

Random matrices: the four-moment theorem for Wigner ensembles

TERENCE TAO AND VAN VU

We survey some recent progress on rigorously establishing the universality of various spectral statistics of Wigner random matrix ensembles, focussing in particular on the four-moment theorem and its applications.

1. Introduction

This paper surveys the *four-moment theorem* and its applications in understanding the asymptotic spectral properties of random matrix ensembles of Wigner type. Due to limitations of space, this survey will be far from exhaustive; an extended version will appear elsewhere. (See also [Erdős 2011; Guionnet 2011; Schlein 2011] for some recent surveys in this area.)

To simplify the exposition (at the expense of stating the results in maximum generality), we shall restrict attention to a model class of random matrix ensembles, in which we assume somewhat more decay and identical distribution hypotheses than are strictly necessary for the main results.

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

Tao is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman award. Vu is supported by research grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

MSC2000: 15A52.

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

Example 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 $\xi_{ii} \equiv N(0, 2)_{\mathbb{R}}$ are real gaussians with mean zero and variance 2 for $1 \leq i \leq n$ (thus $\sigma^2 = 2$ in this case). These ensembles obey Condition C0. These ensembles are invariant with respect to conjugation by unitary and orthogonal matrices respectively.

Given an $n \times n$ Hermitian matrix A , we will denote its n eigenvalues in increasing order 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 $Au_i(A) = \lambda_i(A)u_i(A)$.

We also introduce the *eigenvalue counting function*

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

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 $N_I(A_n)$.

2. The local semicircular law

The most fundamental result about the spectrum of Wigner matrices is the *Wigner semicircular law*. We state here a powerful local version of this law, due to Erdős, Schlein, and Yau [Erdős et al. 2009a; 2009b; 2010d] (see also [Erdős et al. 2012a; 2012c; 2012d; 2013] for further refinements). Denote by ρ_{sc} the semicircle density function with support on $[-2, 2]$:

$$\rho_{sc}(x) := \frac{1}{2\pi} (4 - x^2)_+^{1/2}. \quad (2)$$

Theorem 3 (local semicircle 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 probability,¹ one has²*

$$N_I(W_n) = n \int_I \rho_{\text{sc}}(x) dx + o(n|I|). \quad (3)$$

Proof. See, for example, [Tao and Vu 2010, Theorem 1.10]. For the most precise estimates currently known of this type (and with the weakest decay hypotheses on the entries), see [Erdős et al. 2013]. The proofs are based on the Stieltjes transform method; see [Bai and Silverstein 2006] for an exposition of this method. \square

A variant of Theorem 3, established subsequently³ in [Erdős et al. 2012d], is the extremely useful *eigenvalue rigidity property*

$$\lambda_i(W_n) = \lambda_i^{\text{cl}}(W_n) + O_\varepsilon(n^{-1+\varepsilon}), \quad (4)$$

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). This result is key in some of the strongest applications of the theory. Here the *classical location* $\lambda_i^{\text{cl}}(W_n)$ of the i -th eigenvalue is the element of $[-2, 2]$ defined by the formula

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

Roughly speaking, results such as Theorem 3 and (4) 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 (3). Nevertheless, local semicircle laws are an essential part of the fine-scale theory. One particularly useful consequence of these laws is that of *eigenvector delocalisation*, first established in [Erdős et al. 2010d]:

Corollary 4 (eigenvector 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 $1 \leq i, j \leq n$, where the e_1, \dots, e_n are the standard basis of \mathbb{C}^n .*

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

²We use the asymptotic notation $o(X)$ to denote any quantity that goes to zero as $n \rightarrow \infty$ when divided by X , and $O(X)$ to denote any quantity bounded in magnitude by CX , where C is a constant independent of n .

³The result in [Erdős et al. 2012d] 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.

Note from Pythagoras' theorem that $\sum_{j=1}^n |u_i(M_n)^* e_j|^2 = \|u_i(M_n)\|^2 = 1$; thus Corollary 4 asserts, roughly speaking, that the coefficients of each eigenvector are as spread out (or *delocalised*) as possible.

Corollary 4 can be established in a number of ways. One particularly slick approach 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 $z = E + \sqrt{-1}\eta$; it turns out that the machinery used to prove Theorem 3 also can be used to control the resolvent. See for instance [Erdős 2011] for details of this approach.

3. GUE and gauss divisible ensembles

We now turn to the question of the fine-scale behaviour 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 .

A basic object of study 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

$$\begin{aligned} \int_{\mathbb{R}^k} F(x_1, \dots, x_k) R_n^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k \\ = k! \sum_{1 \leq i_1 < \dots < i_k} \mathbf{E} F(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n)) \end{aligned} \quad (5)$$

for all symmetric continuous compactly supported functions $F : \mathbb{R}^k \rightarrow \mathbb{R}$. 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 $\rho_n := (1/n!)R_n^{(n)}$ is the symmetrised joint probability distribution of all n eigenvalues of A_n .

From the semicircular 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

$$\rho_{n,u}^{(k)}(x_1, \dots, x_k) := R_n^{(k)}\left(nu + \frac{x_1}{\rho_{\text{sc}}(u)}, \dots, nu + \frac{x_k}{\rho_{\text{sc}}(u)}\right). \quad (6)$$

It has been generally believed (and in many cases explicitly conjectured; see e.g., [Mehta 1967, p. 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). This phenomenon was motivated by examples of similarly universal laws in physics, such as the laws of thermodynamics or of critical percolation; see, for example, [Mehta 1967; Deift 1999; Deift 2007] 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 *Ginibre's formula*

$$\rho_n(x_1, \dots, x_n) = \frac{1}{(2\pi n)^{n/2}} e^{-|x|^2/2n} \prod_{1 \leq i < j \leq n} |x_i - x_j|^2, \quad (7)$$

for the joint eigenvalue distribution, as can be verified from a standard calculation; see [Ginibre 1965]. From this formula, the theory of determinantal processes, and asymptotics for Hermite polynomials, one can then obtain the limiting law

$$\lim_{n \rightarrow \infty} \rho_{n,u}^{(k)}(x_1, \dots, x_k) = \rho_{\text{Sine}}^{(k)}(x_1, \dots, x_k) \quad (8)$$

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 [Ginibre 1965; Mehta 1967].

Using a general central limit theorem for determinantal processes due to Costin and Leibowitz [1995] and Soshnikov [2002], one can then 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 [Gustavsson 2005]. 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

$$\frac{\lambda_i(A_n) - \lambda_i^{\text{cl}}(A_n)}{\sqrt{\log n / 2\pi / \rho_{\text{sc}}(u)}} \rightarrow N(0, 1)_{\mathbb{R}} \quad (9)$$

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 [Gustavsson 2005].

Much of the above analysis extends to many other classes of invariant ensembles (such as GOE), for which the joint eigenvalue distribution has a form similar to (7); see [Deift 1999] for further discussion. Another important extension of the above results is to the *gauss divisible* ensembles, which are Wigner matrices M_t of the form

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

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 $t \rightarrow +\infty$. By using explicit formulae for the eigenvalue distribution of a gauss divisible matrix, Johansson [2001] was able⁴ to extend the asymptotic (8) 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 semicircle law), Erdős, Péché, Ramirez, Schlein, and Yau [Erdős et al. 2010a] was able to extend the universality of (8) (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 important alternate approach to these results was developed by Erdős et al. [2010c; 2011; 2012b], based on a stability analysis of the Dyson Brownian motion [Dyson 1970] governing the evolution of the eigenvalues of a matrix

⁴Some additional technical hypotheses were assumed in [Johansson 2001], namely that the diagonal variance σ^2 was equal to 1, that the real and imaginary parts of each entry of M_n^0 were independent, and that the matrix entries had bounded C_0 -th moment for some $C_0 > 6$.

Ornstein–Uhlenbeck process. We refer to [Erdős 2011] for a discussion of this method. Among other things, this argument reproves a weaker version of the result in [Erdős et al. 2010a] mentioned earlier, in which one obtained universality for the asymptotic (8) after an additional averaging in the energy parameter u . However, the method was simpler and more flexible than that in [Erdős et al. 2010a], 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.

4. The four-moment theorem

The results discussed above for invariant or gauss divisible ensembles can be extended to more general Wigner ensembles via a powerful swapping method known as the *Lindeberg exchange strategy*, introduced in Lindeberg’s classic proof [1922] of the central limit theorem, and first applied to Wigner ensembles in [Chatterjee 2006]. This method can be used 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

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

when \tilde{M}_n is formed from M_n by replacing⁵ 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 $\mathbf{E}F(M'_n)$:

$$\mathbf{E}F(M_n) - \mathbf{E}F(M'_n) = o(1). \quad (10)$$

The *four-moment theorem* asserts, roughly speaking, that we can obtain conclusions of the form (10) for suitable statistics F as long as M_n, M'_n match to fourth order. More precisely, we have

⁵Technically, the matrices \tilde{M}_n formed by such a swapping procedure are not Wigner matrices as defined in Definition 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 nonidentically 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.

Definition 5 (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 6 (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 C0. 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*

$$|\nabla^j G(x)| \leq n^{c_0} \quad (11)$$

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

$$|\mathbf{E}(G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))) - \mathbf{E}(G(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n)))| \leq n^{-c_0}. \quad (12)$$

A preliminary version of Theorem 6 was first established by the authors in [Tao and Vu 2011b], in the case⁶ of bulk eigenvalues (thus $\delta n \leq i_1, \dots, i_k \leq (1 - \delta)n$ for some absolute constant $\delta > 0$). In [Tao and Vu 2010], the restriction to the bulk was removed; and in [Tao and Vu 2012a], Condition C0 was relaxed to a finite moment condition. We will discuss the proof of this theorem in Section 5. There is strong evidence that the condition of four matching moments is necessary to obtain the conclusion (12); see [Tao and Vu 2011a].

A key technical result used in the proof of the four-moment theorem, which is also of independent interest, is the *gap theorem*:

Theorem 7 (gap theorem). *Let M_n be a Wigner matrix obeying Condition C0. 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 $1 \leq i < n$.

For reasons of space we will not discuss the proof of this theorem here, but refer the reader to [Tao and Vu 2011b; 2012a]. 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

⁶In 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.

smoothness hypothesis on the atom distributions) in [Erdős et al. 2010d].

Another variant of the four-moment theorem was subsequently introduced in [Erdős et al. 2012c], in which the eigenvalues $\lambda_{i_j}(A_n)$ appearing in Theorem 6 were replaced by the components of the resolvent (or Green's function) $(W_n - z)^{-1}$, but with slightly different technical hypotheses on the matrices M_n, M'_n ; see [Erdős et al. 2012c] for full details. As the resolvent-based quantities 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 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 [Knowles and Yin 2013]. Finally, we remark that the four-moment theorem has also been extended to cover eigenvectors as well as eigenvalues; see [Tao and Vu 2012b; Knowles and Yin 2013] for details.

5. Sketch of proof of the four-moment theorem

In this section we discuss the proof of Theorem 6, following the arguments that originated in [Tao and Vu 2011b] and refined in [Tao and Vu 2012a].

In addition to Theorem 7, a 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 8 (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 C0. 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

$$Q_{i_1}, \dots, Q_{i_k} \leq n^{c_0} \tag{13}$$

and obeys the derivative bounds

$$|\nabla^j G(\lambda_{i_1}, \dots, \lambda_{i_k}, Q_{i_1}, \dots, Q_{i_k})| \leq n^{c_0} \quad (14)$$

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

$$\begin{aligned} & \mathbf{E} G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n), Q_{i_1}(A_n), \dots, Q_{i_k}(A_n)) \\ &= \mathbf{E} G(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n), Q_{i_1}(A'_n), \dots, Q_{i_k}(A'_n)) + O(n^{-1/2+O(c_0)}). \end{aligned} \quad (15)$$

We will discuss the proof of this theorem shortly. Using Theorem 7, one can then deduce Theorem 6 from Theorem 8 by smoothly truncating in the Q variables: see [Tao and Vu 2011b, Section 3.3].

It remains to establish Theorem 8. To simplify the exposition slightly, let us assume that the matrices M_n, M'_n are real symmetric rather than Hermitian.

As indicated in Section 4, 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, 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 (15). (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 linearly⁷ on a real parameter t , with $A(0)$ being a Wigner matrix with one entry (and its adjoint) zeroed out, and

⁷If 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.

$A'(0)$ is the explicit elementary Hermitian matrix

$$A'(0) = e_p e_q^* + e_p^* e_q. \quad (16)$$

We note the crucial fact that the random matrix $A(0)$ is independent of both ξ_{pq} and ξ'_{pq} . Note from Condition C0 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

$$F(t) := \mathbf{E} G(\lambda_{i_1}(A(t)), \dots, \lambda_{i_k}(A(t)), Q_{i_1}(A(t)), \dots, Q_{i_k}(A(t))) \quad (17)$$

then our task is to show that

$$\mathbf{E} F(\xi_{pq}) = \mathbf{E} F(\xi'_{pq}) + O(n^{-5/2+O(c_0)}). \quad (18)$$

Suppose that we have Taylor expansions of the form

$$\lambda_{i_l}(A(t)) = \lambda_{i_l}(A(0)) + \sum_{j=1}^4 c_{l,j} t^j + O(n^{-5/2+O(c_0)}) \quad (19)$$

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 (14) 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 (17), 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 (18).

It remains to establish (19) (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

$$\frac{d^j}{dt^j} \lambda_i(A(t)) = O(n^{-j/2+O(c_0)}) \quad (20)$$

for $j = 1, \dots, 5$. As it turns out, this is not quite true as stated, but it becomes true (with overwhelming probability⁸) 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 (13) on the support of G , although there is a technical issue because one will need to establish the bounds (20) for values of t other than ξ_{pq} or $\tilde{\xi}_{pq}$. This difficulty can be overcome by a continuity argument; see [Tao and Vu 2011b]. For the purposes of this informal discussion, we shall ignore this issue and simply assume that we may restrict to the case where

$$Q_i(A(t)) \ll n^{O(c_0)}. \quad (21)$$

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

To prove (20), 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

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

But recall from eigenvalue delocalisation (Corollary 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 (16), we can then obtain (20) in the $j = 1$ case.

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

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

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 (21) and the local semicircle law (Theorem 3). Using these tools, one can conclude (20) 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 [Tao and Vu 2011b] for details.

⁸Technically, 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 [Tao and Vu 2011b] for details.

6. Distribution of individual eigenvalues

One of the simplest applications of the above machinery is to extend the central limit theorem (9) of Gustavsson [2005] for eigenvalues $\lambda_i(A_n)$ in the bulk from GUE to more general ensembles:

Theorem 9. *The gaussian fluctuation law (9) continues to hold for Wigner matrices obeying Condition C0, and whose atom distributions match that of GUE to second order on the diagonal and fourth order off the diagonal; thus, one has*

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

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 (9) one already has

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

(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

$$\mathbf{P}(\lambda_i(A'_n) \in I_-) - n^{-c_0} \leq \mathbf{P}(\lambda_i(A_n) \in I) \leq \mathbf{P}(\lambda_i(A'_n) \in I_+) + n^{-c_0} \quad (22)$$

for all intervals $I = [a, b]$, and n sufficiently large, where

$$I_+ := [a - n^{-c_0/10}, b + n^{-c_0/10}] \quad \text{and} \quad I_- := [a + n^{-c_0/10}, b - n^{-c_0/10}].$$

We will just prove the second inequality in (22), 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

$$\mathbf{P}(\lambda_i(A_n) \in I) \leq \mathbf{E} G(\lambda_i(A_n)) \quad \text{and} \quad \mathbf{E} G(\lambda_i(A'_n)) \leq \mathbf{P}(\lambda_i(A'_n) \in I).$$

On the other hand, one can choose G to obey (11). Thus by Theorem 6 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 (22) 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 10. In [Gustavsson 2005] 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 [Gustavsson 2005] to broader ensembles of matrices, such as gauss divisible matrices, the same argument would allow some of the moment matching hypotheses to be dropped, using tools such as Lemma 13.

Remark 11. In [Doering and Eichelsbacher 2011], 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 9 by using the four-moment theorem. An analogue of Theorem 9 for real symmetric matrices (using GOE instead of GUE) was established in [O’Rourke 2010].

There are similar results at the edge of the spectrum, though with several additional technicalities; see [Soshnikov 1999; Ruzmaikina 2006; Khorunzhiy 2012; Tao and Vu 2010; Johansson 2012; Erdős et al. 2012d].

7. The Wigner–Dyson–Mehta conjecture

We now consider the extent to which the asymptotic (8), 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., [Mehta 1967]) asserts (informally speaking) that (8) is valid for all fixed k , all Wigner matrices and all fixed energy levels $-2 < u < 2$ in the bulk. However, to make this conjecture precise one has to specify the nature of convergence in (8). 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 (8) holds in the vague sense if one has

$$\begin{aligned} \int_{\mathbb{R}^k} F(x_1, \dots, x_k) \rho_{n,u}^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k \\ = \int_{\mathbb{R}^k} F(x_1, \dots, x_k) \rho_{\text{Sine}}^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k \end{aligned} \quad (23)$$

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.

The Wigner–Dyson–Mehta conjecture is largely resolved in the vague convergence category, in the case of Hermitian Wigner matrices whose variances match those of GUE:

Theorem 12 (Wigner–Dyson–Mehta conjecture for Hermitian matrices in the vague sense). *Let M_n be a Wigner matrix obeying Condition C0 and which matches GUE to second order (i.e., the real and imaginary parts of the off-diagonal entries have variance $1/2$, and the diagonal entries have variance 1), and let $-2 < u < 2$ and $k \geq 1$ be fixed. Then (8) holds in the vague sense.*

This theorem, proven in [Tao and Vu 2011c] (in fact the C0 condition can be relaxed to a finite moment condition [Tao and Vu 2011c]), builds upon a long sequence of partial results towards the Wigner–Dyson–Mehta conjecture [Johansson 2001; Erdős et al. 2010a; 2010b; 2010c; 2012c; 2012d; Tao and Vu 2011b], which we will summarise (in a slightly nonchronological order) below. An alternate proof of the result (and a strengthening to the case of matrices with variable variance) was then established in [Erdős and Yau 2012]. The analogous problem for matrices matching GOE (i.e., real symmetric matrices whose diagonal entries have variance 2) remains open, unless one performs an additional averaging in the energy parameter.

As recalled in Section 3, the asymptotic (8) 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 polynomials.⁹

The next major breakthrough was by Johansson [2001], who, as discussed previously, established (8) for gauss divisible ensembles at some fixed time parameter $t > 0$ independent of n , obtained (8) in the vague sense (in fact, the slightly stronger convergence of *weak convergence* was established in that paper, in which the function F in (23) was allowed to merely be L^∞ and compactly supported, rather than continuous and compactly supported). The main tool used in [Johansson 2001] was an explicit determinantal formula for the correlation functions in the gauss divisible case, essentially due to Brézin and Hikami [1997].

In Johansson’s result, the time parameter $t > 0$ had to be independent of n . It was realised 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 [Erdős et al. 2010c], 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 [Erdős et al. 2010a]. By combining this with the method of reverse heat flow discussed in Section 4, the asymptotic (8) (again in

⁹Analogous results are known for much wider classes of invariant random matrix ensembles, see, for example, [Deift et al. 1999; Pastur and Shcherbina 1997; Bleher and Its 1999]. However, we will not discuss these results further here, as they do not directly impact on the case of Wigner ensembles.

the sense of weak convergence) was established for all Wigner matrices whose distribution obeyed certain smoothness conditions (e.g., when $k = 2$ one needs a C^6 type condition), and also decayed exponentially. The methods used in [Erdős et al. 2010a] were an extension of those in [Johansson 2001], 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 [Erdős et al. 2010c] are based instead on an analysis of the Dyson Brownian motion.

By combining the above observation with the moment matching lemma presented below, one immediately concludes Theorem 12 assuming that the off-diagonal atom distributions are supported on at least three points.

Lemma 13 (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 lemma, see [Tao and Vu 2011b, 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 observe from the proof of the four-moment theorem 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 gauss divisible matrix M_n^t associated to M_n at time t is already close enough to matching the first four moments of M_n to apply (a version of) the four-moment theorem. The results of [Erdős et al. 2010a] give the asymptotic (8) for M_n^t , and the eigenvalue rigidity property (4) then allows one to transfer this property to M_n , giving Theorem 12.

Remark 14. The above presentation (drawn from the most recent paper [Tao and Vu 2011c]) 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 [Tao and Vu 2011b], where the first version of the four-moment theorem was introduced, the asymptotic (8) was established under the additional assumptions of Condition C0, and matching the GUE to fourth order; 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 (4) was not yet established (and was instead deduced from the

results of [Gustavsson 2005] combined with the four-moment theorem, thus necessitating the matching moment hypothesis). For related reasons, the paper in [Erdős et al. 2010b] (which first introduced the use of an approximate four-moment theorem) was only able to establish (8) after an additional averaging in the energy parameter u (and with Condition C0). The subsequent progress in [Erdős et al. 2011] via heat flow methods gave an alternate approach to establishing (8), 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 [Erdős et al. 2012d]. In a very recent paper [Erdős et al. 2012a], Condition C0 has been relaxed to finite $(4 + \varepsilon)$ -th moment of the entries 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 nonconstant variances) were also established in [Ben Arous and Pécché 2005; Tao and Vu 2012a; Erdős et al. 2012a; 2012b; 2012c; 2012d; 2013; Wang 2012].

Remark 15. While Theorem 12 is the “right” result for discrete Wigner ensembles, 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 [2011a; 2011b].

Acknowledgements

We are indebted to the anonymous referee for helpful suggestions and references.

References

- [Bai and Silverstein 2006] Z. D. Bai and J. Silverstein, *Spectral analysis of large dimensional random matrices*, Mathematics Monograph Series **2**, Science Press, Beijing, 2006.
- [Ben Arous and Pécché 2005] G. Ben Arous and S. Pécché, “Universality of local eigenvalue statistics for some sample covariance matrices”, *Comm. Pure Appl. Math.* **58**:10 (2005), 1316–1357.
- [Bleher and Its 1999] P. Bleher and A. Its, “Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and universality in the matrix model”, *Ann. of Math. (2)* **150**:1 (1999), 185–266.
- [Brézin and Hikami 1997] E. Brézin and S. Hikami, “An extension of level-spacing universality”, 1997. arXiv cond-mat/9702213
- [Chatterjee 2006] S. Chatterjee, “A generalization of the Lindeberg principle”, *Ann. Probab.* **34**:6 (2006), 2061–2076.
- [Costin and Lebowitz 1995] O. Costin and J. L. Lebowitz, “Gaussian fluctuation in random matrices”, *Phys. Rev. Lett.* **75**:1 (1995), 69–72.

- [Deift 1999] P. A. Deift, *Orthogonal polynomials and random matrices: a Riemann–Hilbert approach*, Courant Lecture Notes in Mathematics **3**, Courant Institute of Mathematical Sciences, New York, 1999.
- [Deift 2007] P. Deift, “Universality for mathematical and physical systems”, pp. 125–152 in *International Congress of Mathematicians, I*, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2007.
- [Deift et al. 1999] 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**:11 (1999), 1335–1425.
- [Doering and Eichelsbacher 2011] H. Doering and P. Eichelsbacher, “Moderate deviations for the eigenvalue counting function of Wigner matrices”, 2011. arXiv 1104.0221
- [Dyson 1970] F. J. Dyson, “Correlations between eigenvalues of a random matrix”, *Comm. Math. Phys.* **19** (1970), 235–250.
- [Erdős 2011] L. Erdős, “Universality of Wigner random matrices: a survey of recent results”, *Uspekhi Mat. Nauk* **66**:3(399) (2011), 67–198. In Russian; translated in *Russian Mathematical Surveys*, **66**:3 (2011), 507–626. arXiv 1004.0861
- [Erdős and Yau 2012] L. Erdős and H.-T. Yau, “A comment on the Wigner–Dyson–Mehta bulk universality conjecture for Wigner matrices”, *Electron. J. Probab.* **17** (2012), no. 28, 5.
- [Erdős et al. 2009a] L. Erdős, B. Schlein, and H.-T. Yau, “Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices”, *Ann. Probab.* **37**:3 (2009), 815–852.
- [Erdős et al. 2009b] L. Erdős, B. Schlein, and H.-T. Yau, “Local semicircle law and complete delocalization for Wigner random matrices”, *Comm. Math. Phys.* **287**:2 (2009), 641–655.
- [Erdős et al. 2010a] L. Erdős, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau, “Bulk universality for Wigner matrices”, *Comm. Pure Appl. Math.* **63**:7 (2010), 895–925.
- [Erdős et al. 2010b] L. Erdős, J. Ramírez, B. Schlein, T. Tao, V. Vu, and H.-T. Yau, “Bulk universality for Wigner Hermitian matrices with subexponential decay”, *Math. Res. Lett.* **17**:4 (2010), 667–674.
- [Erdős et al. 2010c] L. Erdős, J. A. Ramírez, B. Schlein, and H.-T. Yau, “Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation”, *Electron. J. Probab.* **15** (2010), no. 18, 526–603.
- [Erdős et al. 2010d] L. Erdős, B. Schlein, and H.-T. Yau, “Wegner estimate and level repulsion for Wigner random matrices”, *Int. Math. Res. Not.* **2010**:3 (2010), 436–479.
- [Erdős et al. 2011] L. Erdős, B. Schlein, and H.-T. Yau, “Universality of random matrices and local relaxation flow”, *Invent. Math.* **185**:1 (2011), 75–119.
- [Erdős et al. 2012a] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, “Spectral statistics of Erdős–Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues”, *Comm. Math. Phys.* **314**:3 (2012), 587–640.
- [Erdős et al. 2012b] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, “The local relaxation flow approach to universality of the local statistics for random matrices”, *Ann. Inst. Henri Poincaré Probab. Stat.* **48**:1 (2012), 1–46.
- [Erdős et al. 2012c] L. Erdős, H.-T. Yau, and J. Yin, “Bulk universality for generalized Wigner matrices”, *Probab. Theory Related Fields* **154**:1-2 (2012), 341–407.
- [Erdős et al. 2012d] L. Erdős, H.-T. Yau, and J. Yin, “Rigidity of eigenvalues of generalized Wigner matrices”, *Adv. Math.* **229**:3 (2012), 1435–1515.

- [Erdős et al. 2013] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, “Spectral statistics of Erdős–Rényi graphs, I: local semicircle law”, *Ann. Probab.* **41**:3B (2013), 2279–2375.
- [Ginibre 1965] J. Ginibre, “Statistical ensembles of complex, quaternion, and real matrices”, *J. Mathematical Phys.* **6** (1965), 440–449.
- [Guionnet 2011] A. Guionnet, “Grandes matrices aléatoires et théorèmes d’universalité (d’après Erdős, Schlein, Tao, Vu et Yau)”, in *Séminaire Bourbaki 2009/2010* (Exposé 1019), Soc. Math. de France, Paris, 2011.
- [Gustavsson 2005] J. Gustavsson, “Gaussian fluctuations of eigenvalues in the GUE”, *Ann. Inst. H. Poincaré Probab. Statist.* **41**:2 (2005), 151–178.
- [Johansson 2001] K. Johansson, “Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices”, *Comm. Math. Phys.* **215**:3 (2001), 683–705.
- [Johansson 2012] K. Johansson, “Universality for certain Hermitian Wigner matrices under weak moment conditions”, *Ann. Inst. Henri Poincaré Probab. Stat.* **48**:1 (2012), 47–79.
- [Khorunzhiy 2012] O. Khorunzhiy, “High moments of large Wigner random matrices and asymptotic properties of the spectral norm”, *Random Oper. Stoch. Equ.* **20**:1 (2012), 25–68.
- [Knowles and Yin 2013] A. Knowles and J. Yin, “Eigenvector distribution of Wigner matrices”, *Probab. Theory Related Fields* **155**:3-4 (2013), 543–582.
- [Lindeberg 1922] J. W. Lindeberg, “Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung”, *Math. Z.* **15**:1 (1922), 211–225.
- [Maltsev and Schlein 2011a] A. Maltsev and B. Schlein, “Average density of states of Hermitian Wigner matrices”, *Adv. Math.* **228**:5 (2011), 2797–2836.
- [Maltsev and Schlein 2011b] A. Maltsev and B. Schlein, “A Wegner estimate for Wigner matrices”, 2011. arXiv 1103.1473
- [Mehta 1967] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Academic Press, New York, 1967.
- [O’Rourke 2010] S. O’Rourke, “Gaussian fluctuations of eigenvalues in Wigner random matrices”, *J. Stat. Phys.* **138**:6 (2010), 1045–1066.
- [Pastur and Shcherbina 1997] L. Pastur and M. Shcherbina, “Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles”, *J. Statist. Phys.* **86**:1-2 (1997), 109–147.
- [Ruzmaikina 2006] A. Ruzmaikina, “Universality of the edge distribution of eigenvalues of Wigner random matrices with polynomially decaying distributions of entries”, *Comm. Math. Phys.* **261**:2 (2006), 277–296.
- [Schlein 2011] B. Schlein, “Spectral properties of Wigner matrices”, pp. 79–94 in *Mathematical results in quantum physics*, edited by P. Exner, World Sci. Publ., Hackensack, NJ, 2011.
- [Soshnikov 1999] A. Soshnikov, “Universality at the edge of the spectrum in Wigner random matrices”, *Comm. Math. Phys.* **207**:3 (1999), 697–733.
- [Soshnikov 2002] A. Soshnikov, “Gaussian limit for determinantal random point fields”, *Ann. Probab.* **30**:1 (2002), 171–187.
- [Tao and Vu 2010] T. Tao and V. Vu, “Random matrices: universality of local eigenvalue statistics up to the edge”, *Comm. Math. Phys.* **298**:2 (2010), 549–572.
- [Tao and Vu 2011a] T. Tao and V. Vu, “Random matrices: localization of the eigenvalues and the necessity of four moments”, *Acta Math. Vietnam.* **36**:2 (2011), 431–449.
- [Tao and Vu 2011b] T. Tao and V. Vu, “Random matrices: universality of local eigenvalue statistics”, *Acta Math.* **206**:1 (2011), 127–204.

- [Tao and Vu 2011c] T. Tao and V. Vu, “The Wigner–Dyson–Mehta bulk universality conjecture for Wigner matrices”, *Electron. J. Probab.* **16**:77 (2011), 2104–2121.
- [Tao and Vu 2012a] T. Tao and V. Vu, “Random covariance matrices: universality of local statistics of eigenvalues”, *Ann. Probab.* **40**:3 (2012), 1285–1315.
- [Tao and Vu 2012b] T. Tao and V. Vu, “Random matrices: universal properties of eigenvectors”, *Random Matrices Theory Appl.* **1**:1 (2012), 1150001, 27.
- [Wang 2012] K. Wang, “Random covariance matrices: universality of local statistics of eigenvalues up to the edge”, *Random Matrices Theory Appl.* **1**:1 (2012), 1150005, 24.

tao@math.ucla.edu

*Department of Mathematics,
University of California, Los Angeles, 405 Hilgard Avenue,
Los Angeles, CA 90095-1555, United States*

van.vu@yale.edu

*Department of Mathematics, Yale University,
10 Hillhouse Avenue, New Haven, CT 06511, United States*