \varepsilon_{2k} v_{i_{2k}} = 0,$$

where $\varepsilon_i \in \{-1, 1\}$ *and* $i_1, \dots, i_{2k} \in \{1, 2, \dots, n\}$*. Then*

$$p_{\mathbf{v}} = O_k(n^{-2k-1/2} R_k).$$

Remark 3.4. Several variants of Theorem 3.2 can be found in [27, 30, 16, 28] and the references therein. The connection between the Littlewood-Offord problem and random matrices was first made in [26], in connection with the question of determining how likely a random Bernoulli matrix was to be singular. In fact, the paper [26] inspired much of the work of the authors described in this survey.

3.5. **The inverse Littlewood-Offord problem.** Motivated by inverse theorems from additive combinatorics, in particular Freiman's theorem (see [15], [48, Chapter 5]) and a variant for random sums in [53, Theorem 5.2] (inspired by earlier work in [26]), the authors [49] brought a different view to the problem. Instead of trying to improve the bound further by imposing new assumptions, we aim to provide the full picture by finding the underlying reason for the probability $p_{\mathbf{v}}$ to be large (e.g., larger than n^{-A} for some fixed A).

Notice that the (multi)-set $\mathbf{v}$ has 2^n subsums, and $p_{\mathbf{v}} \geq n^{-C}$ mean that at least $2^n/n^C$ among these take the same value. This suggests that there should be very strong additive structure in the set. In order to determine this structure, one can study examples of $\mathbf{v}$ where $p_{\mathbf{v}}$ is large. For a set A, we denote by lA the set $lA := \{a_1 + \dots + a_l | a_i \in A\}$. A natural example is the following.

Example 3.6. Let $I = [-N, N]$ and $v_1, \dots, v_n$ be elements of I. Since $S \in nI$, by the pigeon hole principle, $p_{\mathbf{v}} \geq \frac{1}{nI} = \Omega(\frac{1}{N})$. In fact, a short consideration yields a better bound. Notice that with probability at least .99, we have $S \in 10\sqrt{n}I$, thus again by the pigeon hole principle, we have $p_{\mathbf{v}} = \Omega(\frac{1}{\sqrt{n}N})$. If we set $N = n^C$ for some constant C, then

$$p_{\mathbf{v}} = \Omega(\frac{1}{n^{C+1/2}}). \tag{5}$$

[6] Intuitively, this is because the less additive structure one has in the v_i, the more likely the sums S are to be distinct from each other. In the most extreme case, if the v_i are linearly independent over the rationals $\mathbf{Q}$, then the sums 2^n sums S are all distinct, and so $p_{\mathbf{v}} = 1/2^n$ in this case.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

The next, and more general, construction comes from additive combinatorics. A very important concept in this area is that of a *generalized arithmetic progression* (GAP). A set Q of numbers is a *GAP of rank* d if it can be expressed as in the form

$$Q = \{a_0 + x_1 a_1 + \cdots + x_d a_d | M_i \leq x_i \leq M_i' \text{ for all } 1 \leq i \leq d\}$$

for some $a_0, \dots, a_d, M_1, \dots, M_d, M_1', \dots, M_d'$.

It is convenient to think of Q as the image of an integer box $B := \{(x_1, \dots, x_d) \in \mathbf{Z}^d | M_i \leq x_i \leq M_i'\}$ under the linear map

$$\Phi : (x_1, \dots, x_d) \mapsto a_0 + x_1 a_1 + \cdots + x_d a_d.$$

The numbers a_i are the *generators* of P, and $\mathrm{Vol}(Q) := |B|$ is the *volume* of B. We say that Q is *proper* if this map is one to one, or equivalently if $|Q| = \mathrm{Vol}(Q)$. For non-proper GAPs, we of course have $|Q| < \mathrm{Vol}(Q)$.

Example 3.7. Let Q be a proper GAP of rank d and volume V. Let $v_1, \dots, v_n$ be (not necessarily distinct) elements of P. The random variable $S = \sum_{i=1}^n \xi_i v_i$ takes values in the GAP nP. Since $|nP| \leq \mathrm{Vol}(nB) = n^d V$, the pigeon hole principle implies that $p_{\mathbf{v}} \geq \Omega(\frac{1}{n^d V})$. In fact, using the same idea as in the previous example, one can improve the bound to $\Omega(\frac{1}{n^{d/2} V})$. If we set $N = n^C$ for some constant C, then

$$p_{\mathbf{v}} = \Omega(\frac{1}{n^{C+d/2}}). \tag{6}$$

The above examples show that if the elements of $\mathbf{v}$ belong to a proper GAP with small rank and small cardinality, then $p_{\mathbf{v}}$ is large. A few years ago, the authors [49] showed that this is essentially the only reason:

Theorem 3.8 (Weak inverse theorem [49])**.** *Let $C, \epsilon > 0$ be arbitrary constants. There are constants d and C' depending on C and ϵ such that the following holds. Assume that $\mathbf{v} = \{v_1, \dots, v_n\}$ is a multiset of integers satisfying $p_{\mathbf{v}} \geq n^{-C}$. Then there is a GAP Q of rank at most d and volume at most $n^{C'}$ which contains all but at most $n^{1-\epsilon}$ elements of $\mathbf{v}$ (counting multiplicity).*

Remark 3.9. The presence of the small set of exceptional elements is not completely avoidable. For instance, one can add $o(\log n)$ completely arbitrary elements to $\mathbf{v}$ and only decrease $p_{\mathbf{v}}$ by a factor of $n^{-o(1)}$ at worst. Nonetheless, we expect the number of such elements to be less than what is given by the results here.

The reason we call Theorem 3.8 *weak* is the fact that the dependence between the parameters is not optimal and does not yet reflect the relations in (5) and (6). Recently, we were able to modify the approach to obtain an almost optimal result.

Theorem 3.10 (Strong inverse theorem [56])**.** *Let C and $1 > \varepsilon$ be positive constants. Assume that*

$$p_{\mathbf{v}} \geq n^{-C}.$$

Then there exists a GAP Q of rank $d = O_{C,\varepsilon}(1)$ which contains all but $O_d(n^{1-\varepsilon})$ elements of $\mathbf{v}$ (counting multiplicity), where

$$|Q| = O_{C,\varepsilon}(n^{C - \frac{d}{2} + \varepsilon}).$$

The bound on $|Q|$ matches (6), up to the ε term. The proofs of Theorem 3.8 and 3.10 use harmonic analysis combined with results from the theory of random walks and several facts about GAPs. Both theorems hold in a more general setting,

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

where the elements of $\mathbf{v}$ are from a torsion-free group. The lower bound n^{-C} on $p_{\mathbf{v}}$ can also be relaxed, but the statement is more technical.

As an application of Theorem 3.10, one can deduce, in a straightforward manner, a slightly weaker version of the *forward* results mentioned above. For instance, let us show if the v_i are different, then $p_{\mathbf{v}} \le n^{-3/2+\delta}$ (for any constant $\delta > 0$). Assume otherwise and set $\varepsilon := \delta/2$. Theorem 3.10 implies that most of $\mathbf{v}$ is contained in a GAP Q of rank d and cardinality at most $O(n^{3/2-\delta-d/2+\varepsilon}) = O(n^{1-\delta/2}) = o(n)$. But since $\mathbf{v}$ has $(1-o(1))n$ elements in Q, we obtain a contradiction.

Next we consider another application of Theorem 3.10, which will be more important in later sections. This theorem enables us to execute very precise counting arguments. Assume that we would like to count the number of (multi)-sets $\mathbf{v}$ of integers with $\max |v_i| \le N$ such that $P(v) \ge p := n^{-C}$.

Fix $d \ge 1$, fix[7] a GAP Q with rank d and volume $V = n^{C-d/2}$. The dominating term will be the number of multi-subsets of size n of Q, which is

$$|Q|^n = n^{(C-d/2+\epsilon)n} \le n^{Cn} n^{-n/2+\epsilon n} = p^{-n} n^{-n(1/2-\epsilon)}. \tag{7}$$

For later purposes, we need a continuous version of this result. Let the v_i be complex numbers. Instead of $p_{\mathbf{v}}$, consider the maximum *small ball* probability

$$p_{\mathbf{v}}(\beta) = \max_{z \in \mathbf{C}} \mathbf{P}(|S - z| \le \beta).$$

Given a small $\beta > 0$ and $p = n^{-O(1)}$, the collection of $\mathbf{v}$ such that $|v| = 1$ and $p_{\mathbf{v}}(\beta) \ge p$ is infinite, but we are able to show that it can be approximated by a small set.

Theorem 3.11 (The β-net Theorem [54]). *Suppose that $p = n^{-O(1)}$. Then the set of unit vectors $\mathbf{v} = (v_1, \dots, v_n)$ such that $p_{\mathbf{v}}(\beta) \ge p$ admits a β-net (in the infinity norm[8] Ω) of size at most*

$$|\Omega| \le p^{-n} n^{-n/2+o(n)}. \tag{8}$$

Remark 3.12. A related result (with different parameters) appears in [38]; in our notation, the probability p is allowed to be much smaller, but the net is coarser (essentially, a $\beta\sqrt{n}$-net rather than a β-net). In terms of random matrices, the results in [38] are better suited to control the extreme tail of such quantities as the least singular value of $A_n - zI$, but require more boundedness conditions on the matrix A_n (and in particular, bounded operator norm) due to the coarser nature of the net.

4. Computer science

Our next stop is computer science and numerical linear algebra, and in particular the problem of dealing with *ill-conditioned* matrices, which is closely related to the issue of pseudospectrum which is of central importance in the circular law problem.

[7] A more detailed version of Theorems 3.8 and 3.10 tells us that there are not too many ways to choose the generators of Q. In particular, if $N = n^{O(1)}$, the number of ways to fix these is negligible.

[8] In other words, for any $\mathbf{v}$ with $p_{\mathbf{v}}(\beta) \ge p$, there exists $\mathbf{v}' \in \Omega$ such that all coefficients of $\mathbf{v} - \mathbf{v}'$ do not exceed β in magnitude.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

4.1. **Theory vs. practice.** Running times of algorithms are frequently estimated by worst-case analysis. But in practice, it has been observed that many algorithms, especially those involving a large matrix, perform significantly better than the worst-case scenario. The most famous example is perhaps the simplex algorithm in linear programming. Here, the basic problem (in its simplest form) is to optimize a goal function $c \cdot x$, under the constraint $Ax \leq b$, where c, b are given vectors of length n and A is an $n \times n$ matrix. In the worst-case scenario, this algorithm takes exponential time. But in practice, the algorithm runs extremally well. It is still very popular today, despite the fact that there are many other algorithms proven to have polynomial complexity.

There have been various attempts to explain this phenomenon. In this section we will discuss an influential recent explanation given by Spielman and Teng [44, 45].

4.2. **The effect of noise.** An important issue in the theory of computing is noise, as almost all computational processes are effected by it. By the word *noise*, we would like to refer to all kinds of errors occurring in a process, due to both humans and machines, including errors in measuring, errors caused by truncations, errors committed in transmitting and inputting the data, etc.

Spielman and Teng [44] pointed out that when we are interested in a solving a certain system of equations, because of noise, our computer actually ends up solving a slightly perturbed version of the system. This is the core of their so-called *smoothed analysis* that they used to explain the effectiveness of a specific algorithm (such as the simplex method). Interestingly, noise, usually a burden, plays a "positive" role here, as it smoothes the inputs randomly, and so prevents a very bad input from ever occurring.

The puzzling question here is, of course, *Why is the perturbed input typically better than the original* (*worst-case*) *input*?

In order to give a mathematical explanation, we need to introduce some notation. For an $n \times n$ matrix M, the *condition number* $\kappa(M)$ is defined as

$$\kappa(M) := \|M\| \|M^{-1}\|,$$

where $\| \|$ denotes the operator norm. (If M is not invertible, we set $\kappa(M) = \infty$.)

The condition number plays a crucial role in numerical linear algebra; in particular, the condition number $\kappa(M)$ of a matrix M serves as a simplified proxy for the accuracy and stability of most algorithms used to solve the equation $Mx = b$ (see [5, 23], for example). The exact solution $x = M^{-1}b$, in theory, can be computed quickly (by Gaussian elimination, say). However, in practice computers can only represent a finite subset of real numbers, and this leads to two difficulties: the represented numbers cannot be arbitrarily large or small, and there are gaps between two adjacent represented numbers. A quantity which is frequently used in numerical analysis is $\varepsilon_{\text{machine}}$ which is half of the distance from 1 to the nearest represented number. A fundamental result in numerical analysis [5] asserts that if one denotes by $\tilde{x}$ the result computed by computers, then the relative error $\frac{\|\tilde{x}-x\|}{\|x\|}$ satisfies

$$\frac{\|\tilde{x} - x\|}{\|x\|} = O\big(\varepsilon_{\text{machine}} \kappa(M)\big).$$

Following the literature, we call M *well conditioned* if $\kappa(M)$ is small. For quantitative purposes, we say that an $n \times n$ matrix M is *well-conditioned* if its condition

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

number is polynomially bounded in n (that is, $\kappa(M) \leq n^C$ for some constant C independent of n).

4.3. **Randomly perturbed matrices are well conditioned.**

The analysis in [44] is guided by the following fundamental intuition:[9]

Conjecture 4.4. *For every input instance, it is unlikely that a slight random perturbation of that instance has a large condition number.*

More quantitatively,

Conjecture 4.5. *Let A be an arbitrary $n \times n$ matrix, and let M_n be a random $n \times n$ matrix. Then with high probability, $A + M_n$ is well conditioned.*

Notice that here one allows A to have a large condition number.

Let us take a look at $\kappa(A+M_n) = \|A+M_n\|\|(A+M_n)^{-1}\|$. In order to have $\kappa(A+M_n) = n^{O(1)}$, we want to upper-bound both $\|A+M_n\|$ and $\|(A+M_n)^{-1}\|$. Bounding $\|A+M_n\|$ is easy, since by the triangle inequality

$$\|A+M_n\| \leq \|A\| + \|M_n\|.$$

In most models of random matrices, $\|M_n\| \leq n^{O(1)}$ with very high probability, so it suffices to assume that $\|A\| \leq n^{O(1)}$; thus, we assume that the matrix A is of polynomial size compared to the noise level. This is a fairly reasonable assumption for high-dimensional matrices for which the effect of noise is non-negligible,[10] and we are going to assume it in the rest of this section.

The remaining problem is to bound the norm of the inverse $\|(A+M_n)^{-1}\|$. An important detail here is how to choose the random matrix M_n. In their works [44, 45, 43], Spielman and Teng (and coauthors) set M_n to have iid Gaussian entries (with variance 1) and obtained the following bound, which played a critical role in their smoothed analysis [44, 45].

Theorem 4.6. *Let A be an arbitrary $n \times n$ matrix, and let M_n be a random matrix with iid Gaussian entries. Then for any $x > 0$,*

$$\mathbf{P}(\|(A+M_n)^{-1}\| \geq x) = O(\frac{\sqrt{n}}{x}).$$

While Spielman-Teng smoothed analysis does seem to have the right philosophy, the choice of M_n is a bit artificial. Of course, the analysis still passes if one replaces the Gaussian distribution by a fine enough approximation. A large fraction of problems in linear programming deal with integral matrices, so the noise is perturbation by integers. In other cases, even when the noise has continuous support, the data is strongly truncated. For example, in many engineering problems, one does not keep more than, say, three to five decimal places. Thus, in many situations, the entries of M_n end up having discrete support with relatively small size, which may not even

[9] This conjecture, of course, does not fully explain the phenomenon of smoothed analysis, since it may be that a well-conditioned matrix still causes a difficulty in one's linear algorithms for some other reason, or perhaps the original ill-conditioned matrix did not cause a difficulty in the first place; we thank Alan Edelman for pointing out this subtlety. Nevertheless, Conjecture 4.4 does provide an informal intuitive justification of smoothed analysis, and various rigorous versions of this conjecture were used in the formal arguments in [44]: see Section 1.4 of that paper for further discussion.

[10] In particular, it is naturally associated to the concept of *polynomially smoothed analysis* from [44].

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

grow with n, while the approximation mentioned above would require this support to have size exponential in n. Therefore, in order to come up with an analysis that better captures real life data, one needs to come up with a variant of Theorem 4.6 where the entries of M_n have discrete support.

This problem was suggested to the authors by Spielman a few years ago. Using the Weak Inverse Theorem, we were able to prove the following variant of Theorem 4.6 [50].

Theorem 4.7. *For any constants $a, c > 0$, there is a constant $b = b(a, c) > 0$ such that the following holds. Let A be an $n \times n$ matrix such that $\|A\| \le n^a$, and let M_n be a random matrix with iid Bernoulli entries. Then*

$$\mathbf{P}(\|(A + M_n)^{-1}\| \ge n^b) \le n^{-c}.$$

Using the stronger β-net Theorem, one can have a nearly optimal relation between the constants a, b and c [51]. These results extend, with the same proof, to a large variety of distributions. For example, one does not need to require the entries of M_n to be iid,[11] although independence is crucially exploited in the proofs. Also, one can allow many of the entries to be 0 [50].

Remark 4.8. Results of this type first appear in [37] (see also [33] for some earlier related work for the least singular value of *rectangular* matrices). In the special case where $A = 0$ and where the entries of M_n are iid and have finite fourth moment, Rudelson and Vershynin [38] (see also [39], [40]) obtained sharp bounds for $\|(A + M_n)^{-1}\|$, using a somewhat different method, which relies on an inverse theorem of a slightly different nature; see Remark 3.12.

The main idea behind the proof of Theorem 4.7, which first appears in [37], is the following. Let d_i be the distance from the ith row vector of $A + M_n$ to the subspace spanned by the rest of the rows. Elementary linear algebra (see also (10) below) then gives the bound

$$\|(A + M_n)^{-1}\| = n^{O(1)}(\min_{1 \le i \le n} d_i)^{-1}.$$

Ignoring various factors of $n^{O(1)}$, the main task is then to understand the distribution of d_i for any given i.

If $v = (v_1, \dots, v_n)$ is the normal vector of a hyperplane V, then the distance from a random vector $(a_1 + \xi_1, \dots, a_n + \xi_n)$ to the hyperplane V is given by the formula

$$|v_1(\xi_1 + a_1) + \cdots + v_n(\xi_n + a_n)| = |\sum_i a_i v_i + S|,$$

where $S := \sum_{i=1}^n v_i \xi_i$ is as in the previous section.

To estimate the chance that $|\sum_{i=1}^n a_i v_i + S| \le \beta$, the notion of the small ball probability $p_{\mathbf{v}}(\beta)$ comes naturally. Of course, this quantity depends on the normal vector $\mathbf{v}$, and so we now divide into cases depending on the nature of this vector.

If $p_{\mathbf{v}}(\beta)$ small, we can be done using a conditioning argument[12]. On the other hand, the β-net Theorem says that there are "few" $\mathbf{v}$ such that $p_{\mathbf{v}}(\beta)$ is large, and

[11] In practice, one would expect the noise at a large entry to have larger variance than one at a small entry, due to multiplicative effects.

[12] Intuitively, the idea of this conditioning argument is to first fix (or "condition") on $n - 1$ of the rows of $A + M_n$, which should then fix the normal vector $\mathbf{v}$. The remaining row is independent of the other $n - 1$ rows, and so should have a probability at most $p_{\mathbf{v}}(\beta)$ of lying within β of the span of the those rows. There are some minor technical issues in making this argument (which

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

in this case a direct counting argument finishes the job.[13] Details can be found in [50], [54], or [51].

5. Back to probability

5.1. The replacement principle. Let us now take another look at the Circular Law Conjecture. Recall that $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $A_n = \frac{1}{\sqrt{n}} M_n$, which generates a normalized counting measure μ_{A_n}. We want to show that μ_{A_n} tends (in probability) to the uniform measure μ on the unit disk.

The traditional way to attack this conjecture is via a Stieltjes transform technique,[14] following [18, 3]. Given a (complex) measure ν, define, for any z with $\mathrm{Im}\, z > 0$,

$$s_\nu(z) := \int \frac{1}{x-z} d\nu(x).$$

For the ESD μ_{A_n}, we have

$$s_{\mu_{A_n}}(z) = \frac{1}{n} \sum \frac{1}{\lambda_i - z}.$$

Thanks to standard results from probability,[15] in order to establish the Circular Law Conjecture in the strong (resp. weak) sense, it suffices to show that $s_{\mu_n}(z)$ converges almost surely (resp. in probability) to $s_\mu(z)$ for almost all z (see [55] for a precise statement).

Set $z =: s + it$ and $s_n(z) =: S + iT$. Since s_n is analytic except at the poles and vanishes at infinity, the Stieltjes transform $s_n(z)$ is determined by its real part S. Let us take a closer look at this variable:

$$\begin{aligned} S &= \frac{1}{n} \sum \frac{\mathrm{Re}\,(\lambda_i) - s}{|\lambda_i - z|^2} \\ &= -\frac{1}{2n} \sum \frac{\partial}{\partial s} \log |\lambda_i - z|^2 \\ &= -\frac{1}{2} \frac{\partial}{\partial s} \int_0^\infty \log x \; \partial \eta_n, \end{aligned}$$

where

$$\eta_n := \mu_{(\frac{1}{\sqrt{n}} M_n - zI)(\frac{1}{\sqrt{n}} M_n - zI)^*}$$

essentially dates back to [29]) rigorous, arising from the fact that the $n-1$ rows may be too degenerate to accurately control $\mathbf{v}$, but these difficulties can be dealt with, especially if one is willing to lose factors of $n^{O(1)}$ in various places.

[13] For instance, one important class of $\mathbf{v}$ for which $p_{\mathbf{v}}(\beta)$ tends to be large are the *compressible* vectors $\mathbf{v}$, in which most of the entries are close to zero. Each compressible $\mathbf{v}$ (e.g., $\mathbf{v} = (1, -1, 0, \dots, 0)$) has a moderately large probability of being close to a normal vector for $A + M_n$ (e.g., in the random Bernoulli case, $\mathbf{v} = (1, -1, 0, \dots, 0)$ has a probability about 2^{-n} of being a normal vector); but the number (or more precisely, the metric entropy) of the set of compressible vectors is small (of size $2^{o(n)}$), and so the net contribution of these vectors is then manageable. Similar arguments (relying heavily on the β-net Theorem) handle other cases when $\mathbf{v}$ is large (e.g., if most entries of $\mathbf{v}$ live near a GAP of controlled size).

[14] The more classical *moment method*, which is highly successful in the Hermitian setting (for instance in proving Theorem 1.2), is not particularly effective in the non-Hermitian setting because moments such as traceA_n^m for $m = 0, 1, 2, \dots$ do not determine the ESD μ_{A_n} (even approximately) unless one takes m to be as large as n; see [3], [4] for further discussion.

[15] One can also use the theory of logarithmic potentials for this, as is done for instance in [21], [35].

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

is the normalised counting measure of the (squares of the) *singular values* of $\frac{1}{\sqrt{n}}M_n - zI$. Notice that in the third equality, we use the fact that $\prod|\lambda_i - z| = |\det(\frac{1}{\sqrt{n}}M_n - zI)|$. This step is critical as it reduces the study of a complex measure to a real one, or in other words to study the ESD of a Hermitian matrix rather than a non-Hermitian matrix.

Putting this observation in the more general setting of Theorem 2.2, we arrived at the following useful result.

Theorem 5.2 (Replacement principle [55]). *Suppose for each n that $A_n, B_n \in M_n(\mathbf{C})$ are ensembles of random matrices. Assume that:*

(i) *The expression*

$$\frac{1}{n^2}\|A_n\|_F^2 + \frac{1}{n^2}\|B_n\|_F^2 \tag{9}$$

is weakly (resp. strongly) bounded;[16] *and*

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

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

converges weakly (resp. strongly) to zero. In particular, for each fixed z, these determinants are non-zero with probability $1 - o(1)$ for all n (resp. almost surely non-zero for all but finitely many n).

Then $\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n}$ converges weakly (resp. strongly) to zero.

At a technical level, this theorem reduces Theorem 2.2 to the comparison of $\log|\det(\frac{1}{\sqrt{n}}A_n - zI)|$ and $\log|\det(\frac{1}{\sqrt{n}}B_n - zI)|$.

Remark 5.3. Note that this expression is large and unstable when z lies in the *pseudospectra* of either $\frac{1}{\sqrt{n}}A_n$ or $\frac{1}{\sqrt{n}}B_n$, which means that the resolvent $(\frac{1}{\sqrt{n}}A_n - zI)^{-1}$ or $(\frac{1}{\sqrt{n}}B_n - zI)^{-1}$ is large. Controlling the probability of the event that z lies in the pseudospectrum is therefore an important portion of the analysis. This technical problem is not an artefact of the method, but is in fact essential to any attempt to control non-Hermitian ESDs for general random matrix models, as such ESDs are extremely sensitive to perturbations in the matrix in regions of pseudospectrum. See [3], [4] for further discussion.

5.4. **Treatment of the pole.**

Using techniques from probability, such as the moment method, one can show that the distributions of the singular values of $\frac{1}{\sqrt{n}}A_n - zI$ and $\frac{1}{\sqrt{n}}B_n - zI$ are asymptotically the same[17] [3, 54, 10, 55, 11]. This, however, is not sufficient to conclude that $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - zI)|$ and $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}B_n - zI)|$ are close. As remarked earlier, the main difficulty here is that some of the singular values can be very small and thus they can significantly influence the value of the logarithm.

[16] A sequence x_n of non-negative random variables is said to be *weakly bounded* if $\lim_{C\to\infty}\liminf_{n\to\infty}\mathbf{P}(x_n \le C) = 1$, and *strongly bounded* if $\limsup_{n\to\infty} x_n < \infty$ with probability 1.

[17] In the setting where the matrices X_n and Y_n have iid entries, one can use the results of [10] to establish this. In the non-iid case, an invariance principle from [11] gives a slightly weaker version of this equivalence; this was observed by Manjunath Krishnapur, and it appears as an appendix to [55].

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Now is where Theorem 4.7 enters the picture. This theorem tells us that (with overwhelming probability), there is no mass between 0 and (say) n^{-C}, for some sufficiently large constant C. Using this critical information, with some more work,[18] we obtain:

Theorem 5.5 ([54]). *The Circular Law Conjecture holds* (*with both strong and weak convergence*) *under the extra condition that the entries have bounded the* $(2+\eta)$*-th moment, for some constant* $\eta > 0$.

Remark 5.6. Shortly after the appearance of [54], Götze and Tikhomirov [22] gave an alternate proof of the Weak Circular Law with these hypotheses, using a variant of Theorem 4.7, which they obtained via a method from [37], [38]. This method is based on a different version of the Weak Inverse Theorem.

5.7. **Negative second moment and sharp concentration.** At the point it was written, the analysis in [54] looked close to the limit of the method. It took some time to realize where the extra moment condition came from and even more time to figure out a way to avoid that extra condition. Consider the sums

$$\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - zI)| = \frac{1}{n}\sum_{i=1}^{n}\log\sigma_i,$$

where $\sigma_1 \geq \cdots \geq \sigma_n$ are the singular values of $\frac{1}{\sqrt{n}}A_n - zI$ and

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

where $\sigma_1' \geq \cdots \geq \sigma_n'$ are the singular values of $\frac{1}{\sqrt{n}}B_n - zI$.

As already mentioned, we know that the bulk of the σ_i and σ_i' are distributed similarly. For the smallest few, we used the lower bound on σ_n as a uniform bound to show that their contribution is negligible. This turned out to be wasteful, and we needed to use the extra moment assumption to compensate the loss in this step.

In order to remove this assumption, we need to find a way to give a better bound on other singular values. An important first step is the discovery of the following simple, but useful, identity.

The Negative Second Moment Identity [55]. Let A be an $m \times n$ matrix, $m \leq n$. Then

$$\sum_{i=1}^{m} d_i^{-2} = \sum_{i=1}^{m}\sigma_i^{-2}, \tag{10}$$

where, as usual, d_i are the distances and σ_i are the singular values.

[18] In particular, the presence of certain factors of $\log n$ arising from inserting Theorem 4.7 into the normalized log-determinant $\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - zI)|$ forces one to establish a *convergence rate* for the ESD of $\frac{1}{\sqrt{n}}A_n - zI$ which is faster than logarithmic in n in a certain sense. This is what ultimately forces one to assume the bounded $(2+\eta)$-th moment hypothesis. Actually the method allows one to relax this hypothesis to that of assuming $\mathbf{E}|\mathrm{x}|^2\log^C(2+|\mathrm{x}|) < \infty$ for some absolute constant C (e.g., $C = 16$ will do).

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

One can prove this identity using undergraduate linear algebra. With this in hand, the rest of the proof falls into place.[19] Consider the singular values $\sigma_1 \geq \cdots \geq \sigma_n$ involved in our analysis, and use A as shorthand for $\frac{1}{\sqrt{n}}A_n - zI$. To bound σ_{n-k} from below, notice that by the interlacing law

$$\sigma_{n-k}(A) \geq \sigma_{m-k}(A'),$$

where $m := n-k$ and A' is an $m \times n$ truncation of A obtained by omitting the last k rows. The Negative Second Moment Identity implies

$$k\sigma_{m-k}(A')^{-2} \leq \sum_{i=1}^{m} \sigma_i(A')^{-2} = \sum_{i=1}^{m} d_i^{-2}.$$

On the other hand, the right-hand side can be bounded efficiently, thanks to the fact that all d_i are large with overwhelming probability, which, in turn, is a consequence of Talagrand's inequality [46]:

Lemma 5.8 (Distance Lemma [52, 55])**.** *With probability $1 - n^{-\omega(1)}$, the distance from a random row vector to a subspace of codimension k is at least $\frac{1}{100}\sqrt{k/n}$, as long as $k \gg \log n$.*

Thus, with overwhelming probability, $\sum_{i=1}^{m} d_i^{-2}$ is $\Omega(m/nk) = \Omega((n-k)/nk)$, which implies

$$\sigma_{n-k}(A) \geq \sigma_{m-k}(A') \gg \frac{k}{\sqrt{(n-k)n}}.$$

This lower bound is now sufficient to establish Theorem 2.2 and with it the Circular Law Conjecture in full generality.

6. Open problems

Our investigation leads to open problems in several areas:

Combinatorics. Our studies of the Littlewood-Offord problem focus on the linear form $S := \sum_{i=1}^{n} v_i xi_i$. What can one say about higher degree polynomials ?

In [6], it was shown that for a quadratic form $Q := \sum_{1 \leq i,j \leq n} c_{ij}\xi_i\xi_j$ with non-zero coefficients, $\mathbf{P}(Q = z)$ is $O(n^{-1/8})$. It is simple to improve this bound to $O(n^{-1/4})$ [7]. On the other hand, we conjecture that the truth is $O(n^{-1/2})$, which would be sharp by taking $Q = (\xi_1 + \cdots + \xi_n)^2$. Costello (personal communication) recently improved the bound to $O(n^{-3/8})$, and it looks likely that his approach will lead to the optimal bound, or something close.

The situation with higher degrees is much less clear. In [6], a bound of the form $O(n^{-c_k})$ was shown, where c_k is a positive constant depending on k, the degree of the polynomial involved. In this bound c_k decreases very fast with k.

Smoothed analysis. Spielman-Teng smoothed analysis of the simplex algorithm [44] was done with Gaussian noise. It is a very interesting problem to see if one can achieve the same conclusion with discrete noise with fixed support, such as

[19] A possible alternate approach would be to bound the intermediate singular values directly, by adapting the results from [39]. This would, however, require some additional effort; for instance, the results in [39] assume zero mean and bounded operator norm, which is not true in general when considering $\frac{1}{\sqrt{n}}A_n - zI$ for non-zero z assuming only a mean and variance condition on the entries of A_n. In any case, the analysis in [39] ultimately goes through a computation of the distances d_i, similar to the approach we present here based on the Negative Second Moment Identity.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Bernoulli. It would give an even more convincing explanation to the efficiency of the simplex method. As discussed earlier, noise that occurs in practice typically has discrete, small support. (This question was mentioned to us by several researchers, including Spielman, few years ago.)

As discussed earlier, we now have the discrete version of Theorem 4.6. While Theorem 4.6 plays a very important part in Spielman-Teng analysis [45], there are several other parts of the proof that make use of the continuity of the support in subtle ways. It is possible to modify these parts to work for fine enough discrete approximations of the continuous (noise) variables in question. However, to do so it seems one needs to make the size of the support very large (typically exponential in n, the size of the matrix).

Another exciting direction is to consider even more realistic models of noise. For instance:

- In several problems, the matrix may have many *frozen* entries, namely those which are not affected by noise. In particular, an entry which is zero (by nature of the problem) is likely to stay zero in the whole computation. It is clear that the *pattern* of the frozen entries will be of importance. For example, if the first column consists of (frozen) zero, then no matter how the noise affects the rest of the matrix, it will always be non-singular (and of course ill conditioned). We hope to classify all patterns where theorems such as Theorem 2.2 are still valid.
- In non-frozen places, the noise could have different distributions. It is natural to think that the error at a large entry should have larger variance than the one occurring at a smaller entry.

Some preliminary results in these directions are obtained in [50]. However, we are still at the very beginning of the road and much needs to be done.

Circular Law Conjecture. A natural question here is to investigate the rate of convergence. In [54], we observed that under the extra assumption that the $(2+\varepsilon)$-moment of the entries are bounded, we can have rate of convergence of order $n^{-\delta}$ for some positive constant δ depending on ε. The exact dependence between ε and δ is not clear.

Another question concerns the determinant of random matrices. It is known, and not hard to prove, that $\log|\det M_n|$ satisfies a central limit theorem when the entries of M_n are iid Gaussian; see [20, 8]. Girko [20] claimed that the same result holds for much more general models of matrices. We, however, are unable to verify his arguments. It would be nice to have an alternative proof.

Acknowledgments

Thanks to Peter Forrester, Kenny Maples, and especially Alan Edelman for corrections.

About the authors

Professor Terrence Tao is the James and Carol Collins Chair and is a professor at the Department of Mathematics at the University of California, Los Angeles.

Professor Van Vu is a professor at the Department of Mathematics at Rutgers University.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

References

[1] L. Arnold, On the asymptotic distribution of the eigenvalues of random matrices, *J. Math. Anal. Appl.*, **20** (1967), 262-268. MR0217833 (36:922)

[2] L. Arnold, On Wigner's semi-cirle law for the eigenvalues of random matrices, *Z. Wahrsch. Verw. Gebiete,* **19** (1971), 191-198. MR0348820 (50:1315)

[3] Z. D. Bai, Circular law, *Ann. Probab.* **25** (1997), 494–529. MR1428519 (98k:60040)

[4] Z. D. Bai and J. Silverstein, Spectral analysis of large dimensional random matrices, Mathematics Monograph Series **2**, Science Press, Beijing 2006.

[5] D. Bau and L. Trefethen, Numerical linear algebra. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. MR1444820 (98k:65002)

[6] K. Costello, T. Tao and V. Vu, Random symmetric matrices are almost surely non-singular, *Duke Math. J.* **135** (2006), 395–413. MR2267289 (2008f:15080)

[7] K. Costello and V. Vu, The ranks of random graphs, *to appear in Random Structures and Algorthms.*

[8] K. Costello and V. Vu, Concentration of random determinants and permanent estimators, *submitted.*

[9] P. Deift, Universality for mathematical and physical systems. *International Congress of Mathematicians* Vol. I, 125–152, Eur. Math. Soc., Zürich, 2007. MR2334189 (2008g:60024)

[10] R. Dozier and J. Silverstein, On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices, *J. Multivar. Anal.* **98** (2007), 678–694. MR2322123

[11] S. Chatterjee, A simple invariance principle. [arXiv:math/0508213]

[12] A. Edelman, Probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the Circular Law, *Journal of Multivariate Analysis* **60** (1997), 203–232. MR1437734 (98b:15025)

[13] P. Erdős, On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** (1945), 898–902. MR0014608 (7:309j)

[14] P. Erdős and L. Moser, Elementary problems and solutions: Solutions: E736. *Amer. Math. Monthly* **54** (1947), no. 4, 229–230. MR1526680

[15] G. Freiman, Foundations of a structural theory of set addition. Translated from the Russian. *Translations of Mathematical Monographs,* Vol 37. American Mathematical Society, Providence, R. I., 1973. vii+108 pp. MR0360496 (50:12944)

[16] P. Frankl and Z. Füredi, Solution of the Littlewood-Offord problem in high dimensions. *Ann. of Math.* (2) **128** (1988), no. 2, 259–270. MR960947 (89m:05002)

[17] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, *Journal of Mathematical Physics* **6** (1965), 440–449. MR0173726 (30:3936)

[18] V. L. Girko, Circular law, *Theory Probab. Appl.* (1984), 694–706.

[19] V. L. Girko, The strong circular law. Twenty years later. II. *Random Oper. Stochastic Equations* **12** (2004), no. 3, 255–312. MR2085255 (2006e:60045)

[20] V. L. Girko, A refinement of the central limit theorem for random determinants, *Teor. Veroyatnost. i Primenen.* **42** (1997), no. 1, 63–73 (Russian); translation in *Theory Probab. Appl.* **42** (1997), no. 1, 121–129 (1998).

[21] F. Götze and A.N. Tikhomirov, On the circular law, *preprint*

[22] F. Götze and A.N. Tikhomirov, The Circular Law for random matrices, *preprint*

[23] G. Golub and C. van Loan, Matrix computations, 3rd Edtion, 1996, John Hopkins Univ. Press. MR1417720 (97g:65006)

[24] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, *Period. Math. Hungar.* **8** (1977), no. 3-4, 197–211. MR0494478 (58:13338)

[25] C. R. Hwang, A brief survey on the spectral radius and the spectral distribution of large random matrices with i.i.d. entries, *Contemp. Math.* vol. 50, Amer. Math. Soc., Providence, RI, 1986, 145–152. MR841088 (87m:60080)

[26] J. Kahn, J. Komlós and E. Szemerédi, On the probability that a random ±1-matrix is singular. *J. Amer. Math. Soc.* **8** (1995), no. 1, 223–240 MR1260107 (95c:15047)

[27] G. Katona, On a conjecture of Erdős and a stronger form of Sperner's theorem. *Studia Sci. Math. Hungar.* **1** (1966), 59–63. MR0205864 (34:5690)

[28] D. Kleitman, On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors, *Advances in Math.* **5** (1970), 155–157. MR0265923 (42:832)

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

[29] J. Komlós, On the determinant of (0, 1) matrices. *Studia Sci. Math. Hungar.* **2** (1967), 7–21. MR0221962 (36:5014)

[30] J. Griggs, J. Lagarias, A. Odlyzko, and J. Shearer, On the tightest packing of sums of vectors, *European J. Combin.* **4** (1983), no. 3, 231–236. MR725071 (84m:52021)

[31] J. W. Lindeberg, Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung, *Math. Z.* **15** (1922), 211–225. MR1544569

[32] J. E. Littlewood and A. C. Offord, On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.* **12** (1943), 277–286. MR0009656 (5:179h)

[33] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, *Adv. Math.* **195** (2005), no. 2, 491–523. MR2146352 (2006g:52009)

[34] M.L. Mehta, Random Matrices and the Statistical Theory of Energy Levels, Academic Press, New York, NY, 1967. MR0220494 (36:3554)

[35] G. Pan and W. Zhou, Circular Law, extreme singular values and potential theory, *preprint*.

[36] L. A Pastur, On the spectrum of random matrices, *Teoret. Mat. Fiz.* **10** (1973), 102-112. MR0475502 (57:15106)

[37] M. Rudelson, Invertibility of random matrices: Norm of the inverse. *Annals of Mathematics, to appear*.

[38] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and the condition number of random matrices, *Advances in Mathematics* **218** (2008), 600-633. MR2407948

[39] M. Rudelson and R. Vershynin, The smallest singular value of a rectangular random matrix, preprint.

[40] M. Rudelson and R. Vershynin, The least singular value of a random square matrix is $O(n^{-1/2})$, preprint.

[41] R. Stanley, Weyl groups, the hard Lefschetz theorem, and the Sperner property, *SIAM J. Algebraic Discrete Methods* **1** (1980), no. 2, 168–184. MR578321 (82j:20083)

[42] A. Sárközy and E. Szemerédi, Über ein Problem von Erdős und Moser, *Acta Arithmetica* **11** (1965), 205-208. MR0182619 (32:102)

[43] A. Sankar, S. H. Teng, and D. A. Spielman, Smoothed Analysis of the Condition Numbers and Growth Factors of Matrices, *preprint*.

[44] D. A. Spielman and S. H. Teng, Smoothed analysis of algorithms, *Proceedings of the International Congress of Mathematicians*, Vol. I (Beijing, 2002), 597–606, Higher Ed. Press, Beijing, 2002. MR1989210 (2004d:90138)

[45] D. A. Spielman and S. H. Teng, Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time, *J. ACM* **51** (2004), no. 3, 385–463. MR2145860 (2006f:90029)

[46] M. Talagrand, A new look at independence, *Ann. Probab.* **24** (1996), no. 1, 1–34. MR1387624 (97d:60028)

[47] T. Tao and V. Vu, On random ±1 matrices: Singularity and determinant, *Random Structures Algorithms* **28** (2006), no. 1, 1–23. MR2187480 (2006g:15048)

[48] T. Tao and V. Vu, Additive combinatorics, Cambridge University Press, 2006. MR2289012 (2008a:11002)

[49] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices, *Annals of Mathematics, to appear*.

[50] T. Tao and V. Vu, The condition number of a randomly perturbed matrix, *STOC 2007*. MR2402448

[51] T. Tao and V. Vu, Random matrices: A general approach for the least singular value problem, preprint.

[52] T. Tao and V. Vu, On random $(-1, 1)$ matrices: Singularity and determinant, *Random Structures and Algorithms* **28** (2006), no 1, 1-23. MR2187480 (2006g:15048)

[53] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices, *J. Amer. Math. Soc.* **20** (2007), 603-673. MR2291914 (2008h:60027)

[54] T. Tao and V. Vu, Random matrices: The Circular Law, *Communication in Contemporary Mathematics* **10** (2008), 261-307. MR2409368

[55] T. Tao and V. Vu, Random matrices: Universality of the ESD and the Circular Law (with an appendix by M. Krishnapur), *submitted*.

[56] T. Tao and V. Vu, *paper in preparation*.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

[57] P. Wigner, On the distribution of the roots of certain symmetric matrices, *The Annals of Mathematics* **67** (1958), 325-327. MR0095527 (20:2029)

[58] J. Wishart, The generalized product moment distribution in samples from a normal multivariate population, *Biometrika* **20** (1928), 32–52.

Department of Mathematics, UCLA, Los Angeles, California 90095-1555
E-mail address: tao@math.ucla.edu

Department of Mathematics, Rutgers University, Piscataway, New Jersey 08854
E-mail address: vanvu@math.rutgers.edu

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use