

Random matrices: The Universality phenomenon for Wigner ensembles

Terence Tao and Van Vu

ABSTRACT. In this paper, we survey some recent progress on rigorously establishing the universality of various spectral statistics of Wigner Hermitian random matrix ensembles, focusing on the Four Moment Theorem and its refinements and applications, including the universality of the sine kernel and the Central limit theorem of several spectral parameters.

We also take the opportunity here to issue some errata for some of our previous papers in this area.

1. Introduction

Random matrix theory is a central topic in probability and mathematical physics, with many connections to various areas such as statistics, number theory, combinatorics, numerical analysis and theoretical computer science.

One of the primary goal of random matrix theory is to derive limiting laws for the eigenvalues and eigenvectors of ensembles of large ($n \times n$) Hermitian random matrices¹, in the asymptotic limit $n \rightarrow \infty$. There are many random matrix ensembles of interest, but to focus the discussion and to simplify the exposition we shall restrict attention to an important model class of ensembles, the *Wigner matrix ensembles*.

DEFINITION 1.1 (Wigner matrices). Let $n \geq 1$ be an integer (which we view as a parameter going off to infinity). An $n \times n$ *Wigner Hermitian matrix* M_n is defined to be a random Hermitian $n \times n$ matrix $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$, in which the ξ_{ij} for $1 \leq i \leq j \leq n$ are jointly independent with $\xi_{ji} = \overline{\xi_{ij}}$ (in particular, the ξ_{ii} are real-valued). For $1 \leq i < j \leq n$, we require that the ξ_{ij} have mean zero and variance one, while for $1 \leq i = j \leq n$ we require that the ξ_{ij} (which are necessarily real) have mean zero and variance σ^2 for some $\sigma^2 > 0$ independent of i, j, n . To simplify some of the statements of the results here, we will also assume that the $\xi_{ij} \equiv \xi$ are identically distributed for $i < j$, and the $\xi_{ii} \equiv \xi'$ are also identically distributed for

2010 *Mathematics Subject Classification*. Primary 15A52.

T. Tao is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman award.

V. Vu is supported by research grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

¹One can of course also study non-Hermitian or non-square random matrices, though in the latter case the concept of an eigenvalue needs to be replaced with that of a singular value. In this survey we will focus almost exclusively on the square Hermitian case.

$i = j$, and furthermore that the real and imaginary parts of ξ are independent. We refer to the distributions $\operatorname{Re}\xi$, $\operatorname{Im}\xi$, and ξ' as the *atom distributions* of M_n .

We say that the Wigner matrix ensemble *obeys Condition C0* if we have the exponential decay condition

$$\mathbf{P}(|\xi_{ij}| \geq t^C) \leq e^{-t}$$

for all $1 \leq i, j \leq n$ and $t \geq C'$, and some constants C, C' (independent of i, j, n). We say that the Wigner matrix ensemble *obeys condition C1 with constant C_0* if one has

$$\mathbf{E}|\xi_{ij}|^{C_0} \leq C$$

for some constant C (independent of n).

Of course, Condition **C0** implies Condition **C1** for any C_0 , but not conversely.

We refer to the matrix $W_n := \frac{1}{\sqrt{n}}M_n$ as the *coarse-scale normalised Wigner Hermitian matrix*, and $A_n := \sqrt{n}M_n$ as the *fine-scale normalised Wigner Hermitian matrix*.

EXAMPLE 1.2 (Invariant ensembles). An important special case of a Wigner Hermitian matrix M_n is the *gaussian unitary ensemble* (GUE), in which $\xi_{ij} \equiv N(0, 1)_{\mathbb{C}}$ are complex gaussians with mean zero and variance one for $i \neq j$, and $\xi_{ii} \equiv N(0, 1)_{\mathbb{R}}$ are real gaussians with mean zero and variance one for $1 \leq i \leq n$ (thus $\sigma^2 = 1$ in this case). Another important special case is the *gaussian orthogonal ensemble* (GOE), in which $\xi_{ij} \equiv N(0, 1)_{\mathbb{R}}$ are real gaussians with mean zero and variance one for $i \neq j$, and $\xi_{ii} \equiv N(0, 1/2)_{\mathbb{R}}$ are real gaussians with mean zero and variance $1/2$ for $1 \leq i \leq n$ (thus $\sigma^2 = 1/2$ in this case). These ensembles obey Condition **C0**, and hence Condition **C1** for any C_0 . For these two ensembles, the probability distribution of M_n can be expressed invariantly in either case as

$$(1.1) \quad \frac{1}{Z_n} e^{-c \operatorname{tr} M_n M_n^*} dM_n$$

for some quantity $Z_n > 0$ depending only on n , where dM_n is Haar measure on the vector space of $n \times n$ Hermitian matrices (in the case of GUE) or real symmetric matrices (in the case of GOE), and c is equal to $1/2$ (for GUE) or $1/4$ (for GOE). From (1.1) we easily conclude that the probability distribution of GUE is invariant with respect to conjugations by unitary matrices, and similarly the probability distribution of GOE is invariant with respect to conjugations by orthogonal matrices. However, a general Wigner matrix ensemble will not enjoy invariances with respect to such large classes of matrices. (For instance, the Bernoulli ensembles described below are only invariant with respect to conjugation by a discrete group of matrices, which include permutation matrices and reflections around the coordinate axes.)

EXAMPLE 1.3 (Bernoulli ensembles). At the opposite extreme from the invariant ensembles are the *Bernoulli ensembles*, which are discrete instead of continuous. In the *real Bernoulli ensemble* (also known as *symmetric random sign matrices*), each of the ξ_{ij} are equal to $+1$ with probability $1/2$ and -1 with probability $1/2$. In the *complex Bernoulli ensemble*, the diagonal entries ξ_{ii} still have this distribution, but the off-diagonal entries now take values² $\pm \frac{1}{\sqrt{2}} \pm \frac{\sqrt{-1}}{\sqrt{2}}$, with each of these four complex numbers occurring with probability $1/4$.

²We use $\sqrt{-1}$ to denote the imaginary unit, in order to free up the symbol i as an index variable.

REMARK 1.4. Many of the results given here have been extended to somewhat broader classes of matrices. For instance, one can consider generalised Wigner ensembles in which entries are not identically distributed; for instance, one can allow the variances σ_{ij}^2 of each entry ξ_{ij} to vary in i, j , and even vanish for some i, j ; the latter situation occurs for instance in band-limited random matrices; one can also consider allowing the mean μ_{ij} of the entries ξ_{ij} to be non-zero (this is for instance the situation with the adjacency matrices of Erdős-Renyi random graphs). One can also consider sparse Wigner random matrices, in which only a small (randomly selected) number of entries are non-zero. For simplicity, though, we shall mostly restrict attention in this survey to ordinary Wigner ensembles. We do remark, however, that in all of these generalisations, it remains crucial that the entries ξ_{ij} for $1 \leq i \leq j \leq n$ are jointly independent, as many of the techniques currently available to control the fine-scale spectral structure of Wigner matrices rely heavily on joint independence.

Given an $n \times n$ Hermitian matrix A , we denote its n eigenvalues in increasing order³ as

$$\lambda_1(A) \leq \dots \leq \lambda_n(A),$$

and write $\lambda(A) := (\lambda_1(A), \dots, \lambda_n(A))$. We also let $u_1(A), \dots, u_n(A) \in \mathbb{C}^n$ be an orthonormal basis of eigenvectors of A with $Au_i(A) = \lambda_i(A)u_i(A)$; these eigenvectors $u_i(A)$ are only determined up to a complex phase even when the eigenvalues are simple (or up to a sign in the real symmetric case), but this ambiguity will not cause much of a difficulty in our results as we will usually only be interested in the *magnitude* $|u_i(A)^*X|$ of various inner products $u_i(A)^*X$ of $u_i(A)$ with other vectors X .

We also introduce the *eigenvalue counting function*

$$(1.2) \quad N_I(A) := |\{1 \leq i \leq n : \lambda_i(A) \in I\}|$$

for any interval $I \subset \mathbb{R}$. We will be interested in both the *coarse-scale* eigenvalue counting function $N_I(W_n)$ and the *fine-scale* eigenvalue counting function $N_I(A_n)$, which are of course transformable to each other by the identity $N_I(W_n) = N_{nI}(A_n)$.

2. Global and local semi-circular laws

We first discuss the coarse-scale spectral structure of Wigner matrices, that is to say the structure of the eigenvalues of W_n at unit scales (or equivalently, the eigenvalues of M_n at scale \sqrt{n} , or A_n at scale n). The fundamental result in this topic is the (*global*) *Wigner semi-circular law*. Denote by ρ_{sc} the semi-circle density function with support on $[-2, 2]$,

$$(2.1) \quad \rho_{sc}(x) := \begin{cases} \frac{1}{2\pi} \sqrt{4 - x^2}, & |x| \leq 2 \\ 0, & |x| > 2. \end{cases}$$

THEOREM 2.1 (Global semi-circular law). *Let M_n be a Wigner Hermitian matrix. Then for any interval I (independent of n), one has*

$$\lim_{n \rightarrow \infty} \frac{1}{n} N_I[W_n] = \int_I \rho_{sc}(y) dy$$

³It is also common in the literature to arrange eigenvalues in decreasing order instead of increasing. Of course, the results remain the same under this convention except for minor notational changes.

in the sense of probability (and also in the almost sure sense, if the M_n are all minors of the same infinite Wigner Hermitian matrix).

REMARK 2.2. Wigner [126] proved this theorem for special ensembles. The general version above is due to Pastur [85] (see [6, 1] for a detailed discussion). The semi-circular law in fact holds under substantially more general hypotheses than those given in Definition 1.1, but we will not discuss this matter further here. One consequence of Theorem 2.1 is that we expect most of the eigenvalues of W_n to lie in the interval $(-2 + \varepsilon, 2 + \varepsilon)$ for $\varepsilon > 0$ small; we shall thus informally refer to this region as the *bulk* of the spectrum.

An essentially equivalent⁴ formulation of the semi-circular law is as follows: if $1 \leq i \leq n$, then one has⁵

$$(2.2) \quad \lambda_i(W_n) = \lambda_i^{\text{cl}}(W_n) + o(1)$$

with probability $1 - o(1)$ (and also almost surely, if the M_n are minors of an infinite matrix), where the *classical location* $\lambda_i^{\text{cl}}(W_n)$ of the i^{th} eigenvalue is the element of $[-2, 2]$ defined by the formula

$$\int_{-2}^{\lambda_i^{\text{cl}}(W_n)} \rho_{\text{sc}}(y) dy = \frac{i}{n}.$$

A remarkable feature of the semi-circular law is its *universality*: the precise distribution of the atom variables ξ_{ij} are irrelevant for the conclusion of the law, so long as they are normalised to have mean zero and variance one (or variance σ^2 , on the diagonal), and are jointly independent on the upper-triangular portion of the matrix. In particular, continuous matrix ensembles such as GUE or GUE, and discrete matrix ensembles such as the Bernoulli ensembles, have the same asymptotic spectral distribution when viewed at the coarse scale (i.e. by considering eigenvalue ranges of size ~ 1 for W_n , or equivalently of size $\sim \sqrt{n}$ for M_n or $\sim n$ for A_n).

However, as stated, the semi-circular law does not give good control on the *fine-scale* behaviour of the eigenvalues, for instance in controlling $N_I(W_n)$ when I is a very short interval (of length closer to $1/n$ than to 1). The fine-scale theory for Wigner matrices is much more recent than the coarse-scale theory given by the semi-circular law, and is the main focus of this survey.

There are several ways to establish the semi-circular law. For invariant ensembles such as GUE or GOE, one can use explicit formulae for the probability distribution of N_I coming from the theory of determinantal processes, giving precise estimates all the way down to infinitesimally small scales; see e.g. [1] or Section 3 below. However, these techniques rely heavily on the invariance of the ensemble, and do not directly extend to more general Wigner ensembles. Another popular technique is the *moment method*, based on the basic moment identities

$$(2.3) \quad \sum_{i=1}^n \lambda_i(W_n)^k = \text{trace } W_n^k = \frac{1}{n^{k/2}} \text{trace } M_n^k$$

⁴This formulation is slightly stronger because it also incorporates the upper bound $\|W_n\|_{\text{op}} \leq 2 + o(1)$ on the operator norm of W_n , which is consistent with, but not implied by, the semi-circular law, and follows from the work of Bai and Yin [7].

⁵We use the asymptotic notation $o(1)$ to denote any quantity that goes to zero as $n \rightarrow \infty$, and $O(X)$ to denote any quantity bounded in magnitude by CX , where C is a constant independent of n .

for all $k \geq 0$. The moment method already is instructive for revealing at least one explanation for the universality phenomenon. If one takes expectations in the above formula, one obtains

$$\mathbf{E} \sum_{i=1}^n \lambda_i(W_n)^k = \frac{1}{n^{k/2}} \sum_{1 \leq i_1, \dots, i_k \leq n} \mathbf{E} \xi_{i_1 i_2} \dots \xi_{i_k i_1}.$$

For those terms for which each edge $\{i_j, i_{j+1}\}$ appears at most twice, the summand can be explicitly computed purely in terms of the mean and variances of the ξ_{ij} , and are thus universal. Terms for which an edge appears three or more times are sensitive to higher moments of the atom distribution, but can be computed to give a contribution of $o(1)$ (at least assuming a decay condition such as Condition **C0**). This already explains universality for quantities such as $\mathbf{E}N_I(W_n)$ for intervals I of fixed size (independent of n), at least if one assumes a suitable decay condition on the entries (and one can use standard truncation arguments to relax such hypotheses substantially).

At the edges ± 2 of the spectrum, the moment method can be pushed further, to give quite precise control on the most extreme eigenvalues of W_n (which dominate the sum in (2.3)) if k is sufficiently large; see [50, 97, 98, 99, 87, 124] and the references therein. However, the moment method is quite poor at controlling the spectrum in the bulk. To improve the understanding of the bulk spectrum of Wigner matrices, Bai [2, 3] (see also [6, Chapter 8]) used the *Stieltjes transform method* (building upon the earlier work of Pastur [85]) to show that the speed of convergence to the semi-circle was $O(n^{-1/2})$. Instead of working with moments, one instead studied the Stieltjes transform

$$s_n(z) = \frac{1}{n} \text{trace}(W_n - z)^{-1} = \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i(W_n) - z}$$

of W_n , which is well-defined for z outside of the spectrum of W_n (and in particular for z in the upper half-plane $\{z \in \mathbb{C} : \text{Im}(z) > 0\}$). To establish the semi-circular law, it suffices to show that $s_n(z)$ converges (in probability or almost surely) to $s_{\text{sc}}(z)$ for each z , where

$$s_{\text{sc}}(z) := \int_{\mathbb{R}} \frac{1}{x - z} \rho_{\text{sc}}(x) dx$$

is the Stieltjes transform of the semi-circular distribution ρ_{sc} . The key to the argument is to establish a *self-consistent equation* for s_n , which roughly speaking takes the form

$$(2.4) \quad s_n(z) \approx \frac{-1}{s_n(z) + z}.$$

One can explicitly compute by contour integration that

$$s_{\text{sc}}(z) = \frac{1}{2}(-z + \sqrt{z^2 - 4})$$

for $z \notin [-2, 2]$, where $\sqrt{z^2 - 4}$ is the branch of the square root that is asymptotic to z at infinity, and in particular that

$$(2.5) \quad s_{\text{sc}}(z) = \frac{-1}{s_{\text{sc}}(z) + z}.$$

From a stability analysis of the elementary equation (2.5) and a continuity argument, one can then use (2.4) to show that

$$(2.6) \quad s_n(z) \approx s_{\text{sc}}(z),$$

which then implies convergence to the semi-circular law. If one can obtain quantitative control on the approximation in (2.4), one can then deduce quantitative versions of the semi-circular law that are valid for certain short intervals.

We briefly sketch why one expects the self-consistent equation to hold. One can expand

$$(2.7) \quad s_n(z) = \frac{1}{n} \sum_{i=1}^n ((W_n - zI)^{-1})_{ii}$$

where $((W_n - zI)^{-1})_{ii}$ denotes the ii^{th} entry of the matrix $(W_n - zI)^{-1}$. Let us consider the $i = n$ term for sake of concreteness. If one expands W_n as a block matrix

$$(2.8) \quad W_n - zI := \begin{pmatrix} \tilde{W}_{n-1} - zI & \frac{1}{\sqrt{n}} X_n \\ \frac{1}{\sqrt{n}} X_n^* & \frac{1}{\sqrt{n}} \xi_{nn} - z \end{pmatrix},$$

where $\tilde{W}_{n-1} = \frac{1}{\sqrt{n}} M_{n-1}$ is the top left $n-1 \times n-1$ minor of M_n , and X_n is the $n-1 \times 1$ column vector with entries $\xi_{n1}, \dots, \xi_{n(n-1)}$, then an application of Schur's complement yields the identity

$$(2.9) \quad ((W_n - zI)^{-1})_{nn} = \frac{-1}{z + \frac{1}{n} X_n^* (\tilde{W}_{n-1} - zI)^{-1} X_n - \frac{1}{\sqrt{n}} \xi_{nn}}.$$

The term $\frac{1}{\sqrt{n}} \xi_{nn}$ is usually negligible and will be ignored for this heuristic discussion. Let us temporarily freeze (or condition on) the entries of the random matrix M_{n-1} , and hence \tilde{W}_{n-1} . Due to the joint independence of the entries of M_n , the entries of X_n remain jointly independent even after this conditioning. As these entries also have mean zero and variance one, we easily compute that

$$\mathbf{E} \frac{1}{n} X_n^* (\tilde{W}_{n-1} - zI)^{-1} X_n = \frac{1}{n} \text{trace}(\tilde{W}_{n-1} - zI)^{-1}.$$

Using the Cauchy interlacing theorem

$$(2.10) \quad \lambda_{i-1}(M_{n-1}) \leq \lambda_i(M_n) \leq \lambda_i(M_{n-1})$$

property between the eigenvalues of M_{n-1} and the eigenvalues of M_n (which easily follows from the Courant-Fisher minimax formula

$$\lambda_i(M_n) = \inf_{V \subset \mathbb{C}^n; \dim(V)=i} \sup_{u \in V; \|u\|=1} u^* M_n u$$

for the eigenvalues), one easily obtains an approximation of the form

$$\frac{1}{n} \text{trace}(\tilde{W}_{n-1} - zI)^{-1} \approx s_n(z).$$

Assuming that the expression $\frac{1}{n} X_n^* (\tilde{W}_{n-1} - zI)^{-1} X_n$ concentrates around its mean (which can be justified under various hypotheses on M_n and z using a variety of concentration-of-measure tools, such as Talagrand's concentration inequality, see e.g. [76]), one thus has

$$(2.11) \quad \frac{1}{n} X_n^* (\tilde{W}_{n-1} - zI)^{-1} X_n \approx s_n(z)$$

and thus

$$((W_n - zI)^{-1})_{nn} \approx \frac{-1}{z + s_n(z)}.$$

Similarly for the other diagonal entries $((W_n - zI)^{-1})_{ii}$. Inserting this approximation back into (2.7) gives the desired approximation (2.4), heuristically at least.

The above argument was optimized⁶ in a sequence of papers [37, 38, 39] by Erdős, Schlein, and Yau (see also [108, Section 5.2] for a slightly simplified proof). As a consequence, one was able to obtain good estimates of the form (2.6) even when z was quite close to the spectrum $[-2, 2]$ (e.g. at distance $O(n^{-1+\varepsilon})$ for some small $\varepsilon > 0$, which in turn leads (by standard arguments) to good control on the eigenvalue counting function $N_I(W_n)$ for intervals I of length as short as $n^{-1+\varepsilon}$. (Note that as there are only n eigenvalues in all, such intervals are expected to only have about $O(n^\varepsilon)$ eigenvalues in them.) Such results are known as *local semi-circular laws*. A typical such law (though not the strongest such law known) is as follows:

THEOREM 2.3 (Local semi-circle law). *Let M_n be a Wigner matrix obeying Condition **C0**, let $\varepsilon > 0$, and let $I \subset \mathbb{R}$ be an interval of length $|I| \geq n^{-1+\varepsilon}$. Then with overwhelming probability⁷, one has*

$$(2.12) \quad N_I(W_n) = n \int_I \rho_{\text{sc}}(x) dx + o(n|I|).$$

PROOF. See e.g. [103, Theorem 1.10]. For the most precise estimates currently known of this type (and with the weakest decay hypotheses on the entries), see [32]. \square

A variant of Theorem 2.3, which was established⁸ in the subsequent paper [44], is the extremely useful *eigenvalue rigidity property*

$$(2.13) \quad \lambda_i(W_n) = \lambda_i^{\text{cl}}(W_n) + O_\varepsilon(n^{-1+\varepsilon}),$$

valid with overwhelming probability in the bulk range $\delta n \leq i \leq (1 - \delta)n$ for any fixed $\delta > 0$ (and assuming Condition **C0**), and which significantly improves upon (2.2). This result is key in some of the strongest applications of the theory. See Section 7.7 for the precise form of this result and recent developments.

Roughly speaking, results such as Theorem 2.3 and (2.13) control the spectrum of W_n at scales $n^{-1+\varepsilon}$ and above. However, they break down at the fine scale n^{-1} ; indeed, for intervals I of length $|I| = O(1/n)$, one has $n \int_I \rho_{\text{sc}}(x) dx = O(1)$, while $N_I(W_n)$ is clearly a natural number, so that one can no longer expect an asymptotic of the form (2.12). Nevertheless, local semicircle laws are an essential

⁶There are further refinements to this method in Erdős, Yau, and Yin [43], [44] and Erdős-Knowles-Yau-Yin [32], [33] that took advantage of some additional cancellation between the error terms in (2.9) (generalised to indices $i = 1, \dots, n$) that could be obtained (via the moment method in [43], [44], and via decoupling arguments in [32], [33]), to improve the error estimates further. These refinements are not needed for the application to Wigner matrices assuming a strong decay condition such as Condition **C0**, but is useful in generalised Wigner matrix models, such as sparse Wigner matrices in which most of the entries are zero, or in models where one only has a weak amount of decay (e.g. Condition **C1** with $C_0 = 4 + \varepsilon$). These refinements also lead to the very useful *eigenvalue rigidity bound* (2.13).

⁷By this, we mean that the event occurs with probability $1 - O_A(n^{-A})$ for each $A > 0$.

⁸The result in [44] actually proves a more precise result that also gives sharp results in the edge of the spectrum, though due to the sparser nature of the $\lambda_i^{\text{cl}}(W_n)$ in that case, the error term $O_\varepsilon(n^{-1+\varepsilon})$ must be enlarged.

part of the fine-scale theory. One particularly useful consequence of these laws is that of *eigenvector delocalisation*, first established in [39]:

COROLLARY 2.4 (Eigenvalue delocalisation). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $\varepsilon > 0$. Then with overwhelming probability, one has $u_i(W_n)^* e_j = O(n^{-1/2+\varepsilon})$ for all $1 \leq i, j \leq n$.*

Note from Pythagoras' theorem that $\sum_{j=1}^n |u_i(M_n)^* e_j|^2 = \|u_i(M_n)\|^2 = 1$; thus Corollary 2.4 asserts, roughly speaking, that the coefficients of each eigenvector are as spread out (or *delocalised*) as possible.

PROOF. (Sketch) By symmetry we may take $e_j = n$. Fix i , and set $\lambda := \lambda_i(W_n)$; then the eigenvector equation $(W_n - \lambda)u_i(W_n) = 0$ can be expressed using (2.8) as

$$\begin{pmatrix} \tilde{W}_{n-1} - \lambda I & \frac{1}{\sqrt{n}} X_n \\ \frac{1}{\sqrt{n}} X_n^* & \frac{1}{\sqrt{n}} \xi_{nn} - \lambda \end{pmatrix} \begin{pmatrix} \tilde{u}_i \\ u_i(W_n)^* e_n \end{pmatrix} = 0$$

where \tilde{u}_i are the first $n-1$ coefficients of $u_i(W_n)$. After some elementary algebraic manipulation (using the normalisation $\|u_i(W_n)\| = 1$), this leads to the identity

$$|u_i(W_n)^* e_n|^2 = \frac{1}{1 + \|(\tilde{W}_{n-1} - \lambda)^{-1} X_n\|^2/n}$$

and hence by eigenvalue decomposition

$$|u_i(W_n)^* e_n|^2 = \frac{1}{1 + \sum_{j=1}^{n-1} (\lambda_j(\tilde{W}_{n-1}) - \lambda)^{-2} |u_j(\tilde{W}_{n-1})^* X_n|^2/n}.$$

Suppose first that we are in the bulk case when $\delta n \leq i \leq (1-\delta)n$ for some fixed $\delta > 0$. From the local semicircle law, we then see with overwhelming probability that there are $\gg n^{\varepsilon/2}$ eigenvalues $\lambda_j(\tilde{W}_{n-1})$ that lie within $n^{-1-\varepsilon/2}$ of λ . Letting V be the span of the corresponding eigenvectors, we conclude that

$$|u_i(W_n)^* e_n|^2 \ll n^{-1+\varepsilon} / \|\pi_V(X_n)\|^2$$

where π_V is the orthogonal projection to V . If we freeze (i.e. condition) on \tilde{W}_{n-1} and hence on V , then the coefficients of X_n remain jointly independent with mean zero and variance one. A short computation then shows that

$$\mathbf{E} \|\pi_V(X_n)\|^2 = \dim(V) \gg n^{\varepsilon/2}$$

and by using concentration of measure tools such as Talagrand's concentration inequality (see e.g. [76]), one can then conclude that $\|\pi_V(X_n)\| \gg 1$ with overwhelming probability. This concludes the claim of the Corollary in the bulk case.

The edge case is more delicate, due to the sparser spectrum near λ . Here, one takes advantage of the identity

$$\lambda + \frac{1}{n} X_n^* (\tilde{W}_{n-1} - \lambda I)^{-1} X_n - \frac{1}{\sqrt{n}} \xi_{nn} = 0,$$

(cf. (2.9)), which we can rearrange as

$$(2.14) \quad \frac{1}{n} \sum_{j=1}^{n-1} (\lambda_j(\tilde{W}_{n-1}) - \lambda)^{-1} |u_j(\tilde{W}_{n-1})^* X_n|^2 = \frac{1}{\sqrt{n}} \xi_{nn} - \lambda.$$

In the edge case, the right-hand side is close to ± 2 , and this can be used (together with the local semicircle law) to obtain the lower bound

$$\sum_{j=1}^{n-1} (\lambda_j(\tilde{W}_{n-1}) - \lambda)^{-2} |u_j(\tilde{W}_{n-1})^* X_n|^2 \gg n^{-1-\varepsilon}$$

with overwhelming probability, which gives the claim. See [103] for details. \square

REMARK 2.5. A slicker approach to eigenvalue delocalisation proceeds via control of the resolvent (or Green's function) $(W_n - zI)^{-1}$, taking advantage of the identity

$$\operatorname{Im}((W_n - zI)^{-1})_{jj} = \sum_{i=1}^n \frac{\eta}{(\lambda_i(W_n) - E)^2 + \eta^2} |u_i(W_n)^* e_j|^2$$

for $z = E + i\eta$; see for instance [31] for details of this approach. Note from (2.9) that the Stieltjes transform arguments used to establish the local semicircle law already yield control on quantities such as $((W_n - zI)^{-1})_{jj}$ as a byproduct.

3. Fine-scale spectral statistics: the case of GUE

We now turn to the question of the fine-scale behavior of eigenvalues of Wigner matrices, starting with the model case of GUE. Here, it is convenient to work with the fine-scale normalisation $A_n := \sqrt{n}M_n$. For simplicity we will restrict attention to the bulk region of the spectrum, which in the fine-scale normalisation corresponds to eigenvalues $\lambda_i(A_n)$ of A_n that are near nu for some fixed $-2 < u < 2$ independent of n .

There are several quantities at the fine scale that are of interest to study. For instance, one can directly study the distribution of individual (fine-scale normalised) eigenvalues $\lambda_i(A_n)$ for a single index $1 \leq i \leq n$, or more generally study the joint distribution of a k -tuple $\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n)$ of such eigenvalues for some $1 \leq i_1 < \dots < i_k \leq n$. Equivalently, one can obtain estimates for expressions of the form

$$(3.1) \quad \mathbf{E}F(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))$$

for various test functions $F : \mathbb{R}^k \rightarrow \mathbb{R}$. By specializing to the case $k = 2$ and to translation-invariant functions $F(x, y) := f(x - y)$, one obtains distributional information on individual eigenvalue gaps $\lambda_{i+1}(A_n) - \lambda_i(A_n)$.

A closely related set of objects to the joint distribution of individual eigenvalues are the k -point correlation functions $R_n^{(k)} = R_n^{(k)}(A_n) : \mathbb{R}^k \rightarrow \mathbb{R}^+$, defined via duality to be the unique symmetric function (or measure) for which one has

$$(3.2) \quad \int_{\mathbb{R}^k} F(x_1, \dots, x_k) R_n^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k = k! \sum_{1 \leq i_1 < \dots < i_k} \mathbf{E}F(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))$$

for all symmetric continuous compactly supported functions $F : \mathbb{R}^k \rightarrow \mathbb{R}$. For discrete ensembles, $R_n^{(k)}$ is only defined as a measure on \mathbb{R}^k (which, with our conventions, has total mass $\frac{n!}{(n-k)!}$); but for continuous ensembles, $R_n^{(k)}$ is a continuous function, and for $x_1 < \dots < x_k$, one can equivalently define $R_n^{(k)}(x_1, \dots, x_k)$ in this

case by the formula

$$R_n^{(k)}(x_1, \dots, x_k) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^k} \mathbf{P}(E_{\varepsilon, x_1, \dots, x_k})$$

where $E_{\varepsilon, x_1, \dots, x_k}$ is the event that there is an eigenvalue of A_n in each of the intervals $[x_i, x_i + \varepsilon]$ for $i = 1, \dots, k$. Alternatively, one can write

$$R_n^{(k)}(x_1, \dots, x_k) = \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} \rho_n(x_1, \dots, x_n) dx_{k+1} \dots dx_n$$

where $\rho_n := \frac{1}{n!} R_n^{(n)}$ is the symmetrized joint probability distribution of all n eigenvalues of A_n .

Note from (3.2) that control on the expressions (3.1) implies (in principle, at least) control on the k -point correlation function by summing over the relevant indices i_1, \dots, i_k ; and from eigenvalue rigidity estimates such as (3.1) we see that (for fixed k , at least) there are only $n^{o(1)}$ choices for the k -tuple (i_1, \dots, i_k) that contribute to this sum.

From the semi-circular law, we expect that at the energy level nu for some $-2 < u < 2$, the eigenvalues of A_n will be spaced with average spacing $1/\rho_{\text{sc}}(u)$. It is thus natural to consider the *normalised k -point correlation function* $\rho_{n,u}^{(k)} = \rho_{n,u}^{(k)}(A_n) : \mathbb{R}^k \rightarrow \mathbb{R}^+$, defined by the formula

$$(3.3) \quad \rho_{n,u}^{(k)}(x_1, \dots, x_k) := \frac{1}{\rho_{\text{sc}}(u)^k} R_n^{(k)} \left(nu + \frac{x_1}{\rho_{\text{sc}}(u)}, \dots, nu + \frac{x_k}{\rho_{\text{sc}}(u)} \right).$$

Informally, for infinitesimal $\varepsilon > 0$, $\varepsilon^k \rho_{n,u}^{(k)}(x_1, \dots, x_k)$ is approximately equal to the probability that there is an eigenvalue in each of the intervals $[nu + \frac{x_i}{\rho_{\text{sc}}(u)}, nu + \frac{x_i + \varepsilon}{\rho_{\text{sc}}(u)}]$ for

The Stieltjes transform $s_n(z) = \text{trace}((W_n - z)^{-1})$ introduced previously is related to the fine-scale normalised eigenvalues by the formula

$$s_n(z) = \sum_{i=1}^n \frac{1}{\lambda_i(A_n) - nz}.$$

More generally, one can consider the random variables

$$(3.4) \quad \text{trace}((W_n - z_1)^{-1} \dots (W_n - z_k)^{-1}) = \sum_{1 \leq i \leq n} \frac{1}{\lambda_i(A_n) - nz_1} \dots \frac{1}{\lambda_i(A_n) - nz_k}.$$

The joint distribution of such random variables can be expressed in terms of the k -point correlations; for instance, one has

$$\mathbf{E} \text{trace}((W_n - z_1)^{-1} \dots (W_n - z_k)^{-1}) = \int_{\mathbb{R}} \frac{R_n^{(1)}(x)}{(x - nz_1) \dots (x - nz_k)} dx.$$

Conversely, it is possible (with some combinatorial effort) to control the k -point correlation function in terms of the joint distribution of such random variables; see [43, §8].

The distribution of the eigenvalue counting functions $N_I(A_n) = N_{I/n}(W_n)$ can be expressed in terms of the distribution of the individual eigenvalues or from the correlation function. For instance, one has (for continuous ensembles, at least) the formula

$$\mathbf{E} \binom{N_I(A_n)}{k} = \frac{1}{k!} \int_{I^k} R_n^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k$$

for any $k \geq 0$ (with the convention that $\binom{n}{k} = 0$ whenever $n < k$).

Finally, to close the circle of relationships, by coupling the eigenvalue counting function $N_I(W_n)$ for intervals such as $I = [-2, x]$ with previously mentioned quantities such as the Stieltjes transforms, as well as additional *level repulsion* estimates that prevent two consecutive eigenvalues from getting too close to each other too often, one can recover control of individual eigenvalues; see [71].

It has been generally believed (and in many cases explicitly conjectured; see e.g. [80, page 9]) that the asymptotic statistics for the quantities mentioned above are *universal*, in the sense that the limiting laws do not depend on the distribution of the atom variables (assuming of course that they have been normalised as stated in Definition 1.1). This phenomenon was motivated by examples of similarly universal laws in physics, such as the laws of thermodynamics or of critical percolation; see e.g. [80, 20, 22] for further discussion.

It is clear that if one is able to prove the universality of a limiting law, then it suffices to compute this law for one specific model in order to describe the asymptotic behaviour for all other models. A natural choice for the specific model is GUE, as for this model, many limiting laws can be computed directly thanks to the availability of an explicit formula for the joint distribution of the eigenvalues, as well as the useful identities of determinantal processes. For instance, one has the formula of Dyson [26]

$$(3.5) \quad \rho_n(x_1, \dots, x_n) = \frac{1}{(2\pi n)^{n/2}} e^{-|x|^2/2n} \prod_{1 \leq i < j \leq n} |x_i - x_j|^2,$$

for the joint eigenvalue distribution, as can be verified from (1.1) and a change of variables. From (3.5) and a downwards induction on k one can then obtain the *Gaudin-Mehta formula*

$$(3.6) \quad R_n^{(k)}(x_1, \dots, x_k) = \det(K_n(x_i, x_j))_{1 \leq i, j \leq k}$$

for all $0 \leq k \leq n$, where K_n is the kernel

$$K_n(x, y) = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} P_k\left(\frac{x}{\sqrt{n}}\right) e^{-x^2/2n} P_k\left(\frac{y}{\sqrt{n}}\right) e^{-y^2/2n}$$

and P_0, P_1, \dots are the Hermite polynomials (thus each P_n is a degree n polynomial, with the P_n being orthonormal with respect to the measure $e^{-x^2/2} dx$). This formula, combined with the classical Plancherel-Rotach asymptotics for Hermite polynomials, gives the limiting law

$$(3.7) \quad \lim_{n \rightarrow \infty} \rho_{n,u}^{(k)}(x_1, \dots, x_k) = \rho_{\text{Sine}}^{(k)}(x_1, \dots, x_k)$$

locally uniformly in x_1, \dots, x_k where

$$\rho_{\text{Sine}}^{(k)}(x_1, \dots, x_k) := \det(K_{\text{Sine}}(x_i, x_j))_{1 \leq i, j \leq k}$$

and K_{Sine} is the *Dyson sine kernel*

$$K_{\text{Sine}}(x, y) := \frac{\sin(\pi(x-y))}{\pi(x-y)}$$

(with the usual convention that $\frac{\sin x}{x}$ equals 1 at the origin); see [51, 80]. Standard determinantal process identities then give an asymptotic

$$(3.8) \quad N_{[nu+a/\rho_{\text{sc}}(u), nu+b/\rho_{\text{sc}}(u)]}(A_n) \rightarrow \sum_{j=1}^{\infty} \xi_j$$

for any fixed real numbers $a < b$ and $-2 < u < 2$, where the $\xi_j \in \{0, 1\}$ are independent Bernoulli variables with $\mathbf{E}(\xi_j) = \lambda_j$, the λ_j are the eigenvalues of the integral operator $T_{[a,b]} : L^2([a, b]) \rightarrow L^2([a, b])$ defined by

$$T_{[a,b]}f(x) := \int_{[a,b]} K_{\text{Sine}}(x, y)f(y) dy,$$

and the convergence is in the sense of probability distributions; see e.g. [60]. Thus, for instance, the probability that the interval $[nu+a/\rho_{\text{sc}}(u), nu+b/\rho_{\text{sc}}(u)]$ is devoid of eigenvalues converges as $n \rightarrow \infty$ to the Fredholm determinant⁹

$$\det(1 - T_{[a,b]}) := \prod_{j=1}^{\infty} (1 - \lambda_j).$$

Using this formula one can obtain a limiting law for an (averaged) eigenvalue spacing. More precisely, given an intermediate scale parameter t_n such that $t_n, n/t_n \rightarrow \infty$ as $n \rightarrow \infty$, define the quantity

$$S_n(s, u, t_n) := \frac{1}{t_n} |\{1 \leq i \leq n : |\lambda_i(A_n) - nu| \leq t_n/\rho_{\text{sc}}(u); \lambda_{i+1}(A_n) - \lambda_i(A_n) \leq s/\rho_{\text{sc}}(u)\}|.$$

Then one can establish for fixed $-2 < u < 2$ that

$$(3.9) \quad \mathbf{E}S_n(s, u, t_n) \rightarrow \int_0^s \rho(\sigma) d\sigma$$

where ρ is the *Gaudin distribution*

$$\rho(s) := \frac{d^2}{ds^2} \det(1 - T_{[0,s]});$$

see [19]. Informally, ρ is the asymptotic distribution of the normalised gap $\rho_{\text{sc}}(u)(\lambda_{i+1}(A_n) - \lambda_i(A_n))$ for typical $\lambda_i(A_n)$ near nu .

A variant of the computations that lead to (3.8) (and more precisely, a general central limit theorem for determinantal processes due to Costin-Leibowitz [14] and Soshnikov [100]) can give a limiting law for $N_I(A_n)$ in the case of the macroscopic intervals $I = [nu, +\infty)$. More precisely, one has the central limit theorem

$$\frac{N_{[nu, +\infty)}(A_n) - n \int_u^{\infty} \rho_{\text{sc}}(y) dy}{\sqrt{\frac{1}{2\pi^2} \log n}} \rightarrow N(0, 1)_{\mathbb{R}}$$

in the sense of probability distributions, for any $-2 < u < 2$; see [57]. By using the counting functions $N_{[nu, +\infty)}$ to solve for the location of individual eigenvalues $\lambda_i(A_n)$, one can then conclude the central limit theorem

$$(3.10) \quad \frac{\lambda_i(A_n) - \lambda_i^{\text{cl}}(A_n)}{\sqrt{\log n / 2\pi^2 / \rho_{\text{sc}}(u)}} \rightarrow N(0, 1)_{\mathbb{R}}$$

whenever $\lambda_i^{\text{cl}}(A_n) := n\lambda_i^{\text{cl}}(W_n)$ is equal to $n(u + o(1))$ for some fixed $-2 < u < 2$; see [57]. Informally, this asserts (in the GUE case, at least) that each eigenvalue

⁹See [63] for a more explicit description of this determinant in terms of a solution to an ODE.

$\lambda_i(A_n)$ typically deviates by $O(\sqrt{\log n}/\rho_{sc}(u))$ around its classical location; this result should be compared with (2.13), which has a slightly worse bound on the deviation (of shape $O_\varepsilon(n^\varepsilon)$ instead of $O(\sqrt{\log n})$) but which holds with overwhelming probability (and for general Wigner ensembles).

Much of the the above analysis extends to many other classes of invariant ensembles (such as GOE¹⁰, for which the joint eigenvalue distribution has a form similar to (3.5) (namely, an exponential factor and a power of a Vandermonde determinant). However, the Hermite polynomials are usually replaced by some other family of orthogonal polynomials, and one needs additional tools (such as the theory of Riemann-Hilbert problems) to obtain enough asymptotic control on those polynomials to recover the other results of the type given here. See [20] for further discussion. However, we will not pursue this important aspect of the universality phenomenon for random matrices here, as our focus is instead on the Wigner matrix models.

4. Extending beyond the GUE case I. Heat flow methods

The arguments used to establish the results in the previous section relied heavily on the special structure of the GUE ensemble, and in particular the fact that the joint probability distribution had a determinantal structure (cf. (3.6)). To go significantly beyond the GUE case, there are two families of techniques used. One family are the *heat flow* methods, based on applying an Ornstein-Uhlenbeck process to a Wigner ensemble M_n^0 to obtain a *gauss divisible* ensemble M_n^t that is closer to the GUE, and showing that the latter obeys approximately the same statistics as the GUE. The other family of methods are the *swapping* methods, in which one replaces the entries of one Wigner ensemble M_n with another ensemble M'_n which are close in some suitable sense (e.g. in the sense of matching moments), and shows that the statistics for both ensembles are close to each other. The two methods are complementary, and many of the strongest results known about universality for Wigner matrices use a combination of both methods.

Our focus shall largely be on the swapping methods (and in particular on the *four moment theorem*), but in this section we will briefly survey the heat flow techniques. (For a more detailed survey of these methods, see [31] or [96].)

Let M_n^0 be a Wigner matrix. One can then define the *matrix Ornstein-Uhlenbeck process* M_n^t for times $t \in [0, +\infty)$ by the stochastic differential equation¹¹

$$dM_n^t = d\beta_t - \frac{1}{2}M_n^t dt$$

with initial data $M_n^t|_{t=0} = M_n^0$, where β_t is a Hermitian matrix process whose entries are standard real Brownian motions on the diagonal and standard complex Brownian motions off the diagonal, with the β_t being independent of M_n^0 , and with the upper-triangular entries of β_t being jointly independent. A standard stochastic calculus computation shows that M_n^t is distributed according to the law

$$M_n^t \equiv e^{-t/2}M_n^0 + (1 - e^{-t})^{1/2}G_n,$$

¹⁰There are some further direct relationships between the GOE and GUE eigenvalue distributions that can be used to deduce control of the former from that of the latter; see [47].

¹¹One can omit the normalising term $-\frac{1}{2}M_n^t dt$ in this process to obtain a Brownian process rather than an Ornstein-Uhlenbeck process, but we retain the normalising term in order to keep the variance of each (off-diagonal) entry of the matrix M_n^t fixed.

where G_n is a GUE matrix independent of M_n^0 . In particular, the random matrix M_n^t is distributed as M_n^0 for $t = 0$ and then continuously deforms towards the GUE distribution as $t \rightarrow +\infty$. We say that a Wigner matrix is *gauss divisible* (also known as a *Johansson matrix*) with parameter t if it has the distribution of $e^{-t/2}M_n^0 + (1 - e^{-t})^{1/2}G_n$ for some Wigner matrix M_n^0 ; thus M_n^t is gauss divisible with parameter t . Note that not every Wigner matrix is gauss divisible; among other things, gauss divisible matrices necessarily have a continuous (and in fact smooth) distribution rather than a discrete one. The larger one makes t , the more restrictive the requirement of gauss divisibility becomes, until in the asymptotic limit $t = +\infty$ the only gauss divisible ensemble remaining is GUE.

The dynamics of the (fine-scale normalised) eigenvalues $\lambda_i(A_n^t) = \sqrt{n}\lambda_i(M_n^t)$ were famously established by Dyson [27] to be governed by the (normalised) *Dyson Brownian motion* equations

$$(4.1) \quad d\lambda_i(A_n^t) = \sqrt{n}dB_i + n \sum_{1 \leq j \leq n: j \neq i} \frac{dt}{\lambda_i(A_n^t) - \lambda_j(A_n^t)} - \frac{1}{2}\lambda_i(A_n^t) dt,$$

where B_1, \dots, B_n are independent standard real Brownian motions.

One can phrase the Dyson Brownian motion in a dual form in terms of the joint eigenvalue distribution function $\rho_n^t : \mathbb{R}^n \rightarrow \mathbb{R}^+$ at time t . Namely, ρ_n^t obeys the *Dyson Fokker-Planck equation*

$$(4.2) \quad \frac{\partial}{\partial t}\rho_n^t = D\rho_n^t$$

where D is the differential operator

$$D\rho := \frac{n}{2} \sum_{i=1}^n \partial_j^2 \rho - n \sum_{1 \leq i, j \leq n: i \neq j} \partial_j \left(\frac{\rho}{x_i - x_j} \right) + \frac{1}{2} \sum_{j=1}^n \partial_j(x_j \rho)$$

and we write ∂_j as shorthand for the partial derivative $\frac{\partial}{\partial x_j}$. Observe that the Ginibre distribution ρ_n^∞ defined by (3.5) is annihilated by D and is thus an equilibrium state of the Dyson Fokker-Planck equation; this is of course consistent with the convergence of the distribution of M_n^t to the distribution of GUE.

The Dyson Fokker-Planck equation (4.2) can in fact be solved explicitly by observing the identity

$$D(\Delta_n u) = \Delta_n \left(Lu + \frac{n(n-1)}{4} u \right)$$

where $\Delta_n(x) := \prod_{1 \leq i < j \leq n} (x_i - x_j)$ is the Vandermonde determinant, and L is the Ornstein-Uhlenbeck operator

$$Lu := \frac{n}{2} \sum_{i=1}^n \partial_j^2 u + \frac{1}{2} \sum_{i=1}^n \partial_j(x_j u).$$

This allows one to reduce the Dyson Fokker-Planck equation by a change of variables to the Ornstein-Uhlenbeck Fokker-Planck equation $\frac{\partial}{\partial t} u = Lu$, which has an explicit fundamental solution. Using this¹² one can obtain an explicit formula for ρ_n^t in terms of ρ_n^0 , and with a bit more effort one can also obtain a (slightly messy) determinantal formula for the associated correlation functions $(R_n^{(k)})^t$; see [13],

¹²The derivation of these formulae in [64] is somewhat different, proceeding via the Harish-Chandra/Itzykson-Zuber formula [59]; however, as noted in that paper, one can also use Dyson Brownian motion to derive the formula, which is the approach taken here.

[64]. By exploiting these explicit formulae, Johansson [64] was able¹³ to extend the asymptotic (3.7) for the k -point correlation function from GUE to the more general class of gauss divisible matrices with fixed parameter $t > 0$ (independent of n).

It is of interest to extend this analysis to as small a value of t as possible, since if one could set $t = 0$ then one would obtain universality for all Wigner ensembles. By optimising Johansson's method (and taking advantage of the local semi-circle law), Erdős, Peche, Ramirez, Schlein, and Yau [35] were able to extend the universality of (3.7) (interpreted in a suitably weak convergence topology, such as vague convergence) to gauss divisible ensembles for t as small as $n^{-1+\varepsilon}$ for any fixed $\varepsilon > 0$.

An alternate approach to these results was developed by Erdős, Ramirez, Schlein, Yau, and Yin [34], [40], [41]. The details are too technical to be given here, but the main idea is to use standard tools such as log-Sobolev inequalities to control the rate of convergence of the Dyson Fokker-Planck equation to the equilibrium measure ρ_n^∞ , starting from the initial data ρ_n^0 ; informally, if one has a good convergence to this equilibrium measure by time t , then one can obtain universality results for gauss divisible ensembles with this parameter t . A simple model to gain heuristic intuition on the time needed to converge to equilibrium is given by the one-dimensional Ornstein-Uhlenbeck process

$$(4.3) \quad dx = \sigma d\beta_t - \theta(x - \mu) dt$$

for some parameters $\sigma, \theta > 0$, $\mu \in \mathbb{R}$ and some standard Brownian motion β_t . Standard computations (or dimensional analysis) suggest that this process should converge to the equilibrium measure (in this case, a normal distribution $N(\mu, \sigma^2/2\theta)$) in time¹⁴ $O(1/\theta)$, in the sense that the probability distribution of x should differ from the equilibrium distribution by an amount that decays exponentially in t/θ .

As was already observed implicitly by Dyson, the difficulty with the Dyson Fokker-Planck equation (4.2) (or equivalently, the Dyson Brownian motion (4.1)) is that different components of the evolution converge to equilibrium at different rates. Consider for instance the trace variable $T := \lambda_1(A_n^t) + \dots + \lambda_n(A_n^t)$. Summing up (4.1) we see that this variable evolves by the Ornstein-Uhlenbeck process

$$dT = nd\beta_t - \frac{1}{2}T dt$$

for some standard Brownian motion β_t , and so we expect convergence to equilibrium for this variable in time $O(1)$. At the other extreme, consider an eigenvalue gap $s_i := \lambda_{i+1}(A_n^t) - \lambda_i(A_n^t)$ somewhere in the bulk of the spectrum. Subtracting two consecutive cases of (4.1), we see that

$$(4.4) \quad ds_i = 2\sqrt{nd}\beta_{t,i} - \theta_i s_i dt + \frac{2n}{s_i} dt$$

¹³Some additional technical hypotheses were assumed in [64], namely that the diagonal variance σ^2 was equal to 1, that the real and imaginary parts of each entry of M'_n were independent, and that Condition C1 held for some $C_0 > 6$.

¹⁴If the initial position $x(0)$ is significantly farther away from the mean μ than the standard deviation $\sqrt{\sigma^2/2\theta}$, say $|x(0) - \mu| \sim K\sqrt{\sigma^2/2\theta}$, then one acquires an additional factor of $\log K$ in the convergence to equilibrium, because it takes time about $\log K/\theta$ for the drift term $-\theta(x - \mu) dt$ in (4.3) to move x back to within $O(1)$ standard deviations of μ . These sorts of logarithmic factors will be of only secondary importance in this analysis, ultimately being absorbed in various $O(n^\varepsilon)$ error factors.

where

$$\theta_i := n \sum_{1 \leq j \leq n: j \neq i, i+1} \frac{1}{(\lambda_{i+1}(A_n^t) - \lambda_j(A_n^t))(\lambda_i(A_n^t) - \lambda_j(A_n^t))} + \frac{1}{2}.$$

Using the heuristic $\lambda_j(A_n) \approx \lambda_j^{\text{cl}}(A_n)$, we expect θ_i to be of size comparable to n and s_i to be of size comparable to 1; comparing (4.4) with (4.3) we thus expect s_i to converge to equilibrium in time $O(1/n)$.

One can use a standard log-Sobolev argument of Bakry and Emery [5] (exploiting the fact that the equilibrium measure ρ_n^∞ is the negative exponential of a strictly convex function H) to show (roughly speaking) that the Dyson Brownian motion converges to global equilibrium in time $O(1)$; see e.g. [31]. Thus the trace variable T is among the slowest of the components of the motion to converge to equilibrium. However, for the purposes of controlling local statistics such as the normalised k -point correlation function

$$\rho_{n,u}^{(k)}(x_1, \dots, x_k) := R_n^{(k)}\left(nu + \frac{x_1}{\rho_{\text{sc}}(u)}, \dots, nu + \frac{x_k}{\rho_{\text{sc}}(u)}\right),$$

and particularly the *averaged* normalised k -point correlation function

$$\frac{1}{2b} \int_{u_0-b}^{u_0+b} R_n^{(k)}\left(nu + \frac{x_1}{\rho_{\text{sc}}(u_0)}, \dots, nu + \frac{x_k}{\rho_{\text{sc}}(u_0)}\right) du,$$

these “slow” variables turn out to essentially be irrelevant, and it is the “fast” variables such as s_i which largely control these expressions. As such, one expects these particular quantities to converge to their equilibrium limit at a much faster rate. By replacing the global equilibrium measure ρ_n^∞ with a localized variant which has better convexity properties in the slow variables, it was shown in [40], [41] by a suitable modification of the Bakry-Emery argument that one in fact had convergence to equilibrium for such expressions in time $O(n^{-1+\varepsilon})$ for any fixed ε ; a weak version¹⁵ of the rigidity of eigenvalues statement (2.13) is needed in order to show that the error incurred by replacing the actual equilibrium measure with a localized variant is acceptable. Among other things, this argument reproves a weaker version of the result in [35] mentioned earlier, in which one obtained universality for the asymptotic (3.7) after an additional averaging in the energy parameter u . However, the method was simpler and more flexible than that in [35], as it did not rely on explicit identities, and has since been extended to many other types of ensembles, including the real symmetric analogue of gauss divisible ensembles in which the role of GUE is replaced instead by GOE. Again, we refer the reader to [31] for more details.

5. Extending beyond the GUE case II. Swapping and the Four Moment Theorem

The heat flow methods discussed in the previous section enlarge the class of Wigner matrices to which GUE-type statistics are known to hold, but do not cover all such matrices, and in particular leave out discrete ensembles such as the Bernoulli ensembles, which are not gauss divisible for any $t > 0$. To complement these methods, we have a family of *swapping* methods to extend spectral asymptotics from one ensemble to another, based on individual replacement of

¹⁵Roughly speaking, the rigidity result that is needed is that one has $\lambda_i(W_n) = \lambda_i^{\text{cl}}(W_n) + O(n^{1/2-c})$ in an ℓ^2 -averaged sense for some absolute constant $c > 0$. See [31] for details.

each coefficient of a Wigner matrix, as opposed to deforming all of the coefficients simultaneously via heat flow.

The simplest (but rather crude) example of a swapping method is based on the *total variation distance* $d(X, Y)$ between two random variables X, Y taking values in the same range R , defined by the formula

$$d(X, Y) := \sup_{E \subset R} |\mathbf{P}(X \in E) - \mathbf{P}(Y \in E)|,$$

where the supremum is over all measurable subsets E of R . Clearly one has

$$|\mathbf{E}F(X) - \mathbf{E}F(Y)| \leq \|F\|_{L^\infty} d(X, Y)$$

for any measurable function $F : R \rightarrow \mathbb{R}$. As such, if $d(M_n, M'_n)$ is small, one can approximate various spectral statistics of M_n by those of M'_n , or vice versa. For instance, one has

$$|\mathbf{E}F(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n)) - F(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n))| \leq \|F\|_{L^\infty} d(M_n, M'_n),$$

and thus from (3.2) we have the somewhat crude bound

$$\begin{aligned} & \left| \int_{\mathbb{R}^k} F(x_1, \dots, x_k) R_n^{(k)}(A_n)(x_1, \dots, x_k) - F(x_1, \dots, x_k) R_n^{(k)}(A'_n)(x_1, \dots, x_k) dx_1 \dots dx_k \right| \\ & \leq n^k \|F\|_{L^\infty} d(M_n, M'_n). \end{aligned}$$

and hence by (3.3)

$$\begin{aligned} & \left| \int_{\mathbb{R}^k} F(x_1, \dots, x_k) \rho_{n,u}^{(k)}(A_n)(x_1, \dots, x_k) - F(x_1, \dots, x_k) \rho_{n,u}^{(k)}(A'_n)(x_1, \dots, x_k) dx_1 \dots dx_k \right| \\ & \leq (\rho_{\text{sc}}(u)n)^k \|F\|_{L^\infty} d(M_n, M'_n) \end{aligned}$$

for any test function F . On the other hand, by swapping the entries of M_n with M'_n one at a time, we see that

$$d(M_n, M'_n) \leq \sum_{i=1}^n \sum_{j=1}^n d(\xi_{ij}, \xi'_{ij}).$$

We thus see that if $d(\xi_{ij}, \xi'_{ij}) \leq n^{-C}$ for a sufficiently large constant C (depending on k), then the k -point correlation functions of M_n and M'_n are asymptotically equivalent. This argument was quite crude, costing many more powers of n than is strictly necessary, and by arguing more carefully one can reduce this power; see [35]. However, it does not seem possible to eliminate the factors of n entirely from this type of argument.

By combining this sort of total variation-based swapping argument with the heat flow universality results for time $t = n^{-1+\varepsilon}$, the asymptotic (3.7) was demonstrated to hold in [35] for Wigner matrices with sufficiently smooth distribution; in particular, the $k = 2$ case of (3.7) was established if the distribution function of the atom variables were C^6 (i.e. six times continuously differentiable) and obeyed a number of technical decay and positivity conditions that we will not detail here. The basic idea is to approximate the distribution ρ of an atom variable ξ_{ij} in total variation distance (or equivalently, in L^1 norm) by the distribution $e^{tL}\tilde{\rho}$ of a gauss-divisible atom variable ξ'_{ij} with an accuracy that is better than n^{-C} for a suitable C , where L is the generator of the Ornstein-Uhlenbeck process and $t = n^{-1+\varepsilon}$; this can be accomplished by setting $\tilde{\rho}$ to essentially be a partial Taylor expansion of the (formal) backwards Ornstein-Uhlenbeck evolution $e^{-tL}\rho$ of ρ to some bounded

order, with the smoothness of ρ needed to ensure that this partial Taylor expansion $\tilde{\rho}$ remains well-defined as a probability distribution, and that $e^{tL}\tilde{\rho}$ approximates ρ sufficiently well. See [35] for more details of this method (referred to as the *method of reverse heat flow* in that paper).

Another fundamental example of a swapping method is the *Lindeberg exchange strategy*¹⁶, introduced in Lindeberg's classic proof [77] of the central limit theorem, and first applied to Wigner ensembles in [17]. We quickly sketch that proof here. Suppose that X_1, \dots, X_n are iid real random variables with mean zero and variance one, and let Y_1, \dots, Y_n be another set of iid real random variables and mean zero and variance one (which we may assume to be independent of X_1, \dots, X_n). We would like to show that the averages $\frac{X_1 + \dots + X_n}{\sqrt{n}}$ and $\frac{Y_1 + \dots + Y_n}{\sqrt{n}}$ have asymptotically the same distribution, thus

$$\mathbf{E}F\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) = \mathbf{E}F\left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}}\right) + o(1)$$

for any smooth, compactly supported function F . The idea is to swap the entries X_1, \dots, X_n with Y_1, \dots, Y_n one at a time and obtain an error of $o(1/n)$ on each such swap. For sake of illustration we shall just establish this for the first swap:

$$(5.1) \quad \mathbf{E}F\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) = \mathbf{E}F\left(\frac{X_1 + \dots + X_{n-1} + Y_n}{\sqrt{n}}\right) + o(1/n).$$

We write $\frac{X_1 + \dots + X_n}{\sqrt{n}} = S + n^{-1/2}X_n$, where $S := \frac{X_1 + \dots + X_{n-1}}{\sqrt{n}}$. From Taylor expansion we see (for fixed smooth, compactly supported F) that

$$F\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) = F(S) + n^{-1/2}X_n F'(S) + \frac{1}{2}n^{-1}X_n^2 F''(S) + O(n^{-3/2}|X_n|^3).$$

We then make the crucial observation that S and X_n are independent. On taking expectations (and assuming that X_n has a bounded third moment) we conclude that

$$\mathbf{E}F\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) = \mathbf{E}F(S) + n^{-1/2}(\mathbf{E}X_n)\mathbf{E}F'(S) + \frac{1}{2}n^{-1}(\mathbf{E}X_n^2)\mathbf{E}F''(S) + O(n^{-3/2}).$$

Similarly one has

$$\mathbf{E}F\left(\frac{X_1 + \dots + X_{n-1} + Y_n}{\sqrt{n}}\right) = \mathbf{E}F(S) + n^{-1/2}(\mathbf{E}Y_n)\mathbf{E}F'(S) + \frac{1}{2}n^{-1}(\mathbf{E}Y_n^2)\mathbf{E}F''(S) + O(n^{-3/2}).$$

But by hypothesis, X_n and Y_n have *matching moments to second order*, in the sense that $\mathbf{E}X_n^i = \mathbf{E}Y_n^i$ for $i = 0, 1, 2$. Thus, on subtracting, we obtain (5.1) (with about a factor of $n^{-1/2}$ to spare; cf. the Berry-Esséen theorem [10], [45]).

Note how the argument relied on the matching moments of the two atom variables X_i, Y_i ; if one had more matching moments, one could continue the Taylor expansion and obtain further improvements to the error term in (5.1), with an additional gain of $n^{-1/2}$ for each further matching moment.

We can apply the same strategy to control expressions such as $\mathbf{E}F(M_n) - F(M'_n)$, where M_n, M'_n are two (independent) Wigner matrices. If one can obtain bounds such as

$$\mathbf{E}F(M_n) - \mathbf{E}F(\tilde{M}_n) = o(1/n)$$

¹⁶We would like to thank S. Chatterjee and M. Krisnapur for introducing this method to us

when \tilde{M}_n is formed from M_n by replacing¹⁷ one of the diagonal entries ξ_{ii} of M_n by the corresponding entry ξ'_{ii} of M'_n , and bounds such as

$$\mathbf{E}F(M_n) - \mathbf{E}F(\tilde{M}_n) = o(1/n^2)$$

when \tilde{M}_n is formed from M_n by replacing one of the off-diagonal entries ξ_{ij} of M_n with the corresponding entry ξ'_{ij} of M'_n (and also replacing $\xi_{ji} = \overline{\xi_{ij}}$ with $\xi'_{ji} = \overline{\xi'_{ij}}$, to preserve the Hermitian property), then on summing an appropriate telescoping series, one would be able to conclude asymptotic agreement of the statistics $\mathbf{E}F(M_n)$ and $\mathbf{E}F(M'_n)$:

$$(5.2) \quad \mathbf{E}F(M_n) - \mathbf{E}F(M'_n) = o(1)$$

As it turns out, the numerology of swapping for matrices is similar to that for the central limit theorem, in that each matching moment leads to an additional factor of $O(n^{-1/2})$ in the error estimates. From this, one can expect to obtain asymptotics of the form (5.2) when the entries of M_n, M'_n match to second order on the diagonal and to fourth order off the diagonal; informally, this would mean that $\mathbf{E}F(M_n)$ depends only on the first four moments of the entries (and the first two moments of the diagonal entries). In the case of statistics arising from eigenvalues or eigenvectors, this is indeed the case, and the precise statement is known as the *Four Moment Theorem*.

We first state the Four Moment Theorem for eigenvalues.

DEFINITION 5.1 (Matching moments). Let $k \geq 1$. Two complex random variables ξ, ξ' are said to *match to order k* if one has $\mathbf{E}\operatorname{Re}(\xi)^a \operatorname{Im}(\xi)^b = \mathbf{E}\operatorname{Re}(\xi')^a \operatorname{Im}(\xi')^b$ whenever $a, b \geq 0$ are integers such that $a + b \leq k$.

In the model case when the real and imaginary parts of ξ or of ξ' are independent, the matching moment condition simplifies to the assertion that $\mathbf{E}\operatorname{Re}(\xi)^a = \mathbf{E}\operatorname{Re}(\xi')^a$ and $\mathbf{E}\operatorname{Im}(\xi)^b = \mathbf{E}\operatorname{Im}(\xi')^b$ for all $0 \leq a, b \leq k$.

THEOREM 5.2 (Four Moment Theorem for eigenvalues). *Let $c_0 > 0$ be a sufficiently small constant. Let $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$ and $M'_n = (\xi'_{ij})_{1 \leq i, j \leq n}$ be two Wigner matrices obeying Condition **C1** for some sufficiently large absolute constant C_0 . Assume furthermore that for any $1 \leq i < j \leq n$, ξ_{ij} and ξ'_{ij} match to order 4 and for any $1 \leq i \leq n$, ξ_{ii} and ξ'_{ii} match to order 2. Set $A_n := \sqrt{n}M_n$ and $A'_n := \sqrt{n}M'_n$, let $1 \leq k \leq n^{c_0}$ be an integer, and let $G : \mathbb{R}^k \rightarrow \mathbb{R}$ be a smooth function obeying the derivative bounds*

$$(5.3) \quad |\nabla^j G(x)| \leq n^{c_0}$$

for all $0 \leq j \leq 5$ and $x \in \mathbb{R}^k$. Then for any $1 \leq i_1 < i_2 \cdots < i_k \leq n$, and for n sufficiently large we have

$$(5.4) \quad |\mathbf{E}(G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))) - \mathbf{E}(G(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n)))| \leq n^{-c_0}.$$

We remark that in the papers [108], [103], [109], a variant of the above result, which we called the *three moment theorem*, was asserted, in which the hypothesis of four matching moments off the diagonal was relaxed to three matching moments

¹⁷Technically, the matrices \tilde{M}_n formed by such a swapping procedure are not Wigner matrices as defined in Definition 1.1, because the diagonal or upper-triangular entries are no longer identically distributed. However, all of the relevant estimates for Wigner matrices can be extended to the non-identically-distributed case at the cost of making the notation slightly more complicated. As this is a relatively minor issue, we will not discuss it further here.

(and no moment matching was required on the diagonal), but for which the bound (5.3) was improved to $|\nabla^j G(x)| \leq n^{-Cj c_0}$ for some sufficiently large absolute constant $C > 0$. Unfortunately, the proof given of the three moment theorem in these papers was not correct as stated, although the claim can still be proven in most cases by other means; see Appendix A.

A preliminary version of Theorem 5.2 was first established by the authors in [108], in the case¹⁸ of bulk eigenvalues (thus $\delta n \leq i_1, \dots, i_k \leq (1 - \delta)n$ for some absolute constant $\delta > 0$) and assuming Condition **C0** instead of Condition **C1**. In [103], the restriction to the bulk was removed; and in [109], Condition **C0** was relaxed to Condition **C1** for a sufficiently large value of C_0 . We will discuss the proof of this theorem in Section 6.

The following technical generalization of the Four Moment Theorem, in which the entries of M_n, M'_n only match approximately rather than exactly, is useful for some applications.

PROPOSITION 5.3 (Four Moment Theorem, approximate moment matching case). *The conclusions of Theorem 5.2 continue to hold if the requirement that ξ_{ij} and ξ'_{ij} match to order 4 is relaxed to the conditions*

$$|\mathbf{ERe}(\xi_{ij})^a \mathbf{Im}(\xi_{ij})^b - \mathbf{ERe}(\xi'_{ij})^a \mathbf{Im}(\xi'_{ij})^b| \leq \varepsilon_{a+b}$$

whenever $a, b \geq 0$ and $a + b \leq 4$, where

$$\varepsilon_0 = \varepsilon_1 = \varepsilon_2 := 0; \quad \varepsilon_3 := n^{-1/2 - C c_0}; \quad \varepsilon_4 := n^{-C c_0}$$

for some absolute constant $C > 0$.

This proposition follows from an inspection of the proof of Theorem 5.2: see Section 6.

A key technical result used in the proof of the Four Moment Theorem, which is also of independent interest, is the *gap theorem*:

THEOREM 5.4 (Gap theorem). *Let M_n be a Wigner matrix obeying Condition **C1** for a sufficiently large absolute constant C_0 . Then for every $c_0 > 0$ there exists a $c_1 > 0$ (depending only on c_0) such that*

$$\mathbf{P}(|\lambda_{i+1}(A_n) - \lambda_i(A_n)| \leq n^{-c_0}) \ll n^{-c_1}$$

for all $1 \leq i < n$.

We discuss this theorem in Section 6. Among other things, the gap theorem tells us that eigenvalues of a Wigner matrix are usually simple. Closely related *level repulsion* estimates were established (under an additional smoothness hypothesis on the atom distributions) in [39].

Another variant of the Four Moment Theorem was subsequently introduced in [43], in which the eigenvalues $\lambda_{i_j}(A_n)$ appearing in Theorem 5.2 were replaced by expressions such as (3.4) that are derived from the resolvent (or Green's function) $(W_n - z)^{-1}$, but with slightly different technical hypotheses on the matrices M_n, M'_n ; see [43] for full details. As the resolvent-based quantities (3.4) are averaged statistics that sum over many eigenvalues, they are far less sensitive to the eigenvalue repulsion phenomenon than the individual eigenvalues, and as such the version of the Four Moment Theorem for Green's function has a somewhat simpler

¹⁸In the paper, k was held fixed, but an inspection of the argument reveals that it extends without difficulty to the case when k is as large as n^{c_0} , for c_0 small enough.

proof (based on resolvent expansions rather than the Hadamard variation formulae and Taylor expansion). Conversely, though, to use the Four Moment Theorem for Green's function to control individual eigenvalues, while possible, requires a significant amount of additional argument; see [71].

We now discuss the extension of the Four Moment Theorem to eigenvectors rather than eigenvalues. Recall that we are using $u_1(M_n), \dots, u_n(M_n)$ to denote the unit eigenvectors of a Hermitian matrix M_n associated to the eigenvalues $\lambda_1(M_n), \dots, \lambda_n(M_n)$, thus $u_1(M_n), \dots, u_n(M_n)$ lie in the unit sphere $S^{2n-1} := \{z \in \mathbb{C}^n : |z| = 1\}$ is the unit sphere of \mathbb{C}^n . We write $u_{i,p}(M_n) = u_i(M_n)^* e_p$ for the p -th coefficient of $u_i(M_n)$ for each $1 \leq i, p \leq n$. If M_n is not Hermitian, but is in fact real symmetric, then we can require the $u_i(M_n)$ to have real coefficients, thus taking values in the unit sphere $S^{n-1} := \{x \in \mathbb{R}^n : |x| = 1\}$ of \mathbb{R}^n .

Unfortunately, the eigenvectors $u_i(M_n)$ are not unique in either the Hermitian or real symmetric cases; even if one assumes that the spectrum of M_n is *simple*, in the sense that

$$\lambda_1(M_n) < \dots < \lambda_n(M_n),$$

one has the freedom to rotate each $u_i(M_n)$ by a unit phase $e^{\sqrt{-1}\theta} \in U(1)$. In the real symmetric case, in which we force the eigenvectors to have real coefficients, one only has the freedom to multiply each $u_i(M_n)$ by a sign $\pm \in O(1)$. There are a variety of ways to eliminate this ambiguity. For sake of concreteness we will remove the ambiguity by working with the orthogonal projections $P_i(M_n)$ to the eigenspace at eigenvalue $\lambda_i(M_n)$; if this eigenvalue is simple, we simply have $P_i(M_n) := u_i(M_n)u_i(M_n)^*$.

THEOREM 5.5 (Four Moment Theorem for eigenvectors). *Let $c_0, M_n, M'_n, C_0, A_n, A'_n, k$ be as in Theorem 5.2. Let $G : \mathbb{R}^k \times \mathbb{C}^k \rightarrow \mathbb{R}$ be a smooth function obeying the derivative bounds*

$$(5.5) \quad |\nabla^j G(x)| \leq n^{c_0}$$

for all $0 \leq j \leq 5$ and $x \in \mathbb{R}^k \times \mathbb{C}^k$. Then for any $1 \leq i_1, i_2, \dots, i_k \leq n$ and $1 \leq p_1, \dots, p_k, q_1, \dots, q_k \leq n$, and for n sufficiently large depending on ε, c_0, C_0 we have

$$(5.6) \quad |\mathbf{E}G(\Phi(A_n)) - \mathbf{E}G(\Phi(A'_n))| \leq n^{-c_0}$$

where for any matrix M of size n , $\Phi(M) \in \mathbb{R}^k \times \mathbb{C}^k$ is the tuple

$$\Phi(M) := ((\lambda_{i_a}(M))_{1 \leq a \leq k}, (nP_{i_a, p_a, q_a}(M))_{1 \leq a \leq k}),$$

and $P_{i,p,q}(M)$ is the pq coefficient of the projection $P_i(M)$. The bounds are uniform in the choice of $i_1, \dots, i_k, p_1, \dots, p_k, q_1, \dots, q_k$.

Theorem 5.5 extends (the first part of) Theorem 5.5, which deals with the case where the function G only depends on the \mathbb{R}^k component of $\mathbb{R}^k \times \mathbb{C}^k$. This theorem

as stated appears in [107]; a slight variant¹⁹ of the theorem (proven via the Four Moment Theorem for Green’s function) was simultaneously²⁰ established in [71].

We also remark that the Four Moment Theorem for eigenvectors (in conjunction with the eigenvalue rigidity bound (2.13)) can be used to establish a variant of the Four Moment Theorem for Green’s function, which has the advantage of being applicable all the way up to the real axis (assuming a level repulsion hypothesis); see [113].

5.1. The necessity of Four Moments. It is a natural question to ask whether the requirement of four matching moments (or four approximately matching moments, as in Proposition 5.3) is genuinely necessary. As far as the distribution of individual eigenvalues $\lambda_i(A_n)$ are concerned, the answer is essentially “yes”, even in the identically distributed case, as the following result from [106] shows.

THEOREM 5.6 (Necessity of fourth moment hypothesis). *Let M_n, M'_n be real symmetric Wigner matrices whose atom variables ξ, ξ' have vanishing third moment $\mathbf{E}\xi^3 = \mathbf{E}(\xi')^3 = 0$ but with distinct fourth moments $\mathbf{E}\xi^4 \neq \mathbf{E}(\xi')^4$. Then for all sufficiently large n , one has*

$$\frac{1}{n} \sum_{i=1}^n |\mathbf{E}\lambda_i(A_n) - \mathbf{E}\lambda_i(A'_n)| \geq \kappa$$

for some κ depending only on the atom distributions.

This result is established by combining a computation of the fourth moment $\sum_{i=1}^n \lambda_i(A_n)^4$ with eigenvalue rigidity estimates such as (2.13). Informally, it asserts that on average, the *mean* value of $\lambda_i(A_n)$ is sensitive to the fourth moment of the atom distributions at the scale of the mean eigenvalue spacing (which is comparable to 1 in the bulk at least). In contrast, the Four Moment Theorem morally²¹ asserts that when the atom variables of M_n and M'_n match to fourth order, then the *median* of $\lambda_i(A_n)$ and of $\lambda_i(A'_n)$ only differ by $O(n^{-c_0})$. Thus, Theorem 5.6 and Theorem 5.2 are not directly comparable to each other. Nevertheless it is expected that the mean and median of $\lambda_i(A_n)$ should be asymptotically equal at the level of the mean eigenvalue spacing, although this is just beyond the known results (such as (2.13)) on the distribution of these eigenvalues. As such, Theorem 5.6 provides

¹⁹Besides the differences in the methods of proof, the hypotheses of the result in [71] differ in some technical aspects from those in Theorem 5.5. For instance, Condition **C1** is replaced with Condition **C0**, and k is restricted to be bounded, rather than being allowed to be as large as n^{c_0} . On the other hand, the result is sharper at the edge of the spectrum (one only requires matching up to two moments, rather than up to four), and the result can be extended to “generalized Wigner matrices” for which the variances of the entries are allowed to differ, provided that a suitable analogue of Theorem 5.4 holds.

²⁰More precisely, the results in [107] were announced at the AIM workshop “Random matrices” in December 2010 and appeared on the arXiv in March 2011. A preliminary version of the results in [71] appeared in February 2011, with a final version appearing in March 2011.

²¹This is not quite true as stated, because of the various error terms in the Four Moment Theorem, and the requirement that the function G in that theorem is smooth. A more accurate statement (cf. the proof of Theorem 7.1 below) is that if the median of $\lambda_i(A_n)$ is M (thus $\mathbf{P}(\lambda_i(A_n) \leq M) = 1/2$, in the continuous case at least), then one has

$$\mathbf{P}(\lambda_i(A'_n) \leq M + n^{-c_0}), \mathbf{P}(\lambda_i(A'_n) \geq M - n^{-c_0}) \geq 1/2 - n^{-c_0},$$

which almost places the median of $\lambda_i(A'_n)$ within $O(n^{-c_0})$ of M .

substantial evidence that the Four Moment Theorem breaks down if one does not have any sort of matching at the fourth moment.

By computing higher moments of $\lambda_i(A_n)$, it was conjectured in [106] that one has an asymptotic of the form

$$(5.7) \quad \mathbf{E}\lambda_i(A_n) = n\lambda_i^{\text{cl}}(W_n) + C_{i,n} + \frac{1}{4}(\lambda_i^{\text{cl}}(W_n)^3 - 2\lambda_i^{\text{cl}}(W_n))\mathbf{E}\xi^4 + O(n^{-c})$$

for all i in the bulk region $\delta n \leq i \leq (1-\delta)n$, where $C_{i,n}$ is a quantity independent of the atom distribution ξ . (At the edge, the dependence on the fourth moment is weaker, at least when compared against the (now much wider) mean eigenvalue spacing; see Section 7.1.)

We remark that while the statistics of *individual* eigenvalues are sensitive to the fourth moment, *averaged* statistics such as the k -point correlation functions $\rho_{n,u}^{(k)}$ are much less sensitive to this moment (or the third moment). Indeed, this is already visible from the results in Section 4, as gauss divisible matrices can have a variety of possible third or fourth moments for their atom distributions (see Lemma 7.6).

6. Sketch of proof of four moment theorem

In this section we discuss the proof of Theorem 5.2 and Theorem 5.4, following the arguments that originated in [108] and refined in [109]. To simplify the exposition, we will just discuss the four moment theorem; the proof of the approximate four moment theorem in Proposition 5.3 is established by a routine modification of the argument.

For technical reasons, the two theorems need to be proven together. Let us say that a Wigner matrix M_n has the *gap property* if it obeys the conclusion of Theorem 5.4; thus Theorem 5.4 asserts that all Wigner matrices obeying Condition **C1** for sufficiently large C_0 have the gap property. We do not know of a direct proof of this result that does not go through the Four Moment Theorem; however, it is possible to establish an independent proof of a more restrictive result:

THEOREM 6.1 (Gap theorem, special case). *Any Wigner matrix obeying Condition **C0** has the gap property.*

We discuss this theorem (which is [108, Theorem 19]) later in this section. Another key ingredient is the following truncated version of the Four Moment Theorem, in which one removes the event that two consecutive eigenvalues are too close to each other. For technical reasons, we need to introduce quantities

$$Q_i(A_n) := \sum_{j \neq i} \frac{1}{|\lambda_j(A_n) - \lambda_i(A_n)|^2}$$

for $i = 1, \dots, n$, which is a regularised measure of extent to which $\lambda_i(A_n)$ is close to any other eigenvalue of A_n .

THEOREM 6.2 (Truncated Four Moment Theorem). *Let $c_0 > 0$ be a sufficiently small constant. Let $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$ and $M'_n = (\xi'_{ij})_{1 \leq i, j \leq n}$ be two Wigner matrices obeying Condition **C1** for some sufficiently large absolute constant C_0 . Assume furthermore that for any $1 \leq i < j \leq n$, ξ_{ij} and ξ'_{ij} match to order 4 and for any $1 \leq i \leq n$, ξ_{ii} and ξ'_{ii} match to order 2. Set $A_n := \sqrt{n}M_n$ and $A'_n := \sqrt{n}M'_n$, let*

$1 \leq k \leq n^{c_0}$ be an integer, and let

$$G = G(\lambda_{i_1}, \dots, \lambda_{i_k}, Q_{i_1}, \dots, Q_{i_k})$$

be a smooth function from $\mathbb{R}^k \times \mathbb{R}_+^k$ to \mathbb{R} that is supported in the region

$$(6.1) \quad Q_{i_1}, \dots, Q_{i_k} \leq n^{c_0}$$

and obeys the derivative bounds

$$(6.2) \quad |\nabla^j G(\lambda_{i_1}, \dots, \lambda_{i_k}, Q_{i_1}, \dots, Q_{i_k})| \leq n^{c_0}$$

for all $0 \leq j \leq 5$. Then

$$(6.3) \quad \begin{aligned} & \mathbf{E}G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n), Q_{i_1}(A_n), \dots, Q_{i_k}(A_n)) = \\ & \mathbf{E}G(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n), Q_{i_1}(A'_n), \dots, Q_{i_k}(A'_n)) + O(n^{-1/2+O(c_0)}). \end{aligned}$$

We will discuss the proof of this theorem shortly. Applying Theorem 6.2 with $k = 1$ and a function G that depends only a single variable Q_i , and using the gap property to bound Q_i (cf. [108, Lemma 49]), one can show a four moment property for the gap theorem: if M_n, M'_n are Wigner matrices obeying Condition **C1** for a sufficiently large C_0 which match to fourth order, and M_n obeys the gap property, then M'_n also obeys the gap property. Using this and Theorem 6.1, one can then obtain Theorem 5.4 in full generality. Using Theorem 5.4, one can then deduce Theorem 5.2 from Theorem 6.2 by smoothly truncating in the Q variables: see [108, §3.3].

It remains to establish Theorem 6.1 and Theorem 6.2. We begin with Theorem 6.2. To simplify the exposition slightly, let us assume that the matrices M_n, M'_n are real symmetric rather than Hermitian. To reduce the number of parameters, we will also set $C_0 := 1/c_0$.

As indicated in Section 5, the basic idea is to use the Lindeberg exchange strategy. To illustrate the idea, let \tilde{M}_n be the matrix formed from M_n by replacing a single entry ξ_{pq} of M_n with the corresponding entry ξ'_{pq} of M'_n for some $p < q$, with a similar swap also being performed at the ξ_{qp} entry to keep \tilde{M}_n Hermitian. Strictly speaking, \tilde{M}_n is not a Wigner matrix as defined in Definition 1.1, as the entries are no longer identically distributed, but this will not significantly affect the arguments. (One also needs to perform swaps on the diagonal, but this can be handled in essentially the same manner.)

Set $\tilde{A}_n := \sqrt{n}\tilde{M}_n$ as usual. We will sketch the proof of the claim that

$$\begin{aligned} & \mathbf{E}G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n), Q_{i_1}(A_n), \dots, Q_{i_k}(A_n)) \\ & = \mathbf{E}G(\lambda_{i_1}(\tilde{A}_n), \dots, \lambda_{i_k}(\tilde{A}_n), Q_{i_1}(\tilde{A}_n), \dots, Q_{i_k}(\tilde{A}_n)) + O(n^{-5/2+O(c_0)}); \end{aligned}$$

by telescoping together $O(n^2)$ estimates of this sort one can establish (6.3). (For swaps on the diagonal, one only needs an error term of $O(n^{-3/2+O(c_0)})$, since there are only $O(n)$ swaps to be made here rather than $O(n^2)$. This is ultimately why there are two fewer moment conditions on the diagonal than off it.)

We can write $A_n = A(\xi_{pq})$, $\tilde{A}_n = A(\xi'_{pq})$, where

$$A(t) = A(0) + tA'(0)$$

is a (random) Hermitian matrix depending linearly²² on a real parameter t , with $A(0)$ being a Wigner matrix with one entry (and its adjoint) zeroed out, and $A'(0)$ is the explicit elementary Hermitian matrix

$$(6.4) \quad A'(0) = e_p e_q^* + e_p^* e_q.$$

We note the crucial fact that the random matrix $A(0)$ is independent of both ξ_{pq} and ξ'_{pq} . Note from Condition **C1** that we expect ξ_{pq}, ξ'_{pq} to have size $O(n^{O(c_0)})$ most of the time, so we should (heuristically at least) be able to restrict attention to the regime $t = O(n^{O(c_0)})$. If we then set

$$(6.5) \quad F(t) := \mathbf{E}G(\lambda_{i_1}(A(t)), \dots, \lambda_{i_k}(A(t)), Q_{i_1}(A(t)), \dots, Q_{i_k}(A(t)))$$

then our task is to show that

$$(6.6) \quad \mathbf{E}F(\xi_{pq}) = \mathbf{E}F(\xi'_{pq}) + O(n^{-5/2+O(c_0)}).$$

Suppose that we have Taylor expansions of the form

$$(6.7) \quad \lambda_{i_l}(A(t)) = \lambda_{i_l}(A(0)) + \sum_{j=1}^4 c_{l,j} t^j + O(n^{-5/2+O(c_0)})$$

for all $t = O(n^{O(c_0)})$ and $l = 1, \dots, k$, where the Taylor coefficients $c_{l,j}$ have size $c_{l,j} = O(n^{-j/2+O(c_0)})$, and similarly for the quantities $Q_{i_l}(A(t))$. Then by using the hypothesis (6.2) and further Taylor expansion, we can obtain a Taylor expansion

$$F(t) = F(0) + \sum_{j=1}^4 f_j t^j + O(n^{-5/2+O(c_0)})$$

for the function $F(t)$ defined in (6.5), where the Taylor coefficients f_j have size $f_j = O(n^{-j/2+O(c_0)})$. Setting t equal to ξ_{pq} and taking expectations, and noting that the Taylor coefficients f_j depend only on F and $A(0)$ and is thus independent of ξ_{ij} , we conclude that

$$\mathbf{E}F(\xi_{pq}) = \mathbf{E}F(0) + \sum_{j=1}^4 (\mathbf{E}f_j)(\mathbf{E}\xi_{pq}^j) + O(n^{-5/2+O(c_0)})$$

and similarly for $\mathbf{E}F(\xi'_{pq})$. If ξ_{pq} and ξ'_{pq} have matching moments to fourth order, this gives (6.6). (Note that a similar argument also would give the Three Moment Theorem, as well as Proposition 5.3.)

It remains to establish (6.7) (as well as the analogue for $Q_{i_l}(A(t))$, which turns out to be analogous). We abbreviate i_l simply as i . By Taylor's theorem with remainder, it would suffice to show that

$$(6.8) \quad \frac{d^j}{dt^j} \lambda_i(A(t)) = O(n^{-j/2+O(c_0)})$$

²²If we were working with Hermitian matrices rather than real symmetric matrices, then one could either swap the real and imaginary parts of the ξ_{ij} separately (exploiting the hypotheses that these parts were independent), or else repeat the above analysis with t now being a complex parameter (or equivalently, two real parameters) rather than a real one. In the latter case, one needs to replace all instances of single variable calculus below (such as Taylor expansion) with double variable calculus, but aside from notational difficulties, it is a routine matter to perform this modification.

for $j = 1, \dots, 5$. As it turns out, this is not quite true as stated, but it becomes true (with overwhelming probability²³) if one can assume that $Q_i(A(t))$ is bounded by $n^{O(c_0)}$. In principle, one can reduce to this case due to the restriction (6.1) on the support of G , although there is a technical issue because one will need to establish the bounds (6.8) for values of t other than ξ_{pq} or $\tilde{\xi}_{pq}$. This difficulty can be overcome by a continuity argument; see [108]. For the purposes of this informal discussion, we shall ignore this issue and simply assume that we may restrict to the case where

$$(6.9) \quad Q_i(A(t)) \ll n^{O(c_0)}.$$

In particular, the eigenvalue $\lambda_i(A(t))$ is simple, which ensures that all quantities depend smoothly on t (locally, at least).

To prove (6.8), one can use the classical *Hadamard variation formulae* for the derivatives of $\lambda_i(A(t))$, which can be derived for instance by repeatedly differentiating the eigenvector equation $A(t)u_i(A(t)) = \lambda_i(A(t))u_i(A(t))$. The formula for the first derivative is

$$\frac{d}{dt} \lambda_i(A(t)) = u_i(A(t))^* A'(0) u_i(A(t)).$$

But recall from eigenvalue delocalisation (Corollary 2.4) that with overwhelming probability, all coefficients of $u_i(A(t))$ have size $O(n^{-1/2+o(1)})$; given the nature of the matrix (6.4), we can then obtain (6.8) in the $j = 1$ case.

Now consider the $j = 2$ case. The second derivative formula reads

$$\frac{d^2}{dt^2} \lambda_i(A(t)) = -2 \sum_{j \neq i} \frac{|u_i(A(t))^* A'(0) u_j(A(t))|^2}{\lambda_j(A(t)) - \lambda_i(A(t))}$$

(compare with the formula (4.1) for Dyson Brownian motion). Using eigenvalue delocalisation as before, we see with overwhelming probability that the numerator is $O(n^{-1+o(1)})$. To deal with the denominator, one has to exploit the hypothesis (6.9) and the local semicircle law (Theorem 2.3). Using these tools, one can conclude (6.8) in the $j = 2$ case with overwhelming probability.

It turns out that one can continue this process for higher values of j , although the formulae for the derivatives for $\lambda_i(A(t))$ (and related quantities, such as $P_i(A(t))$ and $Q_i(A(t))$) become increasingly complicated, being given by a certain recursive formula in j . See [108] for details.

Now we briefly discuss the proof²⁴ of Theorem 6.1. For sake of discussion we restrict attention to the bulk case $\varepsilon n \leq i \leq (1 - \varepsilon)n$; the changes needed to deal with the edge case are relatively minor and are discussed in [103]. The objective here is to limit the probability of the event that the quantity $\lambda_{i+1}(A_n) - \lambda_i(A_n)$

²³Technically, each value of t has a different exceptional event of very small probability for which the estimates fail. Since there are uncountably many values of t , this could potentially cause a problem when applying the union bound. In practice, though, it turns out that one can restrict t to a discrete set, such as the multiples of n^{-100} , in which case the union bound can be applied without difficulty. See [108] for details.

²⁴The argument here is taken from [108]. In the case when the atom distributions are sufficiently smooth, one can also deduce this result from the level repulsion argument in [39, Theorem 3.5] and the eigenvalue rigidity estimate (2.13), and by using Theorem 6.2 one can extend the gap property to several other Wigner ensembles. However, this argument does not cover the case of Bernoulli ensembles, which is perhaps the most difficult case of Theorem 5.4 or Theorem 6.1.

is unexpectedly small. The main difficulty here is the fact that one is comparing two adjacent eigenvalues. If instead one was bounding $\lambda_{i+k}(A_n) - \lambda_i(A_n)$ for a larger value of k , say $k \geq \log^C n$ for a large value of C , then one could obtain such a bound from the local semicircle law (Theorem 2.3) without much difficulty. To reduce k all the way down to 1, the idea is to exploit the following phenomenon:

If $\lambda_{i+1}(A_n) - \lambda_i(A_n)$ is small, then $\lambda_{i+1}(A_{n-1}) - \lambda_{i-1}(A_{n-1})$ is also likely to be small.

Here A_{n-1} denotes²⁵ the top left $(n-1) \times (n-1)$ minor of A_n . This phenomenon can be viewed as a sort of converse to the classical *Cauchy interlacing law*

$$(6.10) \quad \lambda_{i-1}(A_{n-1}) \leq \lambda_i(A_n) \leq \lambda_i(A_{n-1}) \leq \lambda_{i+1}(A_n) \leq \lambda_{i+1}(A_{n-1})$$

(cf. (2.10)), since this law clearly shows that $\lambda_{i+1}(A_n) - \lambda_i(A_n)$ will be small whenever $\lambda_{i+1}(A_{n-1}) - \lambda_{i-1}(A_{n-1})$ is. In principle, if one iterates (generalisations of) the above principle $k = \log^C n$ times, one eventually reaches an event that can be shown to be highly unlikely by the local semicircle law.

To explain why we expect such a phenomenon to be true, let us expand A_n as

$$A_n = \begin{pmatrix} A_{n-1} & X \\ X^* & \sqrt{n}\xi_{nn} \end{pmatrix}$$

where $X \in \mathbb{C}^{n-1}$ is the random vector with entries $\sqrt{n}\xi_{nj}$ for $j = 1, \dots, n-1$. By expanding out the eigenvalue equation $A_n u_i(A_n) = \lambda_i(A_n) u_i(A_n)$, one eventually obtains the formula

$$(6.11) \quad \sum_{j=1}^{n-1} \frac{|u_j(A_{n-1})^* X|^2}{\lambda_j(A_{n-1}) - \lambda_i(A_n)} = \sqrt{n}\xi_{nn} - \lambda_i(A_n)$$

that relates $\lambda_i(A_n)$ to the various eigenvalues $\lambda_j(A_{n-1})$ of A_{n-1} (ignoring for sake of discussion the non-generic case when one or more of the denominators in (6.11) vanish); compare with (2.14). Using concentration of measure tools (such as Talagrand's inequality, see e.g. [76]), one expects $|u_j(A_{n-1})^* X|^2$ to concentrate around its mean, which can be computed to be $n(n-1)$. In view of this and, one expects the largest (and thus, presumably, the most dominant) terms in (6.11) to be the summands on the left-hand side when j is equal to either $i-1$ or i . In particular, if $\lambda_{i+1}(A_n) - \lambda_i(A_n)$ is unexpectedly small (e.g. smaller than n^{-c_0}), then by (6.10) $\lambda_i(A_{n-1}) - \lambda_i(A_n)$ is also small. This causes the $j = i$ summand in (6.11) to (usually) be large and positive; to counterbalance this, one then typically expects the $j = i-1$ summand to be large and negative, so that $\lambda_i(A_n) - \lambda_{i-1}(A_{n-1})$ is small; in particular, $\lambda_i(A_{n-1}) - \lambda_{i-1}(A_{n-1})$ is small. A similar heuristic argument (based on (6.11) but with $\lambda_i(A_n)$ replaced by $\lambda_{i+1}(A_n)$) predicts that $\lambda_{i+1}(A_{n-1}) - \lambda_i(A_{n-1})$ is also small; summing, we conclude that $\lambda_{i+1}(A_{n-1}) - \lambda_{i-1}(A_{n-1})$ is also small, thus giving heuristic support to the above phenomenon.

One can make the above arguments more rigorous, but the details are rather complicated. One of the complications arises from the slow decay of the term $\frac{1}{\lambda_j(A_{n-1}) - \lambda_i(A_n)}$ as i moves away from j . Because of this, a large positive term (such as the $j = i$ summand) in (6.11) need not be balanced primarily by the negative $j = i-1$ summand, but instead by a dyadic block $i - 2^k \leq j < i - 2^{k-1}$ of such summands; but this can be addressed by replacing the gap $\lambda_{i+1}(A_n) - \lambda_i(A_n)$

²⁵Strictly speaking, one has to multiply A_{n-1} also by $\frac{\sqrt{n-1}}{\sqrt{n}}$ to be consistent with our conventions for M_n and M_{n-1} , although this factor turns out to make very little difference to the analysis.

by a more complicated quantity (called the *regularized gap* in [108]) that is an infimum of a moderately large number of (normalised) gaps $\lambda_{i_+}(A_n) - \lambda_{i_-}(A_n)$. A more serious issue is that the numerators $|u_j(A_{n-1})^* X|$ can sometimes be much smaller than their expected value of $\sim n$, which can cause the gap at A_{n-1} to be significantly larger than that at A_n . By carefully counting all the possible cases and estimating all the error probabilities, one can still keep the net error of this situation to be of the form $O(n^{-c})$ for some $c > 0$. It is in this delicate analysis that one must rely rather heavily on the exponential decay hypothesis in Condition **C0**, as opposed to the polynomial decay hypothesis in Condition **C1**.

This concludes the sketch of Theorem 6.1. We remarked earlier that the extension to the edge case is fairly routine. In part, this is because the expected eigenvalue gap $\lambda_{i+1}(A_n) - \lambda_i(A_n)$ becomes much wider at the edge (as large as $n^{1/3}$, for instance, when $i = 1$ or $i = n - 1$), and so Theorem 6.1 and Theorem 5.4 becomes a weaker statement. There is however an interesting “bias” phenomenon that is worth pointing out at the edge, for instance with regard with the interlacing

$$(6.12) \quad \lambda_{n-1}(A_n) \leq \lambda_{n-1}(A_{n-1}) \leq \lambda_n(A_n)$$

of the very largest eigenvalues. On the one hand, the gap $\lambda_n(A_n) - \lambda_{n-1}(A_n)$ between the top two eigenvalues of A_n is expected (and known, in many cases) to be comparable to $n^{1/3}$ on the average; see (7.2) below. On the other hand, from the semi-circular law one expects $\lambda_n(A_n)$ to grow like $2n$, which suggests that $\lambda_n(A_n) - \lambda_{n-1}(A_{n-1})$ should be comparable to 1, rather than to $n^{1/3}$. In other words, the interlacing (6.12) is *biased*; the intermediate quantity $\lambda_{n-1}(A_{n-1})$ should be far closer to the right-most quantity $\lambda_n(A_n)$ than the left-most quantity $\lambda_{n-1}(A_n)$. This bias can in fact be demonstrated by using the fundamental equation (6.11); the point is that in the edge case (when i is close to n) the term $-\lambda_i(A_n)$ on the right-hand side plays a major role, and has to be balanced by $\lambda_i(A_n) - \lambda_i(A_{n-1})$ being as small as $O(1)$.

This bias phenomenon is not purely of academic interest; it turns out to be an essential ingredient in the proof of eigenvalue delocalisation (Corollary 2.4) in the edge case, as discussed in Section 2. See [103] for more discussion. It would be of interest to understand the precise relationship between the various eigenvalues in (6.10) or (6.12); the asymptotic joint distribution for, say, $\lambda_i(A_n)$ and $\lambda_i(A_{n-1})$ is currently not known, even in the GUE case.

7. Applications

By combining the heat flow methods with swapping tools such as the Four Moment Theorem, one can extend a variety of results from the GUE (or gauss divisible) regime to wider classes of Wigner ensembles. We now give some examples of such extensions.

7.1. Distribution of individual eigenvalues. One of the simplest instances of the method arises when extending the central limit theorem (3.10) of Gustavsson [57] for eigenvalues $\lambda_i(A_n)$ in the bulk from GUE to more general ensembles:

THEOREM 7.1. *The gaussian fluctuation law (3.10) continues to hold for Wigner matrices obeying Condition **C1** for a sufficiently large C_0 , and whose atom distributions match that of GUE to second order on the diagonal and fourth order off the*

diagonal; thus, one has

$$\frac{\lambda_i(A_n) - \lambda_i^{\text{cl}}(A_n)}{\sqrt{\log n / 2\pi^2 / \rho_{\text{sc}}(u)}} \rightarrow N(0, 1)_{\mathbb{R}}$$

whenever $\lambda_i^{\text{cl}}(A_n) = n(u + o(1))$ for some fixed $-2 < u < 2$.

PROOF. Let M'_n be drawn from GUE, thus by (3.10) one already has

$$\frac{\lambda_i(A'_n) - \lambda_i^{\text{cl}}(A_n)}{\sqrt{\log n / 2\pi^2 / \rho_{\text{sc}}(u)}} \rightarrow N(0, 1)_{\mathbb{R}}$$

(note that $\lambda_i^{\text{cl}}(A_n) = \lambda_i^{\text{cl}}(A'_n)$). To conclude the analogous claim for A_n , it suffices to show that

$$(7.1) \quad \mathbf{P}(\lambda_i(A'_n) \in I_-) - n^{-c_0} \leq \mathbf{P}(\lambda_i(A_n) \in I) \leq \mathbf{P}(\lambda_i(A'_n) \in I_+) + n^{-c_0}$$

for all intervals $I = [a, b]$, and n sufficiently large, where $I_+ := [a - n^{-c_0/10}, b + n^{-c_0/10}]$ and $I_- := [a + n^{-c_0/10}, b - n^{-c_0/10}]$.

We will just prove the second inequality in (7.1), as the first is very similar. We define a smooth bump function $G : \mathbb{R} \rightarrow \mathbb{R}^+$ equal to one on I_- and vanishing outside of I_+ . Then we have

$$\mathbf{P}(\lambda_i(A_n) \in I) \leq \mathbf{E}G(\lambda_i(A_n))$$

and

$$\mathbf{E}G(\lambda_i(A'_n)) \leq \mathbf{P}(\lambda_i(A'_n) \in I)$$

On the other hand, one can choose G to obey (5.3). Thus by Theorem 5.2 we have

$$|\mathbf{E}G(\lambda_i(A_n)) - \mathbf{E}G(\lambda_i(A'_n))| \leq n^{-c_0}$$

and the second inequality in (7.1) follows from the triangle inequality. The first inequality is similarly proven using a smooth function that equals 1 on I_- and vanishes outside of I . \square

REMARK 7.2. In [57] the asymptotic joint distribution of k distinct eigenvalues $\lambda_{i_1}(M_n), \dots, \lambda_{i_k}(M_n)$ in the bulk of a GUE matrix M_n was computed (it is a gaussian k -tuple with an explicit covariance matrix). By using the above argument, one can extend that asymptotic for any fixed k to other Wigner matrices, so long as they match GUE to fourth order off the diagonal and to second order on the diagonal.

If one could extend the results in [57] to broader ensembles of matrices, such as gauss divisible matrices, then the above argument would allow some of the moment matching hypotheses to be dropped, using tools such as Lemma 7.6.

REMARK 7.3. Recently in [25], a moderate deviations property of the distribution of the eigenvalues $\lambda_i(A_n)$ was established first for GUE, and then extended to the same class of matrices considered in Theorem 7.1 by using the Four Moment Theorem. An analogue of Theorem 7.1 for real symmetric matrices (using GOE instead of GUE) was established in [84].

A similar argument to the one given in Theorem 7.1 also applies at the edge of the spectrum. For sake of discussion we shall just discuss the distribution of the

largest eigenvalue $\lambda_n(A_n)$. In the case of a GUE ensemble, this largest eigenvalue is famously governed by the *Tracy-Widom law* [115, 116], which asserts that

$$(7.2) \quad \mathbf{P}\left(\frac{\lambda_n(A_n) - 2n}{n^{1/3}} \leq t\right) \rightarrow \det(1 - T_{[t, +\infty)})$$

for any fixed $t \in \mathbb{R}$, where $T_{[t, +\infty)} : L^2([t, +\infty)) \rightarrow L^2([t, +\infty))$ is the integral operator

$$T_{[t, +\infty)} f(x) := \int_t^{+\infty} \frac{\text{Ai}(x)\text{Ai}'(y) - \text{Ai}'(x)\text{Ai}(y)}{x - y} f(y) dy$$

and $\text{Ai} : \mathbb{R} \rightarrow \mathbb{R}$ is the Airy function

$$\text{Ai}(x) := \frac{1}{\pi} \int_0^\infty \cos\left(\frac{t^3}{3} + xt\right) dt.$$

Interestingly, the limiting distribution in (7.2) also occurs in many other seemingly unrelated contexts, such as the longest increasing subsequence in a random permutation [4, 116].

It is conjectured that the Tracy-Widom law in fact holds for all Wigner matrices obeying Condition **C1** with $C_0 = 4$; this value of C_0 is optimal, as one does not expect $\lambda_1(A_n)$ to stay near $2\sqrt{n}$ without this hypothesis (see [7]). While this conjecture is not yet fully resolved, there has now been a substantial amount of partial progress on the problem [99, 95, 69, 103, 65, 44]. Soshnikov [99] was the first to obtain the Tracy-Widom law for a large class of Wigner matrices; thanks to subsequent refinements in [95, 69], we know that (7.2) holds for all Wigner matrices whose entries are iid with symmetric distribution and obeying Condition **C1** with $C_0 = 12$. On the other hand, by using the argument used to prove Theorem 7.1, one can also obtain the asymptotic (7.2) for Wigner matrices obeying Condition **C1** for a sufficiently large C_0 , provided that the entries match that of GUE to fourth order. Actually, since the asymptotic (7.2) applies at scale $n^{1/3}$ rather than at scale 1, it is possible to modify the arguments to reduce the amount of moment matching required, that one only needs the entries to match GUE to third order (which in particular subsumes the case when the distribution is symmetric); see Appendix A.

More recently, Johansson [64] established (7.2) for gauss divisible Wigner ensembles obeying Condition **C1** with the optimal decay condition $C_0 = 4$. Combining this result with the three moment theorem (and noting that any Wigner matrix can be matched up to order three with a gauss divisible matrix, see Lemma 7.6 below), one can then obtain (7.2) for any Wigner matrix obeying Condition **C1** for sufficiently large C_0 . An independent proof of this claim (which also applied to generalized Wigner matrix models in which the variance of the entries was non-constant) was also established in [44]. Finally, it was shown very recently in [71] that there is a version of the Four Moment Theorem for the edge that only requires two matching moments, which allows one to establish the Tracy-Widom law for all Wigner matrices obeying Condition **C0**.

7.2. Universality of the sine kernel for Hermitian Wigner matrices.

We now turn to the question of the extent to which the asymptotic (3.7), which asserts that the normalised k -point correlation functions $\rho_{n,u}^{(k)}$ converge to the universal limit $\rho_{\text{Sine}}^{(k)}$, can be extended to more general Wigner ensembles. A long-standing conjecture of Wigner, Dyson, and Mehta (see e.g. [80]) asserts (informally speaking) that (3.7) is valid for all fixed k , all Wigner matrices and all fixed energy levels

$-2 < u < 2$ in the bulk. (They also make the same conjecture for random symmetric and random symplectic matrices.) However, to make this conjecture precise one has to specify the nature of convergence in (3.7). For GUE, the convergence is quite strong (in the local uniform sense), but one cannot expect such strong convergence in general, particularly in the case of discrete ensembles in which $\rho_{n,u}^{(k)}$ is a discrete probability distribution (i.e. a linear combination of Dirac masses) and thus is unable to converge uniformly or pointwise to the continuous limiting distribution $\rho_{\text{Sine}}^{(k)}$. We will thus instead settle for the weaker notion of *vague convergence*. More precisely, we say that (3.7) holds in the vague sense if one has

$$(7.3) \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^k} F(x_1, \dots, x_k) \rho_{n,u}^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k = \int_{\mathbb{R}^k} F(x_1, \dots, x_k) \rho_{\text{Sine}}^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k$$

P for all continuous, compactly supported functions $F : \mathbb{R}^k \rightarrow \mathbb{R}$. By the Stone-Weierstrass theorem we may take F to be a test function (i.e. smooth and compactly supported) without loss of generality.

REMARK 7.4. Vague convergence is not the only notion of convergence studied in the literature. Another commonly studied notion of convergence is *averaged vague convergence*, in which one averages over the energy parameter u as well, thus replacing (7.3) with the weaker claim that

$$(7.4) \quad \lim_{b \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{2b} \int_{E-b}^{E+b} \int_{\mathbb{R}^k} F(x_1, \dots, x_k) \rho_{n,u}^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k du = \int_{\mathbb{R}^k} F(x_1, \dots, x_k) \rho_{\text{Sine}}^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k$$

for all $-2 < E < 2$. It can be argued (as was done in [42]) that as the original conjectures of Wigner, Dyson, and Mehta did not precisely specify the nature of convergence, that any one of these notions of convergence would be equally valid for the purposes of claiming a proof of the conjecture. However, the distinction between such notions is not purely a technical one, as certain applications of the Wigner-Dyson-Mehta conjecture are only available if the convergence notion is strong enough.

Consider for instance the *gap* problem of determining the probability that there is no eigenvalue in the interval $[-t/2n, t/2n]$ (in other words the distribution of the least singular value). This is an important problem which in the GUE case was studied in [63] and discussed in length in a number of important books in the field, including Mehta's (see [80, Chapter 5]), Deift's (see [20, Section 5.4]), Deift-Gioev's (see [21, Section 4.2]) and Anderson-Guionnet-Zeitouni (see [1, Section 3.5]), Forrester's (see [48, Section 9.6]) and the Oxford handbook of matrix theory edited by Akemann et. al. (see [58, Section 4.6] by Anderson). This distribution can be determined from the correlation functions $\rho_{n,0}^{(k)}$ at the energy level $u = 0$ by a standard combinatorial argument. In particular, if one has the Wigner-Dyson-Mehta conjecture in the sense of vague convergence, one can make the limiting law in [63] universal over the class of matrices for which that conjecture is verified; see Corollary 7.12 below. However, if one only knows the Wigner-Dyson-Mehta conjecture in the sense of averaged vague convergence (7.4) instead of vague convergence (7.3), one cannot make this conclusion, because the distributional information at

$u = 0$ is lost in the averaging process. See also Theorem 7.11 for a further example of a statistic which can be controlled by the vague convergence version of the Wigner-Dyson-Mehta conjecture, but *not* the averaged vague convergence version.

For this reason, it is our opinion that a solution of the Wigner-Dyson-Mehta conjecture in the averaged vague convergence sense should be viewed as an important partial resolution to that conjecture, but one that falls short of a complete solution to that conjecture²⁶.

As a consequence of the heat flow and swapping techniques, the Wigner-Dyson-Mehta conjecture for (Hermitian) Wigner matrices is largely resolved in the vague convergence category:

THEOREM 7.5. *Let M_n be a Wigner matrix obeying Condition **C1** for a sufficiently large absolute constant C_0 which matches moments with GUE to second order, and let $-2 < u < 2$ and $k \geq 1$ be fixed. Then (3.7) holds in the vague sense.*

This theorem has been established as a result of a long sequence of partial results towards the Wigner-Dyson-Mehta conjecture [64, 34, 35, 108, 36, 43, 44, 110], which we will summarise (in a slightly non-chronological order) below; see Remark 7.7 for a more detailed discussion of the chronology. The precise result stated above was first proven explicitly in [110], but relies heavily on the previous works just cited. As recalled in Section 3, the asymptotic (3.7) for GUE (in the sense of locally uniform convergence, which is far stronger than vague convergence) follows as a consequence of the Gaudin-Mehta formula and the Plancherel-Rotach asymptotics for Hermite polynomials²⁷.

The next major breakthrough was by Johansson [64], who, as discussed in Section 4, established (3.7) for gauss divisible ensembles at some fixed time parameter $t > 0$ independent of n , obtained (3.7) in the vague sense (in fact, the slightly stronger convergence of *weak convergence* was established in that paper, in which the function F in (7.3) was allowed to merely be L^∞ and compactly supported, rather than continuous and compactly supported). The main tool used in [64] was an explicit determinantal formula for the correlation functions in the gauss divisible case, essentially due to Brézin and Hikami [13].

In Johansson's result, the time parameter $t > 0$ had to be independent of n . It was realized by Erdős, Ramirez, Schlein, and Yau that one could obtain many further cases of the Wigner-Dyson-Mehta conjecture if one could extend Johansson's result to much shorter times t that decayed at a polynomial rate in n . This was first achieved (again in the context of weak convergence) for $t > n^{-3/4+\varepsilon}$ for an arbitrary fixed $\varepsilon > 0$ in [34], and then to the essentially optimal case $t > n^{-1+\varepsilon}$ (for weak convergence, and (implicitly) in the local L^1 sense as well) in [35]. By combining this with the method of reverse heat flow discussed in Section 5, the asymptotic (3.7) (again in the sense of weak convergence) was established for all Wigner matrices whose distribution obeyed certain smoothness conditions (e.g. when $k = 2$ one

²⁶In particular, we view this conjecture as only partially resolved in the real symmetric and symplectic cases, as opposed to the Hermitian cases, because the available results in that setting are either only in the averaged vague convergence sense, or require some additional moment matching hypotheses on the coefficients. Completing the proof of the Wigner-Dyson-Mehta conjecture in these categories remains an interesting future direction of research.

²⁷Analogous results are known for much wider classes of invariant random matrix ensembles, see e.g. [19], [86], [11]. However, we will not discuss these results further here, as they do not directly impact on the case of Wigner ensembles.

needs a C^6 type condition), and also decayed exponentially. The methods used in [35] were an extension of those in [64], combined with an approximation argument (the “method of time reversal”) that approximated a continuous distribution by a gauss divisible one (with a small value of t); the arguments in [34] are based instead on an analysis of the Dyson Brownian motion.

Note from the eigenvalue rigidity property (2.13) that only a small number of eigenvalues (at most $n^{o(1)}$ or so) make²⁸ a significant contribution to the normalised correlation function $\rho_{n,u}^{(k)}$ on any fixed compact set, and any fixed u . Because of this, the Four Moment Theorem (Theorem 5.2) can be used to show that²⁹ if one Wigner matrix M_n obeyed the asymptotics (3.7) in the vague sense, then any other Wigner matrix M'_n that matched M_n to fourth order would also obey (3.7) in the vague sense, assuming that M_n, M'_n both obeyed Condition **C1** for a sufficiently large C_0 (so that the eigenvalue rigidity and four moment theorems are applicable).

By combining the above observation with the moment matching lemma presented below, one immediately concludes Theorem 7.5 assuming that the off-diagonal atom distributions are supported on at least three points.

LEMMA 7.6 (Moment matching lemma). *Let ξ be a real random variable with mean zero, variance one, finite fourth moment, and which is supported on at least three points. Then there exists a gauss divisible, exponentially decaying real random variable ξ' that matches ξ to fourth order.*

For a proof of this elementary lemma, see [108, Lemma 28]. The requirement of support on at least three points is necessary; indeed, if ξ is supported in just two points a, b , then $\mathbf{E}(\xi - a)^2(\xi - b)^2 = 0$, and so any other distribution that matches ξ to fourth order must also be supported on a, b and thus cannot be gauss divisible.

To remove the requirement that the atom distributions be supported on at least three points, one can use the observation from Proposition 5.3 that one only needs the moments of M_n and M'_n to *approximately* match to fourth order in order to be able to transfer results on the distribution of spectra of M_n to that of M'_n . In particular, if $t = n^{-1+\varepsilon}$ for some small $\varepsilon > 0$, then the Ornstein-Uhlenbeck flow M_n^t of M_n by time t is already close enough to matching the first four moments of M_n to apply Proposition 5.3. The results of [35] give the asymptotic (3.7) for M_n^t , and the eigenvalue rigidity property (2.13) then allows one to transfer this property to M_n , giving Theorem 7.5.

REMARK 7.7. The above presentation (drawn from the most recent paper [110]) is somewhat ahistorical, as the arguments used above emerged from a sequence of papers, which obtained partial results using the best technology available at the time. In the paper [108], where the first version of the Four Moment Theorem was introduced, the asymptotic (3.7) was established under the additional

²⁸Strictly speaking, the results in [44] only establish the eigenvalue rigidity property (2.13) assuming Condition **C0**. However, by using the Four Moment Theorem one can relax this to Condition **C1** for a sufficiently large C_0 , at the cost of making (2.13) hold only with high probability rather than overwhelming probability.

²⁹Very recently, it was observed in [42] that if one uses the Green’s function Four Moment Theorem from [43] in place of the earlier eigenvalue Four Moment Theorem from [108] at this juncture, then one can reach the same conclusion here without the need to invoke the eigenvalue rigidity theorem, thus providing a further simplification to the argument.

assumptions of Condition **C0**, and matching the GUE to fourth order³⁰; the former hypothesis was due to the weaker form of the four moment theorem known at the time, and the latter was due to the fact that the eigenvalue rigidity result (2.13) was not yet established (and was instead deduced from the results of Gustavsson [57] combined with the Four Moment Theorem, thus necessitating the matching moment hypothesis). For related reasons, the paper in [36] (which first introduced the use of Proposition 5.3) was only able to establish (3.7) after an additional averaging in the energy parameter u (and with Condition **C0**). The subsequent progress in [40] via heat flow methods gave an alternate approach to establishing (3.7), but also required an averaging in the energy and a hypothesis that the atom distributions be supported on at least three points, although the latter condition was then removed in [44]. In a very recent paper [33], the exponent C_0 in Condition **C1** has been relaxed to as low as $4 + \varepsilon$ for any fixed $\varepsilon > 0$, though still at the cost of averaging in the energy parameter. Some generalisations in other directions (e.g. to covariance matrices, or to generalised Wigner ensembles with non-constant variances) were also established in [8], [109], [41], [43], [44], [32], [33], [125], [88], [89], [42]. For instance, it was very recently observed in [42] that by using a variant of the argument from [110] (replacing the asymptotic four moment theorem for eigenvalues by an asymptotic four moment theorem for the Green function) together with a careful inspection of the arguments in [35] and invoking some results from [43], one can extend Theorem 7.5 to generalised Wigner ensembles in which the entries are allowed to have variable variance (subject to some additional hypotheses); see [42] for details.

To close this section, we remark that while Theorem 7.5 is the “right” result for discrete Wigner ensembles (except for the large value of C_0 in Condition **C1**, which in view of the results in [33] should be reducible to $4 + \varepsilon$), one expects stronger notions of convergence when one has more smoothness hypotheses on the atom distribution; in particular, one should have local uniform convergence of the correlation functions when the distribution is smooth enough. Some very recent progress in this direction in the $k = 1$ case was obtained by Maltsev and Schlein [78], [79].

For results concerning symmetric³¹ and symplectic random matrices, we refer to [31] and the references therein. In these cases, the Wigner-Dyson-Mehta conjecture is established in the context of averaged vague convergence; the case of vague convergence remains open even for gauss divisible distributions (although in the case when four moments agree with GOE, one can recover the vague convergence version of the conjecture thanks to the Four Moment Theorem). This appears to be an inherent limitation to the relaxation flow method, at least with the current state of technology. In the Hermitian case, this limitation is overcome by using the explicit formulae of Brezis-Hikami and Johansson (see [64]), but no such tool appears to be available at present in the symmetric and symplectic cases. It remains an interesting future direction of research to find a way to overcome this

³⁰In [108] it was claimed that one only needed a matching condition of GUE to third order, but this was not rigorously proven in that paper due to an issue with the Three Moment Theorem that we discuss in Appendix A.

³¹With our conventions, the symmetric case is not a sub-case of the Hermitian case, because the matrix would now be required to match GOE to second order, rather than GUE to second order.

limitation and obtain a complete solution to the Wigner-Dyson-Mehta conjecture in these cases, as this could lead to a number of interesting applications, such as the determination of the asymptotic distribution of the least singular value of the adjacency matrix of an Erdős-Renyi graph, which remains open currently despite the significant partial progress (see [33]) on the Wigner-Dyson-Mehta conjecture for such matrices (though see the recent papers [81], [120] for some lower bounds relating to this problem).

7.3. Distribution of the gaps. In Section 1.2 the averaged gap distribution

$$S_n(s, u, t_n) := \frac{1}{t_n} |\{1 \leq i \leq n : |\lambda_i(A_n) - nu| \leq t_n/\rho_{\text{sc}}(u); \lambda_{i+1}(A_n) - \lambda_i(A_n) \leq s/\rho_{\text{sc}}(u)\}|.$$

was defined for a given energy level $-2 < u < 2$ and a scale window t_n with $1/t_n, t_n/n$ both going to zero as $n \rightarrow \infty$. For GUE, it is known that the expected value $\mathbf{E}S_n(s, u, t_n)$ of this distribution converges as $n \rightarrow \infty$ (keeping u, s fixed) to the (cumulative) Gaudin distribution (3.9); see [19]. Informally, this result asserts that a typical normalised gap $\rho_{\text{sc}}(u)(\lambda_{i+1}(A_n) - \lambda_i(A_n))$, where $\lambda_i^{\text{sc}}(W_n) = u + o(1)$, is asymptotically distributed on the average according to the Gaudin distribution.

The eigenvalue gap distribution has received much attention in the mathematics community, partially thanks to the fascinating (numerical) coincidence with the gap distribution of the zeros of the zeta functions. For more discussions, we refer to [20, 67, 22] and the references therein.³²

It is possible to use an inclusion-exclusion argument to deduce information about $\mathbf{E}S_n(s, u, t_n)$ from information on the k -point correlation functions; see e.g. [20], [19], [64], [35], [36]. In particular, one can establish universality of the Gaudin distribution for the Wigner matrix ensembles considered in those papers, sometimes assuming additional hypotheses due to the averaging of the correlation function in the u parameter; for instance, in [36] universality is shown for all Wigner matrices obeying Condition **C0**, assuming that t_n/n decays very slowly to zero, by utilising the universality of the averaged k -point correlation function established in that paper.

A slightly different approach proceeds by expressing the moments of $S_n(s, u, t_n)$ in terms of the joint distribution of the eigenvalues. Observe that

$$\mathbf{E}S_n(s, u, t_n) = \frac{1}{t_n} \sum_{i=1}^n \mathbf{E} \mathbf{1}_{|\lambda_i(A_n) - nu| \leq t_n/\rho_{\text{sc}}(u)} \mathbf{1}_{\lambda_{i+1}(A_n) - \lambda_i(A_n) \leq s/\rho_{\text{sc}}(u)}.$$

Replacing the sharp cutoffs in the expectation by smoothed out versions, and applying Theorem 5.2 (and also using (2.13) to effectively localise the i summation to about $O(t_n) + O(n^{o(1)})$ indices), we see that if M_n, M'_n have four matching moments and both obey Condition **C1** for a sufficiently large C_0 , then one has

$$\mathbf{E}S_n(s, u, t_n) \leq \mathbf{E}S'_n(s + o(1), u, (1 + o(1))t_n) + o(1)$$

and similarly

$$\mathbf{E}S_n(s, u, t_n) \geq \mathbf{E}S'_n(s - o(1), u, (1 - o(1))t_n) - o(1)$$

for suitable choices of decaying quantities $o(1)$. In particular, if M'_n is known to exhibit Gaudin asymptotics (3.9) for any $-2 < u < 2$ and any t_n with $1/t_n, t_n/n =$

³²We would like to thank P. Sarnak for enlightening conversations regarding the gap distribution and for constantly encouraging us to work on the universality problem.

$o(1)$, then M_n does as well. In [64], the Gaudin asymptotics (3.9) were established for gauss divisible ensembles (with fixed time parameter t) and any t_n with $1/t_n, t_n/n = o(1)$, and thus by Lemma 7.6 and the above argument we conclude that (3.9) also holds for Wigner matrices obeying Condition **C1** for a sufficiently large C_0 whose off-diagonal atom distributions are supported on at least three points. This last condition can be removed by using Proposition 5.3 as in the previous section (using the results in [35] instead of [64]), thus giving (3.9) with no hypotheses on the Wigner matrix other than Condition **C1** for a sufficiently large C_0 .

REMARK 7.8. This argument appeared previously in [108], but at the time the eigenvalue rigidity result (2.13) was not available, and the Four Moment Theorem required Condition **C0** instead of Condition **C1**, so the statement was weaker, requiring both Condition **C0** and a vanishing third moment hypothesis (for the same reason as in Remark 7.7). By averaging over all u (effectively setting $t_n = n$), the need for eigenvalue rigidity could be avoided in [108] (but note that the formulation of the averaged universality for the eigenvalue gap in [108, (5)] is not quite correct³³ as stated, due to the absence of the normalisation by $\rho_{sc}(u)$).

REMARK 7.9. A similar argument also applies to higher moments $\mathbf{E}S_n(s, u, t_n)^k$ of $S_n(s, u, t_n)$, and so in principle one can also use the moment method to obtain universal statistics for the full distribution of $S_n(s, u, t_n)$, and not just the expectation. However, this argument would require extending the results in [19], [64], or [35] to control the distribution (and not just the expectation) of $S_n(s, u, t_n)$ for GUE, gauss divisible matrices (with fixed time parameter), or gauss divisible matrices (with $t = n^{-1+\varepsilon}$) respectively.

REMARK 7.10. It is natural to ask whether the averaging over the window t_n can be dispensed with entirely. Indeed, one expects that the distribution of the individual normalised eigenvalue gaps $\rho_{sc}(u)(\lambda_{i+1}(A_n) - \lambda_i(A_n))$, where $-2 < u < 2$ is fixed and $\lambda_i^{sc}(W_n) = u + o(1)$, should asymptotically converge to the Gaudin distribution in the vague topology, without any averaging in i . The Four moment theorem allows one to readily deduce such a fact for general Wigner ensembles once one has established it for special ensembles such as GUE or gauss divisible ensembles (and to remove all moment matching and support hypotheses on the Wigner ensemble, one would need to treat gauss divisible ensembles with time parameter equal to a negative power of n). However, control of the individual normalised eigenvalue gaps in the bulk are not presently in the literature, even for GUE, though they are in principle obtainable from determinantal process methods.

³³More precisely, instead of

$$S_n(s; x) := \frac{1}{n} |\{1 \leq i \leq n : x_{i+1} - x_i \leq s\}|,$$

the gap distribution should instead be expressed as

$$S_n(s; x) := \frac{1}{n} |\{1 \leq i \leq n : x_{i+1} - x_i \leq \frac{s}{\rho_{sc}(t_i)}\}|,$$

where $t_i \in [-2, 2]$ is the classical location of the i^{th} eigenvalue:

$$\int_{-2}^{t_i} \rho_{sc}(x) dx := \frac{i}{n}.$$

7.4. Universality of the counting function and gap probability. Recall for Section 3 that one has an asymptotic (3.8) for the number of eigenvalues of a fine-scale normalised GUE matrix for an interval $I := [nu + a/\rho_{\text{sc}}(u), nu + b/\rho_{\text{sc}}(u)]$ in the bulk, where $-2 < u < 2$ and $a, b \in \mathbb{R}$ are fixed independently of n ; this asymptotic is controlled by the spectral of the Dyson sine kernel K_{Sine} on the interval $[a, b]$. In fact one can allow u to vary in n , so long as it lies inside an interval $[-2 + \varepsilon, 2 - \varepsilon]$ for some fixed $\varepsilon > 0$.

Using the four moment theorem, one can extend this asymptotic to more general Wigner ensembles.

THEOREM 7.11 (Asymptotic for N_I). *Consider a Wigner matrix satisfying Condition **C1** for a sufficiently large constant C_0 . Let $\varepsilon > 0$ and $a < b$ be independent of n . For any n , let $u = u_n$ be an element of $[-2 + \varepsilon, 2 - \varepsilon]$. Then the asymptotic (3.8) (in the sense of convergence in distribution) holds.*

PROOF. See [110, Theorem 8]. The basic idea is to use the moment method, combining Theorem 7.5 with the identity

$$\mathbf{E} \binom{N_I}{k} = \frac{1}{k!} \int_{[a,b]} \rho_{n,u}^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k$$

for any $k \geq 1$. We remark that the method also allows one to control the asymptotic joint distribution of several intervals $N_{I_1}(A_n), \dots, N_{I_k}(A_n)$; for instance, one can show that $N_I(A_n)$ and $N_J(A_n)$ are asymptotically uncorrelated if I, J have bounded lengths, lie in the bulk, and have separation tending to infinity as $n \rightarrow \infty$. We omit the details. \square

Specialising to the case $u = 0$, one can obtain universal behaviour for the probability that M_n has no eigenvalue in an interval of the form $(-t/2\sqrt{n}, t/2\sqrt{n})$ for any fixed t :

COROLLARY 7.12. *For any fixed $t > 0$, and M_n satisfy the conditions of Theorem 7.11, one has*

$$\mathbf{P}\left(\frac{1}{\sqrt{n}}M_n \text{ has no eigenvalues in } (-t/2n, t/2n)\right) \rightarrow \exp\left(\int_0^t \frac{f(x)}{x} dx\right)$$

as $n \rightarrow \infty$, where $f : \mathbb{R} \rightarrow \mathbb{R}$ is the solution of the differential equation

$$(tf'')^2 + 4(tf' - f)(tf' - f + (f')^2) = 0$$

with the asymptotics $f(t) = \frac{-t}{\pi} - \frac{t^2}{\pi^2} - \frac{t^3}{\pi^3} + O(t^4)$ as $t \rightarrow 0$.

PROOF. In the case of GUE, this was established in [1, Theorem 3.1.2], [63]. The general case then follows from Theorem 7.11. A weaker extension was established in [108], assuming matching moments with GUE to fourth order as well as Condition **C0**. \square

This corollary is the Hermitian version of the Goldstine-Von Neumann least singular value problem (see [123, 30, 91, 94, 105] for more details). Recently, some additional bounds on the least singular value of Wigner matrices were established in [81], [120].

The above universality results concerned intervals whose length was comparable to the fine-scale eigenvalue spacing (which, using the fine-scale normalisation

A_n , is $1/\rho_{\text{sc}}(u)$). One can also use the Four Moment Theorem to obtain similar universality results for the counting function on much larger scales, such as $N_I(W_n)$, where $I := [y, +\infty)$ with $y \in (-2, 2)$ in the bulk. For instance, in [57] the mean and variance of this statistic for GUE matrices was computed as

$$(7.5) \quad \begin{aligned} \mathbf{E}[N_I(W_n)] &= n \int_I \rho_{\text{sc}} + O\left(\frac{\log n}{n}\right) \\ \mathbf{Var}(N_I(W_n)) &= \left(\frac{1}{2\pi^2} + o(1)\right) \log n. \end{aligned}$$

Combining this with a general central limit theorem for determinantal processes due to Costin and Lebowitz [14], one obtains

$$(7.6) \quad \frac{N_I(W_n) - \mathbf{E}[N_I(W_n)]}{\sqrt{\mathbf{Var}(N_I(W_n))}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1).$$

and hence

$$(7.7) \quad \frac{N_I(W_n) - n \int_I \rho_{\text{sc}}}{\sqrt{\frac{1}{2\pi^2} \log n}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

This result was extended to more general Wigner ensembles:

THEOREM 7.13 (Central limit theorem for $N_I(W_n)$). *Let M_n be a Wigner matrix obeying Condition **C0** which matches the corresponding entries of GUE up to order 4. Let $y \in (-2, 2)$ be fixed, and set $I := [y, +\infty)$. Then the asymptotic (7.7) hold (in the sense of probability distributions), as do the mean and variance bounds (7.5).*

PROOF. See [18, Theorem 2]. The first component of this theorem is established using the Four Moment Theorem; the second part also uses the eigenvalue rigidity estimate (2.13). \square

7.5. Distribution of the eigenvectors. Let M_n be a matrix drawn from the gaussian orthogonal ensemble (GOE). Then the eigenvalues $\lambda_i(M_n)$ are almost surely simple, and the unit eigenvectors $u_i(M_n) \in S^{n-1} \subset \mathbb{R}^{n-1}$ are well-defined up to sign. To deal with this sign ambiguity, let us select each eigenvector $u_i(M_n)$ independently and uniformly at random among the two possible choices. Since the GOE ensemble is invariant under orthogonal transformations, the eigenvectors $u_i(M_n)$ must each be uniformly distributed on the unit sphere. As is well known, this implies that the normalised coefficients $\sqrt{n}u_{i,j}(M_n) := \sqrt{n}u_i(M_n)e_j^*$ of these eigenvectors are asymptotically normally distributed; see [61], [62] for more precise statements in this direction.

By combining the results of [61], [62] with the four moment theorem for eigenvectors (Theorem 5.5), one can extend this fact to other Wigner ensembles in [113]. Here is a typical result:

THEOREM 7.14. *Let M_n be a random real symmetric matrix obeying hypothesis **C1** for a sufficiently large constant C_0 , which matches GOE to fourth order. Assume furthermore that the atom distributions of M_n are symmetric (i.e. $\xi_{ij} \equiv -\xi_{ji}$ for all $1 \leq i, j \leq n$). Let $i = i_n$ be an index (or more precisely, a sequence of indices) between 1 and n , and let $a = a_n \in S^{n-1}$ be a unit vector in \mathbb{R}^n (or more precisely, a sequence of unit vectors). For each i , let $u_i(M_n) \in S^{n-1}$ be chosen*

randomly among all unit eigenvectors with eigenvalue $\lambda_i(M_n)$. Then $\sqrt{n}u_i(M_n) \cdot a$ tends to $N(0, 1)_{\mathbb{R}}$ in distribution as $n \rightarrow \infty$.

PROOF. See [113, Theorem 13]. \square

As an example to illustrate Theorem 7.14, we can take $a = a_n := \frac{1}{\sqrt{n}}(1, \dots, 1) \in S^{n-1}$, and $i := \lfloor n/2 \rfloor$. Then Theorem 7.14 asserts that the sum of the entries of the middle eigenvector $u_{\lfloor n/2 \rfloor}(M_n)$ is gaussian in the limit.

7.6. Central limit theorem for log-determinant. One of most natural and important matrix functionals is the *determinant*. As such, the study of determinants of random matrices has a long and rich history. The earlier papers on this study focused on the determinant $\det A_n$ of the non-Hermitian iid model A_n , where the entries ζ_{ij} of the matrix were independent random variables with mean 0 and variance 1 [93, 49, 83, 119, 90, 73, 74, 111, 55, 53, 52, 23] (see [82] for a brief survey of these results). See also the early paper of Dixon [29] treating random matrix models in which the distribution is invariant with respect to rotation of the individual rows.

In [55], Goodman considered random gaussian matrices $A_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ where the atom variables ζ_{ij} are iid standard real gaussian variables, $\zeta_{ij} \equiv N(0, 1)_{\mathbb{R}}$. He noticed that in this case the square of the determinant can be expressed as the product of independent chi-square variables. Therefore, its logarithm is the sum of independent variables and thus one expects a central limit theorem to hold. In fact, using properties of the chi-square distribution, it is not hard to prove³⁴

$$(7.8) \quad \frac{\log(|\det A_n|) - \frac{1}{2} \log n! + \frac{1}{2} \log n}{\sqrt{\frac{1}{2} \log n}} \rightarrow N(0, 1)_{\mathbb{R}},$$

where $N(0, 1)_{\mathbb{R}}$ denotes the law of the real gaussian with mean 0 and variance 1.

A similar analysis (but with the real chi distribution replaced by a complex chi distribution) also works for complex gaussian matrices, in which ζ_{ij} remain jointly independent but now have the distribution of the complex gaussian $N(0, 1)_{\mathbb{C}}$ (or equivalently, the real and imaginary parts of ζ_{ij} are independent and have the distribution of $N(0, \frac{1}{2})_{\mathbb{R}}$). In that case, one has a slightly different law

$$(7.9) \quad \frac{\log(|\det A_n|) - \frac{1}{2} \log n! + \frac{1}{4} \log n}{\sqrt{\frac{1}{4} \log n}} \rightarrow N(0, 1)_{\mathbb{R}}.$$

We turn now to real iid matrices, in which the ζ_{ij} are jointly independent and real with mean zero and variance one. In [52], Girko stated that (7.8) holds for such random matrices under the additional assumption that the fourth moment of the atom variables is 3. Twenty years later, he claimed a much stronger result which replaced the above assumption by the assumption that the atom variables have bounded $(4 + \delta)$ -th moment [53]. However, there are several points which are not clear in these papers. Recently, Nguyen and the second author [82] gave a new proof for (7.8). Their approach also results in an estimate for the rate of convergence and is easily extended to handle to complex case.

The analysis of the above random determinants relies crucially on the fact that the rows of the matrix are jointly independent. This independence no longer holds

³⁴Here and in the sequel, \rightarrow denotes convergence in distribution.

for Hermitian random matrix models, which makes the analysis of determinants of Hermitian random matrices more challenging. Even showing that the determinant of a random symmetric Bernoulli matrix is non-zero (almost surely) was a long standing open question by Weiss in the 1980s and solved only five years ago [15]; for more recent developments see [108, Theorem 31], [16], [81], [120]. These results give good lower bound for the absolute value of the determinant or the least singular number, but do not reveal any information about the distribution.

Even in the GUE case, it is highly non-trivial to prove an analogue of the central limit theorem (7.9). Notice that the observation of Goodman does not apply due to the dependence between the rows and so it is not even clear why a central limit theorem must hold for the log-determinant. In [28], Delannay and Le Caer made use of the explicit distribution of GUE and GOE to prove the central limit theorem for these cases.

While it does not seem to be possible to express the log-determinant of GUE as a sum of independent random variables, in [112], the authors found a way to approximate the log-determinant as a sum of weakly dependent terms, based on analysing a tridiagonal form of GUE due to Trotter [118]³⁵. Using stochastic calculus and the martingale central limit theorem, we gave another proof (see [112]) for the central limit theorem for GUE and GOE:

THEOREM 7.15 (Central limit theorem for log-determinant of GUE and GOE). [28] *Let M_n be drawn from GUE. Then*

$$\frac{\log |\det(M_n)| - \frac{1}{2} \log n! + \frac{1}{4} \log n}{\sqrt{\frac{1}{2} \log n}} \rightarrow N(0, 1)_{\mathbb{R}}.$$

Similarly, if M_n is drawn from GOE rather than GUE, one has

$$\frac{\log |\det(M_n)| - \frac{1}{2} \log n! + \frac{1}{4} \log n}{\sqrt{\log n}} \rightarrow N(0, 1)_{\mathbb{R}}.$$

The next task is to extend beyond the GUE or GOE case. Our main tool for this is a four moment theorem for log-determinants of Wigner matrices, analogous to Theorem 5.2.

THEOREM 7.16 (Four moment theorem for determinant). *Let M_n, M'_n be Wigner matrices whose atom distributions have independent real and imaginary parts that match to fourth order off the diagonal and to second order on the diagonal, are bounded by $n^{O(c_0)}$ for some sufficiently small but fixed $c_0 > 0$, and are supported on at least three points. Let $G : \mathbb{R} \rightarrow \mathbb{R}$ obey the derivative estimates*

$$(7.10) \quad \left| \frac{d^j}{dx^j} G(x) \right| = O(n^{c_0})$$

for $0 \leq j \leq 5$. Let $z_0 = E + \sqrt{-1}\eta_0$ be a complex number with $|E| \leq 2 - \delta$ for some fixed $\delta > 0$. Then

$$\mathbf{E}G(\log |\det(M_n - \sqrt{n}z_0)|) - \mathbf{E}G(\log |\det(M'_n - \sqrt{n}z_0)|) = O(n^{-c})$$

for some fixed $c > 0$, adopting the convention that $G(-\infty) = 0$.

³⁵We would like to thank R. Killip for suggesting the use of Trotter's form.

The requirements that M_n, M'_n be supported on at least three points, and that E lie in the bulk region $|E| < 2 - \delta$ are artificial, due to the state of current literature on level repulsion estimates. It is likely that with further progress on those estimates that these hypotheses can be removed. The hypothesis that the atom distributions have independent real and imaginary parts is mostly for notational convenience and can also be removed with some additional effort.

By combining Theorem 7.16 with Theorem 7.15 we obtain

COROLLARY 7.17 (Central limit theorem for log-determinant of Wigner matrices). *Let M_n be a Wigner matrix whose atom distributions ζ_{ij} are independent of n , have real and imaginary parts that are independent and match GUE to fourth order, and obey Condition **C1** for some sufficiently large C_0 . Then*

$$\frac{\log |\det(M_n)| - \frac{1}{2} \log n! + \frac{1}{4} \log n}{\sqrt{\frac{1}{2} \log n}} \rightarrow N(0, 1)_{\mathbb{R}}.$$

If M_n matches GOE instead of GUE, then one instead has

$$\frac{\log |\det(M_n)| - \frac{1}{2} \log n! + \frac{1}{4} \log n}{\sqrt{\log n}} \rightarrow N(0, 1)_{\mathbb{R}}.$$

The deduction of this proposition from Theorem 7.16 and Theorem 7.15 is standard (closely analogous, for instance, to the proof the central limit theorem for individual eigenvalues) and is omitted. (Notice that in order for the atom variables of M_n match those of GUE to fourth order, these variables must have at least three points in their supports.)

7.7. Concentration of eigenvalues. We first discuss the case of the Gaussian Unitary Ensemble (GUE), which is the most well-understood case, as the joint distribution of the eigenvalues is given by a determinantal point process. Because of this, it is known that for any interval I , the random variable $N_I(W_n)$ in the GUE case obeys a law of the form

$$(7.11) \quad N_I(W_n) \equiv \sum_{i=1}^{\infty} \eta_i$$

where the $\eta_i = \eta_{i,n,I}$ are jointly independent indicator random variables (i.e. they take values in $\{0, 1\}$); see e.g. [1, Corollary 4.2.24]. The mean and variance of $N_I(W_n)$ can also be computed in the GUE case with a high degree of accuracy:

THEOREM 7.18 (Mean and variance for GUE). [57] *Let M_n be drawn from GUE, let $W_n := \frac{1}{\sqrt{n}} M_n$, and let $I = [-\infty, x]$ for some real number x (which may depend on n). Let $\varepsilon > 0$ be independent of n .*

(i) *(Bulk case) If $x \in [-2 + \varepsilon, 2 - \varepsilon]$, then*

$$\mathbf{E}N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O\left(\frac{\log n}{n}\right).$$

(ii) *(Edge case) If $x \in [-2, 2]$, then*

$$\mathbf{E}N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(1).$$

- (iii) (*Variance bound*) If one has $x \in [-2, 2 - \varepsilon]$ and $n^{2/3}(2 + x) \rightarrow \infty$ as $n \rightarrow \infty$, one has

$$\mathbf{Var}N_I(W_n) = \left(\frac{1}{2\pi^2} + o(1)\right) \log(n(2 + x)^{3/2}).$$

In particular, one has $\mathbf{Var}N_I(W_n) = O(\log n)$ in this regime.

By combining these estimates with a well-known inequality of Bennett [9] (see [114] for details) we obtain a concentration estimate for $N_I(W_n)$ in the GUE case:

COROLLARY 7.19 (Concentration for GUE). *Let M_n be drawn from GUE, let $W_n := \frac{1}{\sqrt{n}}M_n$, and let I be an interval. Then one has*

$$\mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll \exp(-cT)$$

for all $T \gg \log n$.

From the above corollary we see in particular that in the GUE case, one has

$$N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(\log^{1+o(1)} n)$$

with overwhelming probability for each fixed I , and an easy union bound argument (ranging over all intervals I in, say, $[-3, 3]$ whose endpoints are a multiple of n^{-100} (say)) then shows that this is also true uniformly in I as well.

Now we turn from the GUE case to more general Wigner ensembles. As already mentioned, there has been much interest in recent years in obtaining concentration results for $N_I(W_n)$ (and for closely related objects, such as the Stieltjes transform $s_{W_n}(z) := \frac{1}{n} \text{trace}(W_n - z)^{-1}$ of W_n) for short intervals I , due to the applicability of such results to establishing various universality properties of such matrices; see [37, 38, 39, 108, 103, 40, 43, 44]. The previous best result in this direction was by Erdős, Yau, and Yin [44] (see also [32] for a variant):

THEOREM 7.20. [44] *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Then, for any interval I , one has*

$$(7.12) \quad \mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll \exp(-cT^c)$$

for all $T \geq \log^{A \log \log n} n$, and some constant $A > 0$.

One can reformulate (7.12) equivalently as the assertion that

$$\mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll \exp(\log^{O(\log \log n)} n) \exp(-cT^c)$$

for all $T > 0$.

In particular, this theorem asserts that with overwhelming probability one has

$$N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(\log^{O(\log \log n)} n)$$

for all intervals I . The proof of the above theorem is somewhat lengthy, requiring a delicate analysis of the self-consistent equation of the Stieltjes transform of W_n .

Comparing this result with the previous results for the GUE case, we see that there is a loss of a double logarithm $\log \log n$ in the exponent. It has turned out

that using the swapping method one can remove this double logarithmic loss, at least under an additional vanishing moment assumption³⁶

THEOREM 7.21 (Improved concentration of eigenvalues). [114] *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Assume that M_n matches moments with GUE to third order off the diagonal (i.e. $\text{Re}\xi_{ij}, \text{Im}\xi_{ij}$ have variance 1/2 and third moment zero). Then, for any interval I , one has*

$$\mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll n^{O(1)} \exp(-cT^c)$$

for any $T > 0$.

This estimate is phrased for any T , but the bound only becomes non-trivial when $T \gg \log^C n$ for some sufficiently large C . In that regime, we see that this result removes the double-logarithmic factor from Theorem 7.20. In particular, this theorem implies that with overwhelming probability one has

$$N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(\log^{O(1)} n)$$

for all intervals I ; in particular, for any I , $N_I(W_n)$ has variance $O(\log^{O(1)} n)$.

REMARK 7.22. As we are assuming $\text{Re}(\xi_{ij})$ and $\text{Im}(\xi_{ij})$ to be independent, the moment matching condition simplifies to the constraints that $\mathbf{E}\text{Re}(\xi_{ij})^2 = \mathbf{E}\text{Im}(\xi_{ij})^2 = \frac{1}{2}$ and $\mathbf{E}\text{Re}(\xi_{ij})^3 = \mathbf{E}\text{Im}(\xi_{ij})^3 = 0$. However, it is possible to extend this theorem to the case when the real and imaginary parts of ξ_{ij} are not independent.

REMARK 7.23. The constant c in the bound in Theorem 7.21 is quite decent in several cases. For instance, if the atom variables of M_n are Bernoulli or have sub-gaussian tail, then we can set $c = 2/5 - o(1)$ by optimizing our arguments (details omitted). If we assume 4 matching moments rather than 3, then we can set $c = 1$, matching the bound in Corollary 7.19. It is an interesting question to determine the best value of c . The value of c in [43] is implicit and rather small.

The proof of the above theorem is different from that in [44] in that it only uses a relatively crude analysis of the self-consistent equation to obtain some preliminary bounds on the Stieltjes transform and on N_I (which were also essentially implicit in previous literature). Instead, the bulk of the argument relies on using the Lindeberg swapping strategy to deduce concentration of $N_I(W_n)$ in the non-GUE case from the concentration results in the GUE case provided by Corollary 7.19. In order to keep the error terms in this swapping under control, three matching moments³⁷. We need one less moment here because we are working at “mesoscopic” scales (in which the number of eigenvalues involved is much larger than 1) rather than at “microscopic” scales.

Very roughly speaking, the main idea of the argument is to show that high moments such as

$$\mathbf{E}|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy|^k$$

are quite stable (in a multiplicative sense) if one swaps (the real or imaginary part of) one of the entries of W_n (and its adjoint) with another random variable that

³⁶We thank M. Ledoux for a conversation leading to this study.

³⁷Compare with Theorem 5.2.

matches the moments of the original entry to third order. For technical reasons, however, we do not quite manipulate $N_I(W_n)$ directly, but instead work with a proxy for this quantity, namely a certain integral of the Stieltjes transform of W_n . As observed in [43], the Lindeberg swapping argument is quite simple to implement at the level of the Stieltjes transform (due to the simplicity of the resolvent identities, when compared against the rather complicated Taylor expansions of individual eigenvalues used in [108]). A similar application of the Lindeberg swapping argument to high moments was also recently performed in [72] to control coefficients of the resolvent in arbitrary directions.

As a corollary, we obtain the following rigidity of eigenvalues result, improving upon (2.13) when one has a matching moment hypothesis:

COROLLARY 7.24 (Concentration of eigenvalues). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Assume that M_n matches moments with GUE to three order off the diagonal and second order on the diagonal. Then for any i in the bulk*

$$\mathbf{P}(|\lambda_i(W_n) - \gamma_i| \geq T/n) \ll n^{O(1)} \exp(-cT^c)$$

for any $T > 0$, where the classical location $\gamma_i \in [-2, 2]$ is defined by the formula

$$\int_{-2}^{\gamma_i} \rho_{\text{sc}}(y) dy = \frac{i}{n}.$$

This corollary improves [44, Theorem 2.2] as it allows T to be as small as $\log^{O(1)} n$, instead of $\log^{O(\log \log n)} n$, under the extra third moment assumption. In particular, in the Bernoulli case, this shows that the variance of the bulk eigenvalues is of order $\log^{O(1)} n/n$. We believe that this is sharp, up to the hidden constant in $O(1)$.

This corollary also significantly improves [108, Theorem 29]. (As a matter of fact, the original proof of this theorem has a gap in it; see Appendix A for a further discussion.) One can have analogous results for the edge case, under the four moment assumption; see [114] for details.

8. Open questions

While the universality of many spectral statistics of Wigner matrices have now been established, there are still several open questions remaining. Some of these have already been raised in earlier sections; we collect some further such questions in this section.

In one direction, one can continue generalising the class of matrices for which the universality result holds, for instance by lowering the exponent C_0 in Condition **C1**, and allowing the entries ξ_{ij} of the Wigner matrix to have different variances σ_{ij}^2 , or for the Wigner matrix to be quite sparse. For recent work in these directions, see [32], [33], [43], [44]. With regards to the different variances case, one key assumption that is still needed for existing arguments to work is a *spectral gap hypothesis*, namely that the matrix of variances $(\sigma_{ij}^2)_{1 \leq i, j \leq n}$ has a significant gap between its largest eigenvalue and its second largest one; in addition, for the most complete results one also needs the variances to be bounded away from zero. This omits some interesting classes of Wigner-type matrices, such as those with large

blocks of zeroes. However, the spectral statistics of $p + n \times p + n$ matrices of the form

$$\begin{pmatrix} 0 & M \\ M^* & 0 \end{pmatrix}$$

for rectangular ($p \times n$) matrices with iid matrices M are well understood, as the problem is equivalent to that of understanding the singular values of M (or the eigenvalues of the covariance matrix MM^*); in particular, analogues of the key tools discussed here (such as the four moment theorem, the local semicircle law, and heat flow methods) are known [109], [41]; this suggests that other block-type variants of Wigner matrices could be analysed by these methods. A related problem would be to understand the spectral properties of various self-adjoint polynomial combinations of random matrices, e.g. the commutator $AB - BA$ of two Wigner matrices A, B . The global coarse-scale nature of the spectrum for such matrices can be analysed by the tools of free probability [122], but there are still very few rigorous results for the local theory.

Another direction of generalisation is to consider generalised Wigner matrices whose entries have non-zero mean, or equivalently to consider the spectral properties of a random matrix $M_n + D_n$ that is the sum of an ordinary Wigner matrix M_n and a deterministic Hermitian matrix D_n . Large portions of the theory seem amenable to extension in this direction, although the global and local semicircular law would need to be replaced by a more complicated variant (in particular, the semicircular distribution ρ_{sc} should be replaced by the free convolution of ρ_{sc} with the empirical spectral distribution of D_n , see [122, 121]).

In yet another direction, one could consider non-Hermitian analogues of these problems, for instance by considering the statistics of eigenvalues of iid random matrices (in which the entries are not constrained to be Hermitian, but are instead independent and identically distributed). The analogue of the semicircular law in this setting is the *circular law*, which has been analysed intensively in recent years (see [102] for a survey). There are in fact a number of close connections between the Hermitian and non-Hermitian ensembles, and so it is likely that the progress in the former can be applied to some extent to the latter.

Another natural question to ask is to see if the universality theory for Wigner ensembles can be unified in some way with the older, but very extensively developed, universality theory for invariant ensembles (as covered for instance in [20]). Significant progress in this direction has recently been achieved in [12], in which heat flow methods are adapted to show that the local spectral statistics of β -ensembles are asymptotically independent of the choice of potential function (assuming some analyticity conditions on the potential). This reduces the problem to the gaussian case when the potential is quadratic, which can be handled by existing methods, similarly to how the methods discussed here reduce the statistics of general Wigner matrices to those of invariant ensembles such as GUE or GOE. Note though that these techniques do not provide an independent explanation as to why these invariant ensembles have the limiting statistics they do (e.g. governed by the sine determinantal process in the bulk, and the Airy determinantal process in the edge, in the case of GUE); for that, one still needs to rely on the theory of determinantal processes.

Returning now to Wigner matrices, one of the major limitations of the methods discussed here is the heavy reliance on the hypothesis that the (upper-triangular)

entries are jointly independent; even weak coupling between entries makes many of the existing methods (such as the swapping technique used in the Four Moment Theorem, or the use of identities such as (2.9)) break down. A good test case would be the asymptotic statistics of the adjacency matrices of random regular graphs, where the fixed degree d is a constant multiple of n , such as $n/2$. This is essentially equivalent to a Wigner matrix model (such as the real symmetric Bernoulli matrix ensemble) in which the row and column sums have been constrained to be zero. For this model, the global semicircular law and eigenvector delocalisation has recently been established for such matrices; see [24], [117].

Recall that the Central limit theorem for the log-determinant of non-Hermitian matrices requires only two moment matching. However, in the Hermitian case, Corollary 7.17 requires four matching moments. We believe that this requirement can be weakened. For instance, the central limit theorem must hold for random Bernoulli matrices.

Another interesting problem is to determine the distribution of bulk eigenvalues of a random Bernoulli matrix. This matrix has only three matching moments, but perhaps a central limit theorem like Theorem 7.1 also holds here. We have proved [114] that the variance of any bulk eigenvalue is $\log^{O(1)} n/n$. A good first step would be to determine the right value of the hidden constant in $O(1)$.

Appendix A. Some errata

In this appendix we report an issue with the *Three Moment Theorem*, which first appeared as the second conclusion of [108, Theorem 15]. We extract this theorem from that paper for reference:

THEOREM A.1 (Three Moment Theorem). *There is a small positive constant c_0 such that for every $0 < \varepsilon < 1$ and $k \geq 1$ the following holds. Let $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ and $M'_n = (\zeta'_{ij})_{1 \leq i, j \leq n}$ be two random matrices satisfying **C0**. Assume furthermore that for any $1 \leq i < j \leq n$, ζ_{ij} and ζ'_{ij} match to order 3. Set $A_n := \sqrt{n}M_n$ and $A'_n := \sqrt{n}M'_n$, and let $G : \mathbb{R}^k \rightarrow \mathbb{R}$ be a smooth function obeying the derivative bounds*

$$(A.1) \quad |\nabla^j G(x)| \leq n^{-Cj c_0}$$

for all $0 \leq j \leq 5$ and $x \in \mathbb{R}^k$, and some sufficiently large absolute constant $C > 0$. Then for any $\varepsilon n \leq i_1 < i_2 < \dots < i_k \leq (1 - \varepsilon)n$, and for n sufficiently large depending on ε, k (and the constants C, C' in the definition of Condition **C0**) we have

$$(A.2) \quad |\mathbf{E}(G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))) - \mathbf{E}(G(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n)))| \leq n^{-c_0}.$$

Unfortunately, the proof given of the Three Moment Theorem in [108, §3.3] is not correct, although as we shall see shortly, the theorem can be proven by other means, and the proof itself can be repaired in some other instances. The closely related Four Moment Theorem, which is of more importance in most applications, is not affected by this problem.

The issue affects some subsequent papers [103], [109], [106], [65], [125] where the Three Moment Theorem (or a variant thereof) was used, and we will discuss the alterations needed to correct those papers below.

A.1. The issue. In the proof of the Four Moment Theorem in [108, §3.3], the analogue of the function G that appears in Theorem A.1 is replaced with a truncated variant

$$(A.3) \quad \tilde{G}(\lambda_{i_1}, \dots, \lambda_{i_k}, Q_{i_1}, \dots, Q_{i_k}) := G(\lambda_{i_1}, \dots, \lambda_{i_k}) \prod_{j=1}^k \eta(Q_{i_j})$$

(see [108, §3.3] for notation). This truncated function \tilde{G} then obeys the derivative estimates

$$|\nabla^j \tilde{G}| \ll n^{c_0}$$

for $j = 0, \dots, 5$. These bounds are sufficient to establish the Four Moment Theorem (the first part of [108, Theorem 15]), by invoking the first part of [108, Proposition 46]. However, for the purposes of establishing the Three Moment Theorem (the second part of [108, Theorem 15]), these bounds are not sufficient to be able to invoke the second part of [108, Proposition 46], and so the proof breaks down in this case.

A similar issue also affects the Three Moment Theorem in the bulk of the spectrum for covariance matrices ([109, Theorem 6]), the Three Moment Theorem at the edge of the spectrum of Wigner matrices [103, Theorem 1.5], and the Three Moment Theorem at the edge of the spectrum for covariance matrices [125, Theorem 1.5].

A.2. First fix: strengthen the gap property. The three moment theorem is most useful when applied to eigenvalues at the edge. In this case, we can reprove this theorem by modifying the argument in [108] as follows.

The proof of the Four Moment Theorem relies on the *gap property* (Theorem 5.4). Suppose temporarily that one could improve this property to the bound

$$(A.4) \quad |\lambda_{i+1}(A_n) - \lambda_i(A_n)| \geq n^{C_{c_0}}$$

with high probability. This is not actually possible in the bulk of the spectrum, where the mean eigenvalue spacing is comparable to 1 (by the Wigner semicircle law), but is certainly plausible at the edge of the spectrum (where the mean eigenvalue spacing is $n^{1/3}$). Assuming this improved gap property, one can fix the original proof of the Three Moment Theorem in [108]. To do this, one replaces the quantities Q_i introduced in [108, §3.2] with the modified variant $\tilde{Q}_i := Q_i/n^{C_{c_0}/10}$, and replaces the function \tilde{G} in (A.3) with the variant

$$\tilde{G}(\lambda_{i_1}, \dots, \lambda_{i_k}, \tilde{Q}_{i_1}, \dots, \tilde{Q}_{i_k}) := G(\lambda_{i_1}, \dots, \lambda_{i_k}) \prod_{j=1}^k \eta(\tilde{Q}_{i_j}/n^{C_{c_0}/100}).$$

The function \tilde{G} then obeys the improved bounds

$$|\nabla^j \tilde{G}| \ll n^{-C_{c_0}j/100}.$$

As such, it is possible to adapt the arguments in [108] (replacing Q_i with \tilde{Q}_i , and also making some other minor changes to constants such as C) to recover the Three Moment Theorem.

It remains to establish (A.4) with high probability at the edge of the spectrum. In the case when A_n comes from the GUE ensemble and i is at the edge of the spectrum (thus $i = O(1)$ or $i = n - O(1)$), one can obtain (A.4) with high probability for sufficiently small c_0 by a careful inspection of the analysis used to establish the

Tracy-Widom law for GUE [116] (or alternatively, one can use the more general analysis in [65]). By using the *truncated* Three Moment Theorem at the edge (i.e. [103, Proposition 6.1], after replacing Q_{i_j} with \tilde{Q}_{i_j} as indicated above), we can then deduce (A.4) with high probability for any Wigner matrix (obeying Condition C1 for a sufficiently high moment) that matches GUE to third order. Using this, one can recover the Three Moment Theorem at the edge for any pair M_n, M'_n of matrices that match GUE to third order (i.e. have vanishing third moment).

THEOREM A.2 (Three Moment Theorem). *There is a small positive constant c_0 such that for every $0 < \varepsilon < 1$ and $k \geq 1$ the following holds. Let $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$ be a random matrix satisfying C0 where $\zeta_{ij}, i \neq j$ have vanishing third moment. Set $A_n := \sqrt{n}M_n$ and $A'_n := \sqrt{n}M'_n$, where A'_n is GUE, and let $G : \mathbb{R}^k \rightarrow \mathbb{R}$ be a smooth function obeying the derivative bounds*

$$(A.5) \quad |\nabla^j G(x)| \leq n^{-Cj c_0}$$

for all $0 \leq j \leq 5$ and $x \in \mathbb{R}^k$, and some sufficiently large absolute constant $C > 0$. Then for any $i_1 < i_2 \cdots < i_k = O(1)$, and for n sufficiently large depending on ε, k (and the constants C, C' in the definition of Condition C0) we have

$$(A.6) \quad |\mathbf{E}(G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))) - \mathbf{E}(G(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n)))| \leq n^{-c_0}.$$

The same statement holds at the other edge of the spectrum.

This method allows one to prove one of the main applications of the three moment theorem, [103, Theorem 1.13], which establishes the universality of spectral statistics at the edge of a Wigner matrix with vanishing third moment.

In Theorem A.2, the vanishing third moment assumption was needed only to guarantee (A.4), since GUE satisfies this bound. In a recent paper [65] Johansson considered gauss divisible matrices. While not stated explicitly, his analysis seems to show that the convergence rate of the density of states of gauss divisible matrices at the edge is polynomial, and this would imply that (A.4) also holds for this larger class of matrices. As a consequence, one would be able to remove the vanishing third moment assumption and Theorem A.2 and also obtain [65, Theorem 1.4].

In a similar manner, one can recover the covariance matrix universality in [125] (after one verifies the required polynomial convergence rate of the density of states at the edge for Wishart ensembles (or for gauss divisible covariance matrices), which can be extracted from the analysis in [8]).

A.3. Second fix: use the Knowles-Yin Two Moment Theorem. Another way to handle the above applications is to use the very recent Two Moment Theorem at the edge of Knowles and Yin [71, Theorem 1.1]. The conclusion of this theorem is stronger than that of the Three Moment Theorem for the edge, for it only requires two matching moments rather than three (and it can also control the eigenvectors as well as the eigenvalues, cf. [113]). However, its assumption is more restricted, as the hypotheses on G are stronger than (A.5), being adapted to the scale $n^{1/3}$ of the mean eigenvalue spacing at the edge, rather than to the scale $n^{C c_0}$. As such, the Two Moment Theorem at the edge can be used as a substitute for the Three Moment Theorem to establish universality for the distribution of k eigenvalues at the edge for any fixed $k = O(1)$, as was done in [109] and [65]. This can also recover the analogous universality results for covariance matrices in [125], if one has the covariance matrix version of the Knowles-Yin Two Moment Theorem

in [71] in the covariance matrix case. This was achieved very recently in [89] in the rectangular case and is likely to also be true for covariance matrices arising from square or nearly-square matrices also.

A.4. Third fix: use of eigenvalue rigidity. This fix applies to the original Three Moment Theorem in the bulk for Wigner matrices (Theorem A.1). One can reprove this theorem using the powerful *eigenvalue rigidity* result (2.13). From this bound and (A.1) (with sufficiently large C) we conclude that

$$G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n)) = G(\lambda_{i_1}^{\text{cl}}(A_n), \dots, \lambda_{i_k}^{\text{cl}}(A_n)) + o(n^{-c_0})$$

with overwhelming probability, and similarly for A'_n . Since A_n and A'_n have the same classical locations, the three moment theorem follows.

This fix can be used to resolve all the places where the Three Moment Theorem from [108] are invoked. Of course, in all such cases stronger results can be deduced directly from the subsequent results of [44].

In the case when the third moment is zero, one can also use Corollary 7.24 instead of (2.13). The point is that the proof of this corollary (which relies on the swapping method) is much simpler than that of (2.13).

This fix also repairs the Three Moment Theorem in the bulk for covariance matrices [109, Theorem 6], provided one has an eigenvalue rigidity result for covariance matrices. This was established very recently in [88] in the rectangular case, and it is likely that the results can also be extended to the square case.

REMARK A.3. In [109], it was observed that the exponential decay condition **C0** in the Four Moment Theorem could be relaxed to a finite moment condition **C1** for a sufficiently large moment exponent. If one could obtain a similar relaxation for the eigenvalue rigidity result in [44], then the above argument could then be used to recover the Three Moment Theorem in this case also. In fact, one can achieve this using the Four Moment Theorem. If M_n is a matrix obeying **C1**, then it can be matched to fourth order (or to approximate fourth order, as in [36]) to a matrix M'_n obeying **C0**. The latter matrix obeys (2.13) with overwhelming probability; applying the Four Moment Theorem (cf. the proof of [108, Theorem 32]) we conclude that the former matrix also obeys (2.13) with high probability, and one can then adapt the preceding arguments. It is also likely that one can achieve similar results by direct modification of the arguments in [44].

At the edge of the spectrum, the eigenvalue rigidity property is adapted to a broader scale than that used in the Three Moment Theorem; with respect to the fine-scale normalisation A_n , the latter is adapted to the scale $n^{1/3}$ at the edge whilst the former is adapted to the scale $n^{C_{c_0}}$. As such, the eigenvalue rigidity property cannot be directly used as a fix for applications at the edge.

A.5. Corrections to specific papers. In this section we record specific statements in a number of papers that need repairing, and which of the above fixes are applicable to these papers.

- The second part of [108, Theorem 15] (i.e. the Three Moment Theorem) is true, but the proof requires a result that occurred after the publication of this paper, namely the eigenvalue rigidity result in [44]). References to this part of the theorem in [108, §3.3] should be deleted.
- Similarly for [108, Theorem 32]. This theorem is true, but its proof requires using the rigidity result in [44] or Corollary 7.24.

- Similarly for the asymptotic for the determinant in [108, Theorem 34]. In fact, by using the results in [44], one can even drop the third moment assumption.
- The second part of [109, Theorem 6] (i.e. the Three Moment Theorem) can also be repaired by similar means, using the eigenvalue rigidity results from [88] in the rectangular case (which can also presumably be extended to the square case). If we assume that the third moment vanishes (which is the case in many application), then it suffices to obtain an analogue of Corollary 7.24 and the proof of this would be simpler.
- The eigenvalue localisation result in [106, Theorem 1.4] now requires M_n to match GUE to fourth order rather than third order. Again, the moment matching hypotheses can be dropped entirely if one is willing to use the results in [44].
- The second part of [103, Theorem 1.13] should add the additional hypothesis that M_n, M'_n match GUE to third order (or at least obey the improved gap condition (A.4) with high probability). The proof of this theorem then needs to be modified as per Section A.2.
- In the proof of [65, Theorem 1.4], one needs to replace the invocation of the Three Moment Theorem with the more complicated argument indicated in Section A.2. Alternatively, one can use here the Two Moment Theorem of Knowles and Yin [71].
- The second part of [125, Theorem 1.5] should add the additional hypothesis that M_n, M'_n match the Wishart ensemble to third order (or at least obey the improved gap condition (A.4) with high probability). The proof of this theorem then needs to be modified as per Section A.2. Alternatively, one can use the Two Moment Theorem in [89], in the rectangular case at least.

References

1. G. Anderson, A. Guionnet and O. Zeitouni, An introduction to random matrices, Cambridge Studies in Advanced Mathematics, 118. Cambridge University Press, Cambridge, 2010.
2. Z. D. Bai, Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices, *Ann. Probab.* **21** (1993), no. 2, 625-648.
3. Z. D. Bai, Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.* **21** (1993), no. 2, 649-672
4. J. Baik, P. Deift, and K. Johansson, On the distribution of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12** (1999), 1119–1178.
5. D. Bakry, M. Émery, *Diffusions hypercontractives*, in: Séminaire de probabilités, XIX, 1983/84, 1123 Lecture Notes in Mathematics, Springer, Berlin, 1985, 177-206.
6. Z. D. Bai and J. Silverstein, Spectral analysis of large dimensional random matrices, Mathematics Monograph Series **2**, Science Press, Beijing 2006.
7. Z.D. Bai, Y.Q. Yin, Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix, *Ann. Probab.* **16** (1988), no. 4, 1729-1741.
8. G. Ben Arous and S. Péché, Universality of local eigenvalue statistics for some sample covariance matrices, *Comm. Pure Appl. Math.* **58** (2005), no. 10, 1316–1357.
9. G. Bennett, *Probability Inequalities for the Sum of Independent Random Variables*, Journal of the American Statistical Association **57** (1962), 33-45.
10. A. Berry, The Accuracy of the Gaussian Approximation to the Sum of Independent Variates, *Trans. Amer. Math. Soc.* **49** (1941), 122-136.
11. P. Bleher and A. Its, Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model, *Ann. of Math. (2)* **150** (1999), no. 1, 185–266.
12. P. Bourgade, L. Erdős, H. T. Yau, Universality of General β -Ensembles, *arXiv:1104.2272*

13. E. Brézin, S. Hikami, S., An extension of level-spacing universality, *cond-mat/9702213*.
14. O. Costin and J. Lebowitz, Gaussian fluctuations in random matrices, *Phys. Rev. Lett.* **75** (1) (1995) 69–72.
15. K. Costello, T. Tao and V. Vu, Random symmetric matrices are almost surely singular, *Duke Math. J.* **135** (2006), 395–413.
16. K. Costello, Bilinear and quadratic variants on the Littlewood-Offord problem, *preprint*.
17. S. Chatterjee, A generalization of the Lindenber principle, *Ann. Probab.* **34** (2006), no. 6, 2061–2076.
18. S. Dallaporta, V. Vu, A Note on the Central Limit Theorem for the Eigenvalue Counting Function of Wigner Matrices, *to appear Elect Comm. Prob.*
19. P. Deift, T. Kriecherbauer, K.T.-R. McLaughlin, S. Venakides and X. Zhou, Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory, *Comm. Pure Appl. Math.* **52** (1999), no. 11, 1335–1425.
20. P. Deift, Orthogonal polynomials and random matrices: a Riemann-Hilbert approach. Courant Lecture Notes in Mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.
21. P. Deift and D. Gioev, Random matrix theory: Invariant ensembles and Universality, Courant Lecture Notes in Mathematics, 18. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2009.
22. P. Deift, Universality for mathematical and physical systems. *International Congress of Mathematicians* Vol. I, 125–152, Eur. Math. Soc., Zürich, 2007.
23. A. Dembo, On random determinants, *Quart. Appl. Math.* **47** (1989), no. 2, 185–195.
24. I. Dumitriu, S. Pal, Sparse regular random graphs: spectral density and eigenvectors, *arXiv:0910.5306*
25. H. Doering, P. Eichelsbacher, Moderate deviations for the eigenvalue counting function of Wigner matrices, *arXiv:1104.0221*.
26. F. Dyson, The Threefold Way. Algebraic Structure of Symmetry Groups and Ensembles in Quantum Mechanics, *J. Math. Phys.* **3** (1962), 1199–1215.
27. F. Dyson, Correlations between eigenvalues of a random matrix, *Comm. Math. Phys.* **19** (1970), 235–250.
28. R. Delannay, G. Le Caer, Distribution of the determinant of a random real-symmetric matrix from the Gaussian orthogonal ensemble, *Phys. Rev. E* **62**, 15261536 (2000).
29. J. Dixon, *How good is Hadamard's inequality for determinants?*, *Canad. Math. Bull.* **27** (1984), no. 3, 260–264.
30. A. Edelman, Eigenvalues and condition numbers of random matrices, *SIAM J. Matrix Anal. Appl.* **9** (1988), 543–560.
31. L. Erdős, Universality of Wigner random matrices: a Survey of Recent Results, *arXiv:1004.0861*.
32. L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral Statistics of Erdős-Rényi Graphs I: Local Semicircle Law, *arXiv:1103.1919*.
33. L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral statistics of Erdős-Rényi graphs II: eigenvalue spacing and the extreme eigenvalues, *arXiv:1103.3869*.
34. L. Erdős, J. Ramirez, B. Schlein and H.-T. Yau, Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation, *arXiv:0905.2089*.
35. L. Erdős, S. Peche, J. Ramirez, B. Schlein and H.-T. Yau, Bulk universality for Wigner matrices, *arXiv:0905.4176*
36. L. Erdős, J. Ramirez, B. Schlein, T. Tao, V. Vu, and H.-T. Yau, Bulk universality for Wigner hermitian matrices with subexponential decay, *arxiv:0906.4400*, *To appear in Math. Research Letters*.
37. L. Erdős, B. Schlein and H.-T. Yau, Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37** (2009), 815–852 .
38. L. Erdős, B. Schlein and H.-T. Yau, Local semicircle law and complete delocalization for Wigner random matrices, *Comm. Math. Phys.* **287** (2009), no. 2, 641655.
39. L. Erdős, B. Schlein and H.-T. Yau, Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Notices* **2010** (2010), 436–479.
40. L. Erdős, B. Schlein and H.-T. Yau, Universality of Random Matrices and Local Relaxation Flow, *arXiv:0907.5605*

41. L. Erdős, B. Schlein, H.-T. Yau and J. Yin, The local relaxation flow approach to universality of the local statistics for random matrices. *arXiv:0911.3687*
42. L. Erdős and H.-T. Yau, A comment on the Wigner-Dyson-Mehta bulk universality conjecture for Wigner matrices. *arXiv:1201.5619*
43. L. Erdős, H.-T. Yau, and J. Yin, Bulk universality for generalized Wigner matrices. *arXiv:1001.3453*
44. L. Erdős, H.-T. Yau, and J. Yin, Rigidity of Eigenvalues of Generalized Wigner Matrices. *arXiv:1007.4652*
45. C.-G. Esseen, On the Liapunoff limit of error in the theory of probability, *Arkiv för matematik, astronomi och fysik* A28 (1942), 1-19.
46. O. Feldheim and S. Sodin, A universality result for the smallest eigenvalues of certain sample covariance matrices, *Geom. Funct. Anal.* 20 (2010), no. 1, 88–123.
47. P. J. Forrester, E. M. Rains, Interrelationships between orthogonal, unitary and symplectic matrix ensembles, *Random matrix models and their applications*, Math. Sci. Res. Inst. Publ., vol. 40, Cambridge Univ. Press, Cambridge, 2001, pp. 171-207.
48. P. J. Forrester, Log-gases and random matrices. London Mathematical Society Monographs Series, 34. Princeton University Press, Princeton, NJ, 2010.
49. G. E. Forsythe and J. W. Tukey, The extent of n random unit vectors, *Bull. Amer. Math. Soc.* 58 (1952), 502.
50. Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices, *Combinatorica* 1 (1981), no. 3, 233–241.
51. J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, *J. Mathematical Phys.* 6 (1965), 440-449.
52. V. L. Girko, The central limit theorem for random determinants (Russian), translation in *Theory Probab. Appl.* 24 (1979), no. 4, 729–740.
53. V. L. Girko, A Refinement of the central limit theorem for random determinants (Russian), translation in *Theory Probab. Appl.* 42 (1998), no. 1, 121–129.
54. V. L. Girko, Theory of random determinants, Kluwer Academic Publishers, 1990.
55. N. R. Goodman, Distribution of the determinant of a complex Wishart distributed matrix, *Annals of Statistics* 34 (1963), 178–180.
56. A. Guionnet, Grandes matrices aléatoires et théorèmes d’universalité, *Séminaire BOURBAKI*. Avril 2010. 62ème année, 2009-2010, no 1019.
57. J. Gustavsson, Gaussian fluctuations of eigenvalues in the GUE, *Ann. Inst. H. Poincaré Probab. Statist.* 41 (2005), no. 2, 151–178.
58. The Oxford handbook of random matrix theory, edited by Akeman, Baik and Di Fancesco, Oxford 2011.
59. Harish-Chandra, Differential operators on a semisimple Lie algebra, *Amer. J. Math.*, 79, (1957) 87–120.
60. J. Ben Hough, M. Krishnapur, Y. Peres, B. Virág, Determinantal processes and independence. *Probab. Surv.* 3 (2006), 206-229.
61. T. Jiang. How Many Entries of A Typical Orthogonal Matrix Can Be Approximated By Independent Normals?, *Ann. Probab.* 34 (2006), 1497–1529.
62. T. Jiang, The Entries of Haar-invariant Matrices from the Classical Compact Groups, *Journal of Theoretical Probability*, 23 (2010), 1227–1243.
63. M. Jimbo, T. Miwa, Tetsuji, Y. Mori and M. Sato, Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent, *Phys. D 1.* (1980), no. 1, 80–158.
64. K. Johansson, Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices, *Comm. Math. Phys.* 215 (2001), no. 3, 683–705.
65. K. Johansson, Universality for certain Hermitian Wigner matrices under weak moment conditions, *preprint*. re
66. J. Kahn, J. Komlós and E. Szemerédi, On the probability that a random ± 1 matrix is singular, *J. Amer. Math. Soc.* 8 (1995), 223–240.
67. N. Katz, P. Sarnak, Random matrices, Frobenius eigenvalues, and monodromy. American Mathematical Society Colloquium Publications, 45. American Mathematical Society, Providence, RI, 1999.
68. J. P. Keating, N. C. Snaith, Random matrix theory and $\zeta(1/2 + it)$, *Comm. Math. Phys.* 214 (2000), no. 1, 57-89.

69. O. Khorunzhiy, High Moments of Large Wigner Random Matrices and Asymptotic Properties of the Spectral Norm, preprint.
70. R. Killip, Gaussian fluctuations for β ensembles, *Int. Math. Res. Not.* 2008, no. 8, Art. ID rnn007, 19 pp.
71. A. Knowles, J. Yin, Eigenvector Distribution of Wigner Matrices, *arXiv:1102.0057*
72. A. Knowles, J. Yin, The isotropic semicircle law and deformation of Wigner matrices, *arXiv:1110.6449*.
73. J. Komlós, On the determinant of $(0, 1)$ matrices, *Studia Sci. Math. Hungar.* **2** (1967), 7–21.
74. J. Komlós, On the determinant of random matrices, *Studia Sci. Math. Hungar.* **3** (1968), 387–399.
75. I. V. Krasovsky, Correlations of the characteristic polynomials in the Gaussian unitary ensemble or a singular Hankel determinant, *Duke Math. J.* **139** (2007), no. 3, 581–619.
76. M. Ledoux, The concentration of measure phenomenon, *Mathematical Surveys and Monographs*, 89. American Mathematical Society, Providence, RI, 2001.
77. J. W. Lindeberg, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, *Math. Z.* **15** (1922), 211–225.
78. A. Maltsev, B. Schlein, Average Density of States for Hermitian Wigner Matrices, *arXiv:1011.5594*.
79. A. Maltsev, B. Schlein, A Wegner estimate for Wigner matrices, *arXiv:1103.1473*.
80. M.L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels*, Academic Press, New York, NY, 1967.
81. H. Nguyen, On the least singular value of random symmetric matrices, *arXiv:1102.1476*.
82. H. Nguyen and V. Vu, Random matrix: Law of the determinant, *submitted*.
83. H. Nyquist, S. O. Rice and J. Riordan, The distribution of random determinants, *Quart. Appl. Math.* **12** (1954), 97–104.
84. S. O’Rourke, Gaussian fluctuations of eigenvalues in Wigner random matrices, *J. Stat. Phys.* **138** (2010), no. 6, 1045–1066.
85. L. Pastur, On the spectrum of random matrices, *Teoret. Mat.Fiz.* **10**, 102–112 (1973).
86. L. Pastur and M. Shcherbina, Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles, *J. Statist. Phys.* **86** (1997), no. 1–2, 109–147.
87. S. Peche and A. Soshnikov, Wigner Random Matrices with Non-symmetrically Distributed Entries *J. Stat. Phys.* , vol. 129, No. 5/6, 857–884, (2007).
88. N. Pillai, J. Yin, Universality of Covariance Matrices, *arXiv:1110.2501*
89. N. Pillai, J. Yin, Edge Universality of Correlation Matrices, *arXiv:1112.2381*
90. A. Prékopa, On random determinants I, *Studia Sci. Math. Hungar.* **2** (1967), 125–132.
91. M. Rudelson, R. Vershynin, The Littlewood-Offord Problem and invertibility of random matrices, *Advances in Mathematics* **218** (2008), 600–633.
92. A. Rouault, Asymptotic behavior of random determinants in the Laguerre, Gram and Jacobi ensembles, *Latin American Journal of Probability and Mathematical Statistics (ALEA)*, **3** (2007) 181–230.
93. G. Szekeres and P. Turán, On an extremal problem in the theory of determinants, *Math. Naturwiss. Am. Ungar. Akad. Wiss.* **56** (1937), 796–806.
94. M. Rudelson, R. Vershynin, The least singular value of a random square matrix is $O(n^{-1/2})$, *C. R. Math. Acad. Sci. Paris* **346** (2008), no. 15–16, 893–896.
95. A. Ruzmaikina, Universality of the edge distribution of eigenvalues of Wigner random matrices with polynomially decaying distributions of entries, *Comm. Math. Phys.* **261** (2006), no. 2, 277–296.
96. B. Schlein, Spectral Properties of Wigner Matrices, *Proceedings of the Conference QMath 11*, Hradec Kralove, September 2010.
97. Y. Sinai, A. Soshnikov, Central limit theorem for traces of large symmetric matrices with independent matrix elements, *Bol. Soc. Brazil. Mat.* **29**, 1–24 (1998).
98. Y. Sinai, A. Soshnikov, A refinement of Wigners semicircle law in a neighborhood of the spectrum edge for random symmetric matrices, *Func. Anal. Appl.* **32**, 114–131 (1998).
99. A. Soshnikov, Universality at the edge of the spectrum in Wigner random matrices, *Comm. Math. Phys.* **207** (1999), no. 3, 697–733.
100. A. Soshnikov, Gaussian limit for determinantal random point fields, *Ann. Probab.* **30** (1) (2002) 171–187.

101. G. Szegő, On certain Hermitian forms associated with the Fourier series of a positive function, *Comm. Sem. Math. Univ. Lund* 1952(1952), tome supplementaire, 228–238.
102. T. Tao and V. Vu, From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices, *Bull. Amer. Math. Soc. (N.S.)* 46 (2009), no. 3, 377–396.
103. T. Tao and V. Vu, Random matrices: universality of local eigenvalue statistics up to the edge, *Comm. Math. Phys.*, **298** (2010), 549–572.
104. T. Tao and V. Vu, Random matrices: universality of ESDs and the circular law. With an appendix by Manjunath Krishnapur, *Ann. Probab.* 38 (2010), no. 5, 2023–2065.
105. T. Tao and V. Vu, Random matrices: the distribution of the smallest singular values, *Geom. Funct. Anal.* 20 (2010), no. 1, 260–297.
106. T. Tao and V. Vu, Random matrices: Localization of the eigenvalues and the necessity of four moments, *to appear in Acta Math. Vietnamica*.
107. T. Tao and V. Vu, Random matrices: Universal properties of eigenvectors, [arXiv:arXiv:1103.2801](https://arxiv.org/abs/1103.2801).
108. T. Tao and V. Vu, Random matrices: Universality of the local eigenvalue statistics, *Acta Mathematica* 206 (2011), 127–204.
109. T. Tao, V. Vu, Random covariance matrices: universality of local statistics of eigenvalues, *to appear in Annals of Probability*.
110. T. Tao and V. Vu, The Wigner-Dyson-Mehta bulk universality conjecture for Wigner matrices, *Electronic Journal of Probability*, vol 16 (2011), 2104–2121.
111. T. Tao and V. Vu, On random ± 1 matrices: singularity and determinant, *Random Structures Algorithms* **28** (2006), 1–23.
112. T. Tao and V. Vu, A central limit theorem for the determinant of a Wigner matrix, *preprint* <http://arxiv.org/pdf/1111.6300.pdf>.
113. T. Tao and V. Vu, Random matrices: Universal properties of eigenvectors, *to appear in Random Matrices: Theory and Applications*.
114. T. Tao and V. Vu, Random matrices: Sharp concentration of eigenvalues, *preprint* <http://arxiv.org/pdf/1201.4789.pdf>.
115. C. Tracy and H. Widom, On orthogonal and symplectic matrix ensembles, *Commun. Math. Phys.* 177 (1996) 727–754.
116. C. Tracy and H. Widom, Distribution functions for largest eigenvalues and their applications, Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002), 587–596.
117. L. Tran, V. Vu, K. Wang, Sparse random graphs: Eigenvalues and Eigenvectors, *to appear in Random Structures and Algorithms*.
118. H. Trotter, Eigenvalue distributions of large Hermitian matrices; Wigner’s semicircle law and a theorem of Kac, Murdock, and Szegő, *Adv. in Math.* **54**(1):67–82, 1984.
119. P. Turán, On a problem in the theory of determinants (Chinese), *Acta Math. Sinica* **5** (1955), 411–423.
120. R. Vershynin, Invertibility of symmetric random matrices, *to appear in Random Structures and Algorithms*
121. Voiculescu, D., Addition of certain non-commuting random variables, *J. Funct. Anal.* **66** (1986), 323–346.
122. D. Voiculescu, Limit laws for random matrices and free products, *Invent. Math.* **104** (1991), no. 1, 201–220.
123. J. von Neumann and H. Goldstine, *Numerical inverting matrices of high order*, Bull. Amer. Math. Soc. 53, 1021–1099, 1947.
124. V. Vu, Spectral norm of random matrices, *Combinatorica* 27 (6), 2007, 721–736.
125. K. Wang, Random covariance matrices: Universality of local statistics of eigenvalues up to the edge, [arXiv:1104.4832](https://arxiv.org/abs/1104.4832)
126. P. Wigner, On the distribution of the roots of certain symmetric matrices, *The Annals of Mathematics* **67** (1958) 325–327.

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555
E-mail address: tao@math.ucla.edu

DEPARTMENT OF MATHEMATICS, YALE, NEW HAVEN, CT 06520
E-mail address: van.vu@yale.edu