

## RANDOM MATRICES: LOCALIZATION OF THE EIGENVALUES AND THE NECESSITY OF FOUR MOMENTS

TERENCE TAO AND VAN VU

*Dedicated to Tran Duc Van on the occasion of his sixtieth birthday*

ABSTRACT. Consider the eigenvalues  $\lambda_i(M_n)$  (in increasing order) of a random Hermitian matrix  $M_n$  whose upper-triangular entries are independent with mean zero and variance one, and are exponentially decaying. By Wigner's semicircular law, one expects that  $\lambda_i(M_n)$  concentrates around  $\gamma_i\sqrt{n}$ , where  $\int_{-\infty}^{\gamma_i} \rho_{sc}(x)dx = \frac{i}{n}$  and  $\rho_{sc}$  is the semicircular function.

In this paper, we show that if the entries have vanishing third moment, then for all  $1 \leq i \leq n$

$$\mathbf{E}|\lambda_i(M_n) - \sqrt{n}\gamma_i|^2 = O(\min(n^{-c} \min(i, n+1-i)^{-2/3} n^{2/3}, n^{1/3+\varepsilon})),$$

for some absolute constant  $c > 0$  and any absolute constant  $\varepsilon > 0$ . In particular, for the eigenvalues in the bulk ( $\min\{i, n-i\} = \Theta(n)$ ),

$$\mathbf{E}|\lambda_i(M_n) - \sqrt{n}\gamma_i|^2 = O(n^{-c}).$$

A similar result is achieved for the rate of convergence.

As a corollary, we show that the four moment condition in the Four Moment Theorem is necessary, in the sense that if one allows the fourth moment to change (while keeping the first three moments fixed), then the *mean* of  $\lambda_i(M_n)$  changes by an amount comparable to  $n^{-1/2}$  on the average. We make a precise conjecture about how the expectation of the eigenvalues vary with the fourth moment.

### 1. INTRODUCTION

This note is concerned with the local eigenvalue statistics of the following random matrix model.

**Definition 1.1** (Wigner matrices). A *Wigner matrix* is a random hermitian matrix  $M_n = (\zeta_{ij})_{1 \leq i, j \leq n}$  such that

- The  $\zeta_{ij}$  for  $1 \leq i \leq j \leq n$  are independent with mean zero and variance one, and  $\zeta_{ji} = \overline{\zeta_{ij}}$ ;

---

Received January 10, 2011.

2000 *Mathematics Subject Classification*. 15B52, 60B20.

*Key words and phrases*. Random matrices, localization, rate of convergence, four moment theorem.

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.

- For  $1 \leq i < j \leq n$ ,  $\zeta_{ij}$  are identically distributed, with the real and imaginary parts of  $\zeta_{ij}$  being independent and identically distributed with distribution  $\eta$ ;
- For  $1 \leq i \leq n$ , the  $\zeta_{ii}$  are identically distributed with distribution  $\tilde{\eta}$ ;
- (Uniform exponential decay) There exist constants  $C, C' > 0$  such that

$$(1) \quad \mathbf{P}(|\zeta_{ij}| \geq t^C) \leq \exp(-t)$$

for all  $t \geq C'$  and  $1 \leq i, j \leq n$ .

We refer to  $\eta, \tilde{\eta}$  as the *atom distributions* of  $M_n$ .

A classical example of a Wigner matrix is the *Gaussian unitary ensemble* (GUE), in which  $\eta$  and  $\tilde{\eta}$  are the normal distributions with mean zero and variances  $1/2, 1$  respectively.

**Remark 1.2.** Wigner's matrices are not the most general random matrix model for which the results here are applicable, but we restrict to this case for simplicity. The results, for example, hold for real matrices, in particular Bernoulli matrices.

A Wigner matrix  $M_n$  has  $n$  real eigenvalues

$$\lambda_1(M_n) \leq \dots \leq \lambda_n(M_n).$$

The *global* distribution of these eigenvalues has been known since the 1950s, and is described by the famous *Wigner semicircular law*, which asserts that the empirical spectral measure

$$\frac{1}{n} \sum_{i=1}^n \delta_{\frac{1}{\sqrt{n}} \lambda_i(M_n)}$$

converges almost surely (in the vague topology) to the semicircular distribution  $\rho_{sc}(x) dx$ , where

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

It is of interest to understand the distribution of individual eigenvalues  $\lambda_i(M_n)$ . If for each  $1 \leq i \leq n$  we define the *classical location*  $\gamma_i$  of the normalised  $i^{\text{th}}$  eigenvalue by the formula

$$(2) \quad \int_{-\infty}^{\gamma_i} \rho_{sc}(x) dx = \frac{i}{n},$$

then the Wigner semicircular law (combined with an almost sure bound of  $(2 + o(1))\sqrt{n}$  for the operator norm of  $M_n$ , due to Bai and Yin[4]) is equivalent to the assertion that one has

$$(3) \quad \lambda_i(M_n) = \gamma_i \sqrt{n} + o(\sqrt{n})$$

uniformly for  $1 \leq i \leq n$ , almost surely as  $n \rightarrow \infty$ . If we ignore the  $o(\sqrt{n})$  error in (3), we are thus led to the heuristic

$$(4) \quad \lambda_{i+1}(M_n) - \lambda_i(M_n) \approx \min(i, n-i)^{-1/3} n^{-1/6}$$

for the  $i^{\text{th}}$  eigenvalue spacing. In particular, this spacing should be comparable to  $n^{-1/2}$  in the bulk region  $\delta n \leq i \leq (1 - \delta)n$  (for any fixed  $\delta > 0$ ), and as

large as  $n^{-1/6}$  at the edge of the spectrum. Note though that this derivation is non-rigorous as the  $o(\sqrt{n})$  error could be much larger than the expected gap size (4).

In the last few years, there has been much progress in formalising the above heuristics and obtaining more precise control on the distribution of the eigenvalues and their spacings: see e.g. [16], [5], [17], [6], [18], [14]. A recent survey of these topics can be found in [10].

It is natural to ask whether the  $o(\sqrt{n})$  error in (3) can be improved. In [1], the Talagrand concentration inequality was used to establish (among other things<sup>1</sup>) that

$$(5) \quad \lambda_i(M_n) = \mathbf{M}\lambda_i(M_n) + O(n^\varepsilon \min(i, n+1-i))$$

with probability at least  $1 - O_{A,\varepsilon}(n^{-A})$  (say) for any  $A, \varepsilon > 0$ , where  $\mathbf{M}\lambda_i(M_n)$  is the median of  $\lambda_i(M_n)$ ; see also [11] for closely related results. This was improved in [15] to

$$(6) \quad \lambda_i(M_n) = \mathbf{M}\lambda_i(M_n) + O(n^\varepsilon \min(i, n+1-i)^{1/2})$$

with the same probability of  $1 - O_{A,\varepsilon}(n^{-A})$ . In the bulk region  $\delta n < i < (1-\delta)n$ , the concentration of measure arguments in [11] gives the bound

$$(7) \quad \lambda_i(M_n) = \sqrt{n}\gamma_i + O(n^{1/2+\varepsilon} \min(i, n+1-i)^{-1/3} n^{-1/6})$$

with probability  $1 - O_{A,\varepsilon,\delta}(n^{-A})$  whenever  $\min(i, n+1-i) > n^{1/2+\varepsilon}$  (see Section 2 for further discussion of this bound).

In all the above estimates, the error term is larger than 1. In the recent paper [8, Theorem 7.1], the bound

$$(8) \quad \sum_{i=1}^n \mathbf{E}|\lambda_i(M_n) - \gamma_i \sqrt{n}|^2 = O(n^{1-c})$$

was established for some absolute constant  $c > 0$ . (See also the earlier result in [7, Theorem 6.3], which established (8) under an additional log-Sobolev hypothesis on the distribution.) In the bulk region  $\delta n < i < (1-\delta)n$ , a significantly stronger localisation was obtained in [8] (see the equation preceding (7.8) in that paper), namely that  $\lambda_i(M_n) = \gamma_i \sqrt{n} + O(n^{-1/2+\varepsilon})$  with probability  $O_{\varepsilon,A,\delta}(n^{-A})$  for any  $\varepsilon, A$ , with variants of this result also being obtained closer to the edge. This result was established via a strong bound on the convergence of the Stieltjes transform, which in turn was obtained by a lengthy moment method computation.

Our first main result gives an alternate method to establish eigenvalue localisation, based on the *three moment theorem* rather than on combining the Stieltjes

---

<sup>1</sup>Strictly speaking, the results in [1], [15] only establish the bounds (5), (6) implicitly, and require in addition that the matrix entries are uniformly bounded, rather than exponentially decaying. However, the arguments in these papers can be easily extended to the exponentially decaying case, after a standard truncation argument to reduce to the case when the entries are bounded in magnitude by  $n^{\varepsilon/2}$  (say), and replacing the median with a slightly shifted variant. We omit the details.

transform method with the moment method, and which gives a non-averaged version (8) (in the case when the third moment vanishes):

**Theorem 1.3** (Localisation). *There is an absolute constant  $c > 0$  such that the following holds for any constant  $\varepsilon > 0$ . Let  $M_n$  be a Wigner matrix whose atom distribution  $\eta$  has vanishing third moment  $\mathbf{E}\eta^3 = 0$ . Then for all  $1 \leq i \leq n$ ,*

$$(9) \quad \mathbf{E}|\lambda_i(M_n) - \sqrt{n}\gamma_i|^2 = O(\min(n^{-c} \min(i, n+1-i)^{-2/3} n^{2/3}, n^{1/3+\varepsilon})).$$

One can set  $c$  to be  $1/1000$  (say) and we make no attempt to optimize this constant. By summing over  $i$ , one obtains (8). Furthermore, in the bulk region  $\delta n \leq i \leq (1-\delta)n$ , the theorem implies

$$\mathbf{E}|\lambda_i(M_n) - \sqrt{n}\gamma_i|^2 = O_\delta(n^{-c}).$$

This is not as strong as the recent localisation result obtained in [8], but the proof is shorter (assuming the three moment theorem) and will suffice for our applications. In view of (4), the optimal bound on the right-hand side of (9) should be  $O(\min(i, n-i)^{-2/3} n^{-1/3+\varepsilon})$ .

Let  $N_I$  be the number of eigenvalues (of  $\frac{1}{\sqrt{n}}M_n$ ) in  $I$ , and define  $F_n(x) = \frac{1}{n}\mathbf{E}N_{[-2,x]}$ . The quantity  $\Delta := \sup_x |F_n(x) - \int_{-\infty}^x \rho_{sc}(t) dt|$  is of interest and has been investigated by many researchers (see [3, Chapter 8], [9] and the references therein). In these papers, it has been shown that  $\Delta = O(n^{-1/2})$  under various assumptions (the most general one seems to be in [9] which only requires bounded fourth moment).

The arguments in the proof of Theorem 1.3 can be used to break the  $n^{-1/2}$  barrier, under the extra third moment condition:

**Theorem 1.4.** *There is an absolute constant  $c > 0$  such that the following holds. Let  $M_n$  be a Wigner matrix whose atom distribution  $\eta$  has vanishing third moment  $\mathbf{E}\eta^3 = 0$ . Then  $\Delta = O(n^{-1/2-c})$ .*

We prove this theorem in Section 2.1. The bound  $n^{-1/2-c}$  can be improved to  $n^{-1+\varepsilon}$  if we use median instead of expectation in the definition of  $F_n$  (see Remark 2.6). A related result was proven (using different methods) recently in [8, Theorem 6.3], namely that

$$\frac{1}{n}N_{[-2,x]} = \int_{-\infty}^x \rho_{sc}(t) dt + O(n^{-1/2-c}/|x-2|)$$

with probability  $1 - O(n^{-A})$  for any fixed  $A$ , without a third moment hypothesis.

Next, we give an application of Theorem 1.3 to demonstrate the sharpness (in some sense) of the *four moment theorem*, introduced by the authors in [17, 18] in order to study the distribution of eigenvalues of random matrices. We state a special case of this theorem here:

**Theorem 1.5** (Four Moment Theorem). *For all sufficiently small  $c_0 > 0$  the following holds. Let  $M_n, M'_n$  be two Wigner random matrices whose atom distributions  $\eta, \eta'$  have matching moments to fourth order, thus  $\mathbf{E}\eta^j = \mathbf{E}(\eta')^j$  for*

$j = 3, 4$ . Let  $G : \mathbb{R} \rightarrow \mathbb{R}$  be a smooth function obeying the derivative bounds

$$(10) \quad |G^{(j)}(x)| \leq n^{c_0}$$

for all  $0 \leq j \leq 5$  and  $x \in \mathbb{R}$ . We abbreviate  $\lambda_i := \lambda_i(M_n)$  and  $\lambda'_i := \lambda_i(M'_n)$ .

Then for  $n$  sufficiently large (depending on  $c_0$  and the constants  $C, C'$  in (1)) and all  $1 \leq i \leq n$  one has

$$(11) \quad |\mathbf{E}G(\sqrt{n}\lambda_i) - \mathbf{E}G(\sqrt{n}\lambda'_i)| \leq n^{-c_0}.$$

If the atom distributions  $\eta, \eta'$  have matching moments only to order 3 rather than 4 (i.e.  $\mathbf{E}\eta^3 = \mathbf{E}(\eta')^3$ ), then (11) still holds provided that one strengthens (10) to

$$(12) \quad |G^{(j)}(x)| \leq n^{-Cj c_0}$$

for all  $0 \leq j \leq 5$  and  $x \in \mathbb{R}$ , and some absolute constant  $C$ .

*Proof.* See [17, Theorem 15] (which handled the bulk case when  $\delta n < i < (1-\delta)n$ ) and [18, Theorem 1.13] (which handled the edge case). These theorems can also handle the joint distribution of several eigenvalues at once, as well as somewhat more general ensembles than those in Definition 1.1, but we will not discuss these generalisations here.  $\square$

We will refer to the second part of Theorem 1.5 as the *three moment theorem*.

Roughly speaking, Theorem 1.5 asserts that the distributions of  $\lambda_i(M_n)$  and  $\lambda_i(M'_n)$  differ by  $O(n^{-1/2-c})$  for some  $c > 0$  if the atom distributions have matching moments to order 4, and by  $O_c(n^{-1/2+c})$  for any  $c > 0$  if the atom distributions only have matching moments to order 3. For instance, for sufficiently large  $n$  one has

$$\mathbf{P}(\lambda_i \leq a) \leq \mathbf{P}(\lambda'_i \leq a + n^{-1/2-c}) + n^{-c}$$

for some  $c > 0$  if one has matching moments to order 4, and

$$\mathbf{P}(\lambda_i \leq a) \leq \mathbf{P}(\lambda'_i \leq a + n^{-1/2+c}) + n^{-c}$$

for all  $c > 0$  (with  $n$  sufficiently large depending on  $c$ ) if one has matching moments to order 3. Morally speaking, this means that the medians  $\mathbf{M}\lambda_i, \mathbf{M}\lambda'_i$  of  $\lambda_i, \lambda'_i$  differ by  $O(n^{-1/2-c})$  when there are four matching moments and by  $O_c(n^{-1/2+c})$  when there are three matching moments, although this is not quite rigorous due to the presence of the  $n^{-c}$  error in the above bounds. (For some rigorous bounds on the median of  $\lambda_i$ , see Section 2.)

The matching moment conditions are essential to the method of proof of Theorem 1.5, which uses a Taylor expansion argument. But it is natural to ask if these conditions are in fact necessary. Indeed, if one is not interested in the distribution of *individual* eigenvalues  $\lambda_i(M_n)$ , but instead in the *k-point correlation functions* of these eigenvalues, then in the asymptotic limit  $n \rightarrow \infty$  (and with appropriate normalisations), these correlation functions have a universal distribution regardless of how many matching moments there are; see [17, 6] (with earlier partial results in this direction in [13], [5]). It is not hard to see that the universality for the limiting distributions (or joint distributions) of individual eigenvalues imply the universality of the *k-point correlation function*.

Our second main result below is evidence that the four moment hypothesis is indeed necessary if one wishes to control individual eigenvalues at the scale of the eigenvalue spacing (4).

**Theorem 1.6** (Necessity of fourth moment hypothesis). *Let  $M_n, M'_n$  be Wigner matrices whose atom variables  $\eta, \eta'$  satisfy  $\mathbf{E}\eta^3 = \mathbf{E}(\eta')^3 = 0$  but their fourth moments are different  $\mathbf{E}\eta^4 \neq \mathbf{E}(\eta')^4$ . As before, write  $\lambda_i := \lambda_i(M_n)$  and  $\lambda'_i := \lambda_i(M'_n)$ . Then for all sufficiently large  $n$ , one has*

$$\sum_{i=1}^n |\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i| \geq \kappa n^{1/2}$$

for some  $\kappa$  depending only on the atom distributions. In particular (by the pigeonhole principle), there exists  $1 \leq i \leq n$  such that

$$|\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i| \geq \kappa' \min(i, n+1-i)^{-1/3} n^{-1/6},$$

where  $\kappa' > 0$  depends only on the atom distributions.

Theorem 1.6 is not exactly comparable to the four moment theorem, as it pertains to the *mean* of the eigenvalues  $\lambda_i$ , whereas the four moment theorem instead controls quantities such as the *median*. However, it is expected that the mean and median of  $\lambda_i$  should be quite close (in particular, closer than the expected eigenvalue spacing (4)), but the best known concentration results for  $\lambda_i$  (such as Theorem 1.3) are not strong enough to establish this yet. If one assumes that the mean and median are sufficiently close, then Theorem 1.6 is strong evidence that the four moment theorem breaks down if one only assumes three matching moments.

The three moment theorem implies (roughly speaking) that the medians of  $\lambda_i, \lambda'_i$  should only differ by  $O(n^{-1/2+c_0})$ , for arbitrarily small  $c_0 > 0$ . In view of this, one expects the index  $i$  provided by Theorem 1.6 to lie in the bulk region  $\delta n < i < (1-\delta)n$ , and indeed the conclusion of Theorem 1.6 should in fact hold for *most*  $i$  in this bulk region. However, we were unable to demonstrate this. Nevertheless, concentration bounds such as those given earlier in this section should be able to establish some non-trivial lower bound on  $\min(i, n+1-i)$ .

The question that how each particular eigenvalue reacts to a change in the fourth moment looks very interesting. By utilising higher moments and making some heuristic arguments (see the last section of the paper) we are led to the following precise conjecture.

**Conjecture 1.7** (Conjectured asymptotic). *For  $\delta n \leq i \leq (1-\delta)n$  with  $\delta > 0$  fixed, one has*

$$\mathbf{E}\lambda_i = n^{1/2}\gamma_i + n^{-1/2}C_{i,n} + \frac{1}{4\sqrt{n}}(\gamma_i^3 - 2\gamma_i)\mathbf{E}\eta^4 + O_\delta(n^{-1/2-c})$$

for some absolute constant  $c > 0$ , where  $C_{i,n}$  is some bounded quantity depending only on  $i, n$  (and is in particular independent of  $\eta$ ). The same statement should also be true for the median  $\mathbf{M}\lambda_i$ .

We ran a numerical test to check the conjecture. We generated two random matrices models whose entry's distributions are Gaussian  $N(0, 1)$  and Laplace  $(0, 1/\sqrt{2})$ , and investigated the behaviour of the difference  $\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i$  as a function of  $i$ .

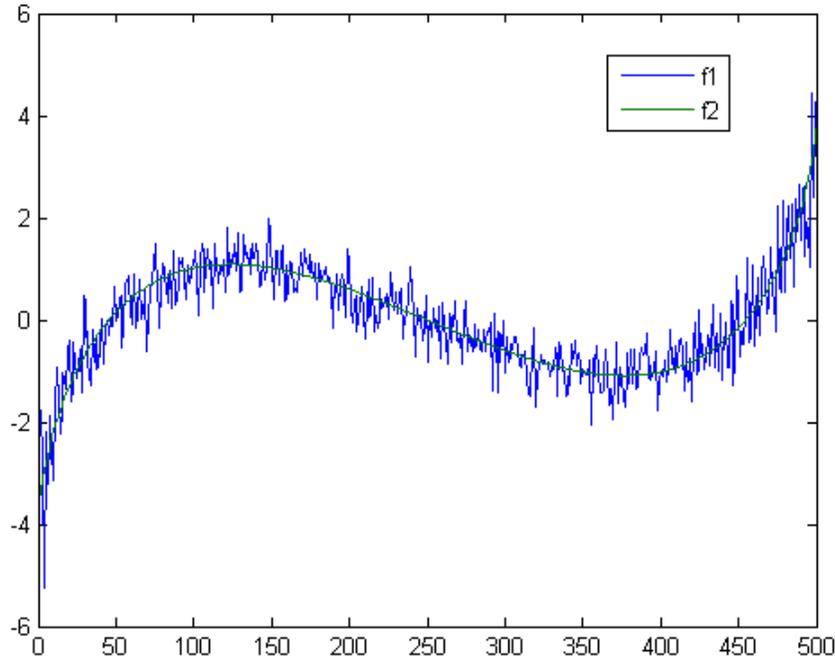


Figure 1: Comparison between 2 curves  $f_1 = 4\sqrt{n} \frac{\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i}{\mathbf{E}\eta^4 - \mathbf{E}\eta'^4}$  and  $f_2 = \gamma_i^3 - 2\gamma_i$ .  
( $n = 500$ )

This conjecture would imply that if one increases the fourth moment, then (in expectation) those  $\lambda_i$  with  $\gamma_i \leq -\sqrt{2}$  or  $0 \leq \gamma_i \leq \sqrt{2}$  are shifted to the left (decreasing), while those  $\lambda_i$  with  $\gamma_i \geq \sqrt{2}$  or  $0 \geq \gamma_i \geq -\sqrt{2}$  are shifted to the right (increasing). In other words, the eigenvalues in the middle move toward the center of the spectrum, while those closer to the edge move outward.

We prove Theorem 1.6 in Section 3. Apart from Theorem 1.3, the main ingredient is a standard moment computation (Lemma 3.1) that compares  $\sum_{i=1}^n \lambda_i^4$  with  $\sum_{i=1}^n (\lambda'_i)^4$ .

*Notation.* We use the usual asymptotic notation as  $n \rightarrow \infty$ , thus  $O(f(n))$  denotes a quantity  $g(n)$  bounded in magnitude by  $Cf(n)$ , and  $o(f(n))$  denotes a quantity  $g(n)$  bounded in magnitude by  $c(n)f(n)$ , where  $c(n) \rightarrow 0$  as  $n \rightarrow \infty$ . If we need,  $c, C$  to depend on additional parameters, we indicate this by subscripts,

e.g.  $o_k(f(n))$  is a quantity bounded in magnitude by  $c_k(n)f(n)$ , where  $c_k(n) \rightarrow 0$  as  $n \rightarrow \infty$  for each fixed  $k$ .

## 2. PROOFS OF THEOREMS 1.3 AND 1.4

We now prove Theorem 1.3. We may assume without loss of generality that  $\varepsilon > 0$  is small; all implied constants can depend on  $\varepsilon$ , and we assume that  $n$  is sufficiently large depending on  $\varepsilon$ .

We first observe that the claim is true in the edge cases  $i = 1, n$ . Indeed, in those cases the estimate (6) (or (5)) gives

$$\lambda_i(M_n) = \mathbf{M}\lambda_i(M_n) + O(n^\varepsilon)$$

with probability  $1 - O(n^{-100})$ . On the other hand, the Tracy-Widom law for Wigner matrices (see [18]) gives

$$\mathbf{M}\lambda_i(M_n) = \gamma_i\sqrt{n} + O(n^{-1/6})$$

at the edge, and the claim follows (with plenty of room to spare).

We may now reduce to the bulk case  $\min(i, n - i) \geq n^{1/2+\varepsilon}$ . Indeed, one can deduce the edge case  $1 < i < n^{1/2+\varepsilon}$  from the bulk case by setting  $i_0$  to be the least integer greater than  $n^{1/2+\varepsilon}$ , and using the crude pointwise bound

$$|\lambda_i(M_n) - \sqrt{n}\gamma_i| \leq |\lambda_1(M_n) - \sqrt{n}\gamma_1| + |\lambda_{i_0}(M_n) - \sqrt{n}\gamma_{i_0}| + |\sqrt{n}\gamma_1 - \sqrt{n}\gamma_{i_0}|$$

and observing that  $\sqrt{n}\gamma_1 - \sqrt{n}\gamma_{i_0} = O(n^{1/6+\varepsilon/3})$ . Similarly to deal with the case  $n - n^{(1+\varepsilon)/2} < i < n$ .

Henceforth, we fix  $i$  with  $\min(i, n - i) \geq n^{1/2+\varepsilon}$ . The next step is to verify the theorem in the model case that  $M_n$  is the GUE random matrix ensemble. In this case, much sharper concentration results are known. Indeed, we have

**Lemma 2.1** (Concentration for GUE). *Let  $M_n$  be a GUE matrix, and let  $I \subset \mathbb{R}$  be an interval. Let  $N_I$  be the counting function  $N_I := \{1 \leq i \leq n : \frac{1}{\sqrt{n}}\lambda_i(M_n) \in I\}$ . Then one has*

$$\mathbf{P}(|N_I - n \int_I \rho_{sc}(x) dx| \geq n^\varepsilon) \leq n^{-100}$$

(say) uniformly in  $I$ , if  $n$  is sufficiently large depending on  $\varepsilon$ .

*Proof.* This follows from the fact that the number of eigenvalues of GUE in  $I$  can be expressed<sup>2</sup> as the sum of independent random variables (see [2, Corollary 4.2.24]), with variance of logarithmic size (see [12, Lemma 2.3]) and thus strongly concentrated. For details, see [2]. In the bulk region  $I \subset [-2 + \delta, 2 - \delta]$ , this type of result (for more general Wigner ensembles) was established in [8, Theorem 6.3].  $\square$

<sup>2</sup>In fact, in the GUE case  $N_I$  has a binomial distribution, though we will not need this fact here.

From this lemma and a standard computation, we conclude that in the GUE case one has

$$(13) \quad \lambda_i(M_n) = \sqrt{n}\gamma_i + O(n^\varepsilon \min(i, n + 1 - i)^{-1/3}n^{-1/6})$$

with probability  $1 - O(n^{-100})$ . From this bound (and using very crude estimates to control the tail event of probability  $O(n^{-100})$ , e.g. controlling  $\lambda_i(M_n)$  by the Frobenius norm of  $M_n$ ), we have

$$\mathbf{E}|\lambda_i(M_n) - \sqrt{n}\gamma_i|^2 = O(n^\varepsilon \min(i, n + 1 - i)^{-1/3}n^{-1/6})^2 + O(n^{-10})$$

(say), which easily implies (9) (with some room to spare).

**Remark 2.2.** In the bulk case  $\delta n < i < (1 - \delta)n$ , a result of Gustavsson [12] shows the related statement that  $\lambda_i(M_n)$  for GUE is asymptotically normally distributed around  $\sqrt{n}\gamma_i$  with variance  $\frac{2 \log n}{(4 - \gamma_i^2)n}$ , which is consistent with the above concentration results. Using the four moment theorem, the result of Gustavsson was extended to other Wigner matrices in [17]. In particular, this gives a bound

$$\mathbf{M}\lambda_i(M_n) = \sqrt{n}\gamma_i + o\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right)$$

for the median uniformly in the bulk region  $\delta n \leq i \leq (1 - \delta)n$  for fixed  $\delta > 0$ , whenever the atom distribution of  $M_n$  has vanishing third moment. As a consequence of the recent results in [8], a similar result (with  $n^\varepsilon$  instead of  $\sqrt{\log n}$ ) holds without the vanishing third moment hypothesis.

Now we pass from the GUE case to more general Wigner matrices with vanishing third moment. The main tool here is the three moment theorem (the second part of Theorem 1.5). We will also need a weak version of Lemma 2.1 in the non-GUE case:

**Lemma 2.3** (Weak concentration for Wigner). *Let  $M_n$  be a Wigner matrix, and let  $I \subset \mathbb{R}$  be an interval. Let  $N_I$  be the counting function  $N_I := \{1 \leq i \leq n : \frac{1}{\sqrt{n}}\lambda_i(M_n) \in I\}$ . Then for any fixed  $\varepsilon > 0$ , one has*

$$\mathbf{P}(|N_I - n \int_I \rho_{sc}(x) dx| \geq n^{1/2+\varepsilon}) \leq n^{-100}$$

(say) uniformly in  $I$ , if  $n$  is sufficiently large depending on  $\varepsilon$ .

*Proof.* This follows from the concentration of measure approach first developed in [11], and then modified in [19, Appendix F] to deal with exponentially decaying entries and with the discontinuous nature of the indicator function  $1_I$ . As mentioned earlier, in the bulk case  $I \subset [-2 + \delta, 2 - \delta]$ , stronger results of this type have also recently been obtained in [8, Theorem 6.3].  $\square$

As a consequence of this lemma and our hypothesis  $\min(i, n - i) > n^{1/2+\varepsilon}$ , one has the bound (7) with probability  $1 - O(n^{-100})$ . In other words, we can find a quantity  $R$  comparable to  $n^{1/2+\varepsilon} \min(i, n + 1 - i)^{-1/3}n^{-1/6}$  such that

$$(14) \quad \mathbf{P}(|\lambda_i(M_n) - \sqrt{n}\gamma_i| \geq R) \leq n^{-100}.$$

We now introduce the function

$$G(x) := \psi\left(\frac{x - n\gamma_i}{\sqrt{n}R}\right)$$

where  $\psi : \mathbb{R} \rightarrow [0, 1]$  is a smooth cutoff function supported on  $[-2, 2]$  that equals  $\psi(x) := \frac{1}{10}x^2$  on  $[-1, 1]$ . Note that  $\sqrt{n}R \geq \sqrt{n}$ . As such, for sufficiently small  $c_0 > 0$  (independent of  $\varepsilon$ ), one easily verifies that

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

for  $0 \leq j \leq 5$  and all  $x \in \mathbb{R}$ , where  $C$  is the constant in the three moment theorem. We may thus apply that theorem and conclude that

$$\mathbf{E}G(\sqrt{n}\lambda_i(M_n)) = \mathbf{E}G(\sqrt{n}\lambda_i(M'_n)) + O(n^{-c_0})$$

for some absolute constant  $c_0 > 0$ , where  $M'_n$  is drawn from GUE. On the other hand, from (13) one easily sees that

$$\mathbf{E}G(\sqrt{n}\lambda_i(M'_n)) = \frac{1}{10R^2} \mathbf{E}|\lambda_i(M'_n) - \sqrt{n}\gamma_i|^2 + O(n^{-10})$$

(say), and similarly from (14) one has

$$\mathbf{E}G(\sqrt{n}\lambda_i(M_n)) = \frac{1}{10R^2} \mathbf{E}|\lambda_i(M_n) - \sqrt{n}\gamma_i|^2 + O(n^{-10}).$$

We conclude that

$$\mathbf{E}|\lambda_i(M_n) - \sqrt{n}\gamma_i|^2 = \mathbf{E}|\lambda_i(M'_n) - \sqrt{n}\gamma_i|^2 + O(n^{-c_0}R^2) + O(n^{-10}).$$

Substituting in the definition of  $R$ , we obtain Theorem 1.3.

**Remark 2.4.** From (13) and the three moment theorem one can also show that for any Wigner matrix  $M_n$  whose atom distribution has the third vanishing moment, and any  $1 \leq i \leq n$ , one has

$$(15) \quad \lambda_i(M_n) = \sqrt{n}\gamma_i + O(n^\varepsilon \min(i, n+1-i)^{-1/3} n^{-1/6})$$

with probability  $1 - O_\varepsilon(n^{-c})$  for some absolute constant  $c > 0$ ; in particular, one has

$$\mathbf{M}\lambda_i(M_n) = \sqrt{n}\gamma_i + O_\varepsilon(n^\varepsilon \min(i, n+1-i)^{-1/3} n^{-1/6}).$$

We omit the details, which are similar to the above calculations and also to the proof of [17, Theorem 32] (which is essentially the bulk case of (15)). Heuristically, this suggests that one can take  $c = 1 - \varepsilon$  in (9), which would be consistent with the results in [8]; however, the available bound  $O(n^{-c})$  of the tail probability for (15) is too weak to make this heuristic rigorous. Unfortunately, even if one assumes more than three matching moments, the methods of proof in [17], [18] do not seem strong enough to establish this conjecture; the main technical obstacle arises from the need to truncate away the event that an eigenvalue gap such as  $\lambda_{i+1}(M_n) - \lambda_i(M_n)$  is unexpectedly small, such events occur with a probability of size  $O(n^{-c})$  but no better.

**Remark 2.5.** The vanishing third moment was needed since we compare to GUE. One can omit this assumption if one can extend Lemma 2.1 to Johansson matrices (i.e. Wigner matrices whose atom distribution is gauss divisible, see [13]). In fact, one only needs this lemma with some  $\varepsilon < 1/2$ . A similar remark applies to the proof of Theorem 1.4 below. The techniques in the recent paper [8] also imply a version of Theorem 1.3 in which no condition on the third moment is required (but with a slightly different right-hand side).

**2.1. Proof of Theorem 1.4.** We now prove Theorem 1.4. The method of proof is only a slight variant of that used above.

Fix  $x$ , and let  $\varepsilon > 0$  be a small absolute constant to be chosen later. From Lemma 2.3 one has

$$i_x - R \leq N_{[-2,x]} \leq i_x + R$$

with probability  $1 - O(n^{-100})$ , where  $R = O(n^{1/2+\varepsilon})$  and

$$i_x := n \int_{-\infty}^x \rho_{sc}(t) dt.$$

Inside the interval  $I_x := [i_x - R, i_x + R] \cap [1, n]$ , we locate  $m = O(n^{1/4+2\varepsilon})$  integers  $i_1, \dots, i_m$  such that every integer in  $I_x$  lies within  $O(n^{1/4-\varepsilon})$  of one of the  $i_j$ . Then with probability  $1 - O(n^{-100})$ , one has

$$N_{[-2,x]} = i_x - R + \sum_{i_x - R < j \leq i_x + R} \mathbf{I}\left(\frac{1}{\sqrt{n}} \lambda_{i_j}(M_n) \leq x\right)$$

where  $\mathbf{I}(E)$  is the indicator of an event  $E$ , and we adopt the conventions that  $\lambda_i(M_n) = -\infty$  for  $i < 1$  and  $\lambda_i(M_n) = +\infty$  for  $i > n$ . Taking expectations, we conclude that

$$F_n(x) = \frac{1}{n}(i_x - R) + \frac{1}{n} \sum_{i_x - R < j \leq i_x + R} \mathbf{P}\left(\frac{1}{\sqrt{n}} \lambda_j(M_n) \leq x\right) + O(n^{-10})$$

(say). Let  $M'_n$  be sampled using GUE. Using the three moment theorem as in [17, Corollary 21], we have

$$\mathbf{P}\left(\frac{1}{n} \lambda_j(M_n) \leq x\right) \leq \mathbf{P}\left(\frac{1}{\sqrt{n}} \lambda_j(M'_n) \leq x + n^{-1+c'}\right) + O(n^{-c})$$

uniformly in  $j$ , for some small absolute  $c, c' > 0$  independent of  $\varepsilon$ . Using this bound, we conclude that

$$\begin{aligned} F_n(x) &\leq \frac{1}{n}(i_x - R) + \frac{1}{n} \sum_{i_x - R < j \leq i_x + R} \mathbf{P}\left(\frac{1}{\sqrt{n}} \lambda_j(M'_n) \leq x + n^{-1+c'}\right) \\ &\quad + O\left(\frac{R}{n} n^{-c}\right) + O(n^{-10}). \end{aligned}$$

The second error term is  $O(n^{-1/2-\varepsilon})$  if  $\varepsilon$  is small enough depending on  $c$ . On the other hand, using Lemma 2.3 for  $M'_n$  instead of  $M_n$ , we conclude from a variant

of the above arguments that

$$F'_n(x+n^{-1+c'}) = \frac{1}{n}(i_x-R) + \frac{1}{n} \sum_{i_x-R < j \leq i_x+R} \mathbf{P}\left(\frac{1}{\sqrt{n}}\lambda_j(M'_n) \leq x+n^{-1+c'}\right) + O(n^{-10}),$$

where  $F'_n$  is the counterpart of  $F_n$  for  $M'_n$  instead of  $M_n$ . We conclude that

$$F_n(x) \leq F'_n(x+n^{-1+c'}) + O(n^{-1/2-\varepsilon});$$

a similar argument also gives

$$F_n(x) \geq F'_n(x-n^{-1+c'}) - O(n^{-1/2-\varepsilon}).$$

On the other hand, from Lemma 2.1 one easily sees that

$$F'_n(x) = \int_{-\infty}^x \rho_{sc}(t) dt + O(n^{-1+\varepsilon})$$

for all  $x$ , and the claim follows.

**Remark 2.6.** Using the three moment theorem, one can control the *median*  $\mathbf{MN}_{[-2,x]}$  with much higher accuracy than the *mean*  $\mathbf{EN}_{[-2,x]}$ . Indeed, using (15) it is not difficult to show that

$$\mathbf{MN}_I = n \int_I \rho_{sc}(x) dx + O_\varepsilon(n^\varepsilon)$$

uniformly for all intervals  $I$  and any  $\varepsilon > 0$ , assuming vanishing third moment of the atom distribution; we leave the details to the interested reader. In view of this, it is reasonable to conjecture that one can take  $c$  arbitrarily close to 1 in Theorem 1.4. In [8], this claim is established in the bulk region  $I \subset [-2+\delta, 2-\delta]$ .

### 3. PROOF OF THEOREM 1.6

We now begin the proof of Theorem 1.6. Let  $M_n, M'_n, \lambda_i, \lambda'_i$  be as in that theorem. The starting point is the following fourth moment calculation:

**Lemma 3.1** (Fourth moment calculation). *Set  $\kappa_0 := \mathbf{E}(\eta^4) - \mathbf{E}((\eta')^4)$ , thus  $\kappa_0 \neq 0$  by hypothesis. Then*

$$\sum_{i=1}^n \mathbf{E}(\lambda_i^4) - \mathbf{E}((\lambda'_i)^4) = 2\kappa_0(n^2 - n).$$

*Proof.* We expand

$$\begin{aligned} \sum_{i=1}^n \mathbf{E}(\lambda_i^4) &= \mathbf{E}(\text{trace } M_n^4) \\ &= \sum_{1 \leq a,b,c,d \leq n} \mathbf{E}\zeta_{ab}\zeta_{bc}\zeta_{cd}\zeta_{da}. \end{aligned}$$

Of course, there is a similar formula for  $\sum_{i=1}^n \mathbf{E}\lambda_i'^4$ , in which the  $\zeta_{ij}$  are replaced by  $\zeta'_{ij}$ .

Consider the four sets  $\{a, b\}, \{b, c\}, \{c, d\}, \{d, a\}$ . If one of these sets occurs with multiplicity one, then the expectation  $\mathbf{E}\zeta_{ab}\zeta_{bc}\zeta_{cd}\zeta_{da}$  vanishes from the mean

zero and independence properties of the coefficients of  $M_n$ . If instead one has two pairs of sets occurring with multiplicity two, then the expectation of  $\mathbf{E}\zeta_{ab}\zeta_{bc}\zeta_{cd}\zeta_{da}$  is equal to that of  $\mathbf{E}\zeta'_{ab}\zeta'_{bc}\zeta'_{cd}\zeta'_{da}$ . From this we see that

$$\sum_{i=1}^n \mathbf{E}(\lambda_i^4) - \sum_{i=1}^n \mathbf{E}((\lambda'_i)^4) = 2 \sum_{1 \leq a < b \leq n} \mathbf{E}(|\zeta_{ab}|^4) - \mathbf{E}(|\zeta'_{ab}|^4).$$

But a short calculation (using the fact that  $\eta, \eta'$  match to the third order) reveals that

$$\mathbf{E}(|\zeta_{ab}|^4) - \mathbf{E}(|\zeta'_{ab}|^4) = 2(\mathbf{E}(\eta^4) - \mathbf{E}((\eta')^4)).$$

The claim follows. □

The value  $\max\{|\lambda_1|, |\lambda_n|\}$  is called the spectral norm of  $M_n$  and will be denoted by  $\|M_n\|$ . The following result is well-known:

**Lemma 3.2** (Concentration of the spectral norm). *For any  $A \geq 0$ , one has*

$$\mathbf{P}(\|M_n\| \geq 3n^{1/2}) = O_A(n^{-A}).$$

*In particular,*

$$\mathbf{P}(|\lambda_i(M_n)| \geq 3n^{1/2}) = O_A(n^{-A})$$

*and*

$$\mathbf{E}|\lambda_i(M_n)|^A = O_A(n^{A/2}).$$

*Proof.* This follows easily from (5) or (6), combined with (3) (as well as using crude estimates, such as Hölder’s inequality, to deal with the rare tail event in which the estimates (5) or (6) fail). Note that this argument allows us to replace the coefficient 3 in the above large deviation inequality by  $2 + o(1)$ , but we will not need this improvement here. For even sharper concentration results, see the recent paper [8]. □

We can now invoke Theorem 1.3 to establish

**Proposition 3.3** (Fourth moment concentration). *We have*

$$\left| \sum_{i=1}^n \mathbf{E}(\lambda_i^4) - (\sqrt{n}\gamma_i)^4 - 4(\sqrt{n}\gamma_i)^3(\mathbf{E}\lambda_i - \sqrt{n}\gamma_i) \right| = O(n^{2-c})$$

*for some absolute constant  $c > 0$ , and similarly for  $\lambda'_i$ .*

*Proof.* We begin with the Taylor expansion

$$\lambda_i^4 = (\sqrt{n}\gamma_i)^4 + 4(\sqrt{n}\gamma_i)^3(\lambda_i - \mathbf{E}\lambda_i) + O(|\lambda_i - \sqrt{n}\gamma_i|^2(|\lambda_i| + \sqrt{n}\gamma_i)^2).$$

From Lemma 3.2, we see that with probability  $1 - O(n^{-100})$  (say), we have  $|\lambda_i| + \sqrt{n}\gamma_i = O(\sqrt{n})$ . Taking expectations and summing using Theorem 1.3 (and using crude estimates, such as Hölder’s inequality, to deal with the tail event of probability  $O(n^{-100})$ ) we obtain the claim. □

Combining Proposition 3.3 with Lemma 3.1 and the triangle inequality, we conclude (for  $n$  large enough) that

$$\left| \sum_{i=1}^n 4\gamma_i^3 (\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i) \right| \geq |\kappa_0| n^{1/2}.$$

Since  $\gamma_i = O(1)$ , Theorem 1.6 now follows from the triangle inequality.

**Remark 3.4.** One can use [8, Theorem 7.1] as a substitute for Theorem 1.3 in the arguments above.

#### 4. HIGHER MOMENT COMPUTATIONS

In this section we discuss a higher moment computation that lead to Conjecture 4.2. We will restate this conjecture at the end of this section for the reader's convenience.

**Lemma 4.1** (Higher moment computations). *Let  $M_n, M'_n, \lambda_i, \lambda'_i$  be as in Theorem 1.6. Set  $\kappa_0 := \mathbf{E}(\eta^4) - \mathbf{E}((\eta')^4)$ , and let  $k \geq 0$  be an integer. Then we have*

$$\sum_{i=1}^n \mathbf{E}(\lambda_i^k) - \mathbf{E}((\lambda'_i)^k) = (2D_{(k-2)/2} \kappa_0 + O_k(n^{-1})) n^{k/2}$$

where the modified Catalan number  $D_m$  is defined to be equal to

$$(16) \quad D_m = \binom{2m+2}{m-1} = \frac{(2m+2)!}{(m-1)!(m+3)!}$$

when  $m = 1, 2, \dots$  is a positive integer, and  $D_m = 0$  otherwise, thus

$$D_0 = 0; \quad D_1 = 1; \quad D_2 = 6; \quad D_3 = 28; \quad D_4 = 120; \quad \dots$$

*Proof.* This is a standard moment method computation (which was the method used by Wigner to prove the semi-circle law [3]; this is also related to the *genus expansion* from string theory). We have

$$\sum_{i=1}^n \mathbf{E}(\lambda_i)^k = \sum_{1 \leq a_1, \dots, a_k \leq n} \mathbf{E} \zeta_{a_1 a_2} \dots \zeta_{a_k a_1}$$

and similarly for  $\sum_{i=1}^n \mathbf{E}(\lambda'_i)^k$ .

Consider the  $k$  sets  $\{a_1, a_2\}, \{a_2, a_3\}, \dots, \{a_k, a_1\}$ . If one of these sets appears with multiplicity one, then the expectation vanishes. If none of the sets appears with multiplicity at least four, then the contributions to  $\sum_{i=1}^n \mathbf{E}(\lambda_i)^k - \sum_{i=1}^n \mathbf{E}(\lambda'_i)^k$  cancel each other out. Thus the only terms that survive are those in which each set appears with multiplicity at least two, and at least one set appears with multiplicity four. In particular, there are at most  $(k-2)/2$  distinct sets  $\{a_i, a_{i+1}\}$  (with the convention  $a_{k+1} = a_1$ ), and thus at most  $k/2$  distinct values of  $a_i$ .

If there are fewer than  $k/2$  distinct values of  $a_i$ , then the total contribution here is easily seen to be  $O_k(n^{k/2-1})$ , which is acceptable. Thus we may restrict attention to the case when there are exactly  $k/2$  distinct values of  $a_i$ , which

forces there to be exactly  $(k - 2)/2$  distinct sets  $\{a_i, a_{i+1}\}$ , and furthermore the (connected) graph formed by these edges cannot contain any cycles and is thus a tree. Finally, each set  $\{a_i, a_{i+1}\}$  must appear with multiplicity two, with the exception of one set that appears with multiplicity four. The summand  $\mathbf{E}\zeta_{a_1 a_2} \dots \zeta_{a_k a_1}$  is then equal to  $2\mathbf{E}\eta^4$ , and similarly  $\mathbf{E}\zeta'_{a_1 a_2} \dots \zeta'_{a_k a_1}$  is then equal to  $2\mathbf{E}(\eta')^4$ .

We now assign each  $a_i$  a label  $j = j(a_i)$  from 1 to  $k/2$  by order of appearance; thus  $a_1$  will be assigned a label  $j(a_1)$  of 1, the first  $a_i$  that is distinct from  $a_1$  will be assigned a label of  $j(a_i) = 2$ , the first  $a_i$  has not already been labeled 1 or 2 will be labeled 3, and so forth. The closed path  $\gamma = ((j(a_1), j(a_2)), \dots, (j(a_{k/2}), j(a_1)))$  then traverses a tree  $T_\gamma$  of  $(k - 2)/2$  edges spanning the vertices  $\{1, \dots, k/2\}$ , where the path  $\gamma$  traverses each edge of  $T_\gamma$  with multiplicity two, with the exception of one edge of  $T_\gamma$  that is traversed four times. Furthermore, the path  $\gamma$  only encounters a vertex  $j$  in  $\{1, \dots, k/2\}$  after it has first encountered  $1, \dots, j - 1$ ; in particular, the starting (and ending) vertex of  $\gamma$  is necessarily 1.

Call a closed path  $\gamma = ((j_1, j_2), \dots, (j_{k/2}, j_1))$  of length  $k/2$  in  $\{1, \dots, k/2\}$  *4-admissible* if it traverses a tree of  $(k - 2)/2$  edges spanning  $\{1, \dots, k/2\}$ , so that each edge is traversed twice with the exception of one edge that is traversed four times, and such that each vertex  $j$  is encountered only after encountering  $1, \dots, j - 1$ . It is not hard to see that each such 4-admissible path contributes

$$(n^{k/2} + O_k(n^{k/2-1})) \times 2\mathbf{E}\eta^4$$

to  $\sum_{i=1}^n \mathbf{E}(\lambda_i)^k$ , and similarly contributes

$$(n^{k/2} + O_k(n^{k/2-1})) \times 2\mathbf{E}(\eta')^4$$

to  $\sum_{i=1}^n \mathbf{E}(\lambda'_i)^k$ . Subtracting, we see that each  $\gamma$  contributes

$$(n^{k/2} + O_k(n^{k/2-1})) \times 2\kappa_0$$

to  $\sum_{i=1}^n \mathbf{E}(\lambda_i)^k - \sum_{i=1}^n \mathbf{E}(\lambda'_i)^k$ . Thus, it will suffice to show that for any  $m$ , the number of 4-admissible paths on trees of  $m$  edges is equal to  $D_m$ .

The claim is trivial unless  $m$  is a positive integer. We observe the recurrence

$$(17) \quad D_m = 2 \sum_{i,j \geq 0: i+j=m-1}^{m-1} C_i D_j + \sum_{i,j,k,l \geq 0: i+j+k+l=m-1} C_i C_j C_k C_l$$

for  $m = 1, 2, \dots$ , where

$$C_m := \frac{(2m)!}{m!(m+1)!}$$

are the Catalan numbers, thus

$$C_0 = 0; \quad C_1 = 1; \quad C_2 = 2; \quad C_3 = 5; \quad C_4 = 14; \quad \dots$$

Indeed, writing

$$c(x) := \sum_{m=0}^{\infty} C_m x^m = \frac{1 - \sqrt{1 - 4x}}{2x}$$

and

$$(18) \quad d(x) := \sum_{m=0}^{\infty} D_m x^m = \frac{(1 - \sqrt{1 - 4x})^4}{16x^3 \sqrt{1 - 4x}}$$

a brief calculation shows that

$$d(x) = 2xc(x)d(x) + xc(x)^4$$

whence the claim.

Call a path  $((j_1, j_2), \dots, (j_{2m}, j_1))$  *2-admissible* if it traverses a tree of  $m$  edges spanning  $\{1, \dots, m + 1\}$ , such that each edge is traversed exactly twice, and each vertex  $j$  is encountered only after encountering  $1, \dots, j - 1$ . It is a classical fact that the number of 2-admissible paths is  $C_m$ .

Now suppose inductively that there are  $D_j$  4-admissible paths on trees of  $j$  edges for all  $j < m$ . It suffices to show that the number of 4-admissible paths on trees of  $m$  edges is given by the right-hand side of (17). To do this, consider the first edge  $(j_1, j_2)$  of an admissible path  $\gamma$  on a tree with  $m$  edges. This edge is traversed either two or four times. Suppose first that it is traversed two times. Then one can split  $\gamma$  into the following pieces: the first edge  $(j_1, j_2)$ , a (reabeled) 2-admissible or 4-admissible path on a tree with  $i$  edges that starts and ends at  $j_2$ , a return edge  $(j_2, j_1)$ , and a (reabeled) 4-admissible or 2-admissible path on a tree with  $j$  edges that is disjoint from the first tree that starts and ends at  $j_1$ , where  $i, j \geq 0$  add up to  $m - 1$ . This case gives a net contribution of

$$\sum_{i,j \geq 0: i+j=m-1} C_i D_j + D_i C_j = 2 \sum_{i,j \geq 0: i+j=m-1} C_i D_j$$

which is the first term of (17).

Now suppose that  $(j_1, j_2)$  is traversed four times. Then we can split  $\gamma$  into the following pieces: the first edge  $(j_1, j_2)$ , a (reabeled) 2-admissible path on a tree with  $i$  edges that starts and ends at  $j_2$ , a return edge  $(j_2, j_1)$ , a (reabeled) 2-admissible path with  $j$  edges that starts and ends at  $j_1$ , a repeated edge  $(j_1, j_2)$ , a (reabeled) 2-admissible path with  $k$  edges that starts and ends at  $j_2$ , a repeated return edge  $(j_2, j_1)$ , and a (reabeled) 2-admissible path on a tree with  $l$  edges that starts and ends at  $j_1$ , where all trees are disjoint (except at their roots) and  $i, j, k, l \geq 0$  add up to  $m - 1$ . This gives the second contribution to the right-hand side of (17).  $\square$

We now repeat the arguments from the previous section. A routine generalisation of Proposition 3.3 yields the bound

$$\left| \sum_{i=1}^n \mathbf{E}(\lambda_i^k) - (\sqrt{n}\gamma_i)^k - k(\sqrt{n}\gamma_i)^{k-1}(\mathbf{E}\lambda_i - \sqrt{n}\gamma_i) \right| = O_k(n^{k-c})$$

for any  $k \geq 1$  and some absolute constant  $c > 0$  (note that the left-hand side vanishes for  $k = 1$ ). We conclude from this and Lemma 4.1 that

$$(19) \quad \frac{1}{n} \sum_{i=1}^n k\gamma_i^{k-1} \sqrt{n}(\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i) = 2D_{(k-2)/2} \kappa_0 + O_k(n^{-c}).$$

Next, we observe from (18) that

$$-\sum_{k=0}^{\infty} \frac{D_{(k-2)/2}}{z^k} = -\frac{(z - \sqrt{z^2 - 4})^4}{16\sqrt{z^2 - 4}}$$

for  $z$  near infinity, where we pick the branch of the square root of  $\sqrt{z^2 - 4}$  that equals  $z$  near infinity and is analytic away from the interval  $[-2, 2]$ . Then the right-hand side continues analytically to the exterior of this interval. Calling this analytic function  $f(z)$ , we compute the jump formula

$$\lim_{b \rightarrow 0} \frac{f(x + ib) - f(x - ib)}{2\pi i} = g(x),$$

where  $g : \mathbb{R} \rightarrow \mathbb{R}$  vanishes outside of the interval  $[-2, 2]$ , and is equal to

$$g(x) := \frac{1}{2\pi} \frac{x^4 - 4x^2 + 2}{\sqrt{4 - x^2}}$$

on this interval. From the Cauchy integral formula we conclude the moment formula

$$D_{(k-2)/2} = \int_{-2}^2 g(x)x^k dx$$

for  $k = 0, 1, 2, \dots$ . The antiderivative of  $g(x)$  is

$$-\frac{1}{8\pi}(x^3 - 2x)\sqrt{4 - x^2} = -\frac{1}{8}(x^3 - 2x)\rho_{sc}(x)$$

so by an integration by parts we have

$$D_{(k-2)/2} = \int_{-2}^2 \frac{1}{8}(x^3 - 2x)kx^{k-1} \rho_{sc}(x)dx.$$

By Riemann integration (or more precisely, the trapezoid rule), the right-hand side is equal to

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{8}(\gamma_i^3 - 2\gamma_i)k\gamma_i^{k-1} + O_k(n^{-c})$$

for some absolute constant  $c > 0$ .

Thus if we introduce the normalised shift

$$s_i := \sqrt{n}(\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i) - \frac{1}{4}(\gamma_i^3 - 2\gamma_i)\kappa_0$$

we can rewrite (19) as

$$(20) \quad \frac{1}{n} \sum_{i=1}^n k\gamma_i^{k-1} s_i = O_k(n^{-c}).$$

This suggests (but does not rigorously prove<sup>3</sup>) that the  $s_i$  are small, of size  $O(n^{-c})$ , at least in the bulk region  $\delta n \leq i \leq (1 - \delta)n$ . In particular, we are led to the

<sup>3</sup>Specifically, the difficulty is that there could be cancellation between nearby values of  $s_i$ . If one could show some assertion to the effect that  $s_i \approx s_{i'}$  when  $i, i'$  are close together, then this would go a long way towards establishing Conjecture 4.2 below.

conjecture (Conjecture 4.2) that

$$(\mathbf{E}\lambda_i - \mathbf{E}\lambda'_i) = \frac{1}{4\sqrt{n}}(\gamma_i^3 - 2\gamma_i)\kappa_0 + O(n^{-1/2-c})$$

and this in turn suggests the following asymptotic for the expected value of  $\lambda_i$ :

**Conjecture 4.2** (Conjectured asymptotic). *For  $\delta n \leq i \leq (1 - \delta)n$  with  $\delta > 0$  fixed, one has*

$$\mathbf{E}\lambda_i = n^{1/2}\gamma_i + n^{-1/2}C_{i,n} + \frac{1}{4\sqrt{n}}(\gamma_i^3 - 2\gamma_i)\mathbf{E}\eta^4 + O_\delta(n^{-1/2-c})$$

for some absolute constant  $c > 0$ , where  $C_{i,n}$  is some bounded quantity depending only on  $i, n$  (and is in particular independent of  $\eta$ ). The same statement should also be true for the median  $\mathbf{M}\lambda_i$ .

The bound on  $C_{i,n}$  is plausible in view of results such as Lemma 2.1. It should in fact be possible to obtain (at least conjecturally) a precise value for  $C_{i,n}$  from an analysis of the GUE case. Such an asymptotic would demonstrate more precisely the dependence of the  $i^{\text{th}}$  eigenvalue on the fourth moment  $\mathbf{E}\eta^4$  at the scale  $\Theta(n^{-1/2})$  of the mean eigenvalue spacing.

#### ACKNOWLEDGEMENT

We would like to thank K. Johansson for a useful conversation and O. Zeitouni for confirming Lemma 2.1. Thanks to R. Killip, M. Meckes, and H.T. Yau for corrections and to Linh Tran for the numerical experiments.

#### REFERENCES

- [1] N. Alon, M. Krivelevich, and V. Vu, On the concentration of eigenvalues of random symmetric matrices, *Israel J. Math.* **131** (2002), 259–267.
- [2] G. Anderson, A. Guionnet and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2009.
- [3] Z. D. Bai and J. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Mathematics Monograph Series **2**, Science Press, Beijing 2006.
- [4] Z. D. Bai and Y. Q. Yin, Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix, *Ann. Probab.* **16** (1988), 1729–1741.
- [5] L. Erdős, J. Ramirez, S. Peche, B. Schlein and H-T. Yau, Bulk universality for Wigner matrices, *Comm. Pure Appl. Math.* **63** (2010), no. 7, 895–925.
- [6] L. Erdős, J. Ramirez, B. Schlein, T. Tao, V. Vu, and H-T. Yau, Bulk universality for Wigner hermitian matrices with subexponential decay, *Math. Res. Lett.* **17** (2010), no. 4, 667–674.
- [7] L. Erdős, H-T. Yau, and J. Yin, Bulk universality for generalized Wigner matrices, preprint.
- [8] L. Erdős, H-T. Yau, and J. Yin, Universality for generalized Wigner matrices with Bernoulli distributions, preprint.
- [9] F. Götze and A. Tikhomirov, Rate of convergence to the semi-circular law. *Probab. Theory Related Fields* **127** (2003), no. 2, 228–276.
- [10] A. Guionnet, Grandes matrices aléatoires et théorèmes d’universalité, *Séminaire Bourbaki*, 62<sup>ème</sup> année, 2009-2010, no. 1019.
- [11] A. Guionnet and O. Zeitouni, Concentration of the spectral measure for large matrices, *Electron. Comm. Probab.* **5** (2000), 119–136 (electronic).
- [12] J. Gustavsson, Gaussian fluctuations of eigenvalues in the GUE, *Ann. Inst. H. Poincaré Probab. Statist.* **41** (2005), no. 2, 151–178.

- [13] K. Johansson, Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices, *Comm. Math. Phys.* **215** (2001), no. 3, 683–705.
- [14] K. Johansson, Universality for certain Hermitian Wigner matrices under weak moment conditions, preprint.
- [15] M. Meckes, Concentration of norms and eigenvalues of random matrices, *J. Funct. Anal.* **211** (2004), no. 2, 508–524.
- [16] A. Soshnikov, Universality at the edge of the spectrum in Wigner random matrices, *Comm. Math. Phys.* **207** (1999), no. 3, 697–733.
- [17] T. Tao and V. Vu, Random matrices: Universality of the local eigenvalue statistics, *to appear in Acta Mathematica*.
- [18] T. Tao and V. Vu, Random matrices: universality of local eigenvalue statistics up to the edge, *Comm. Math. Phys.* **298** (2010), no. 2, 549–572.
- [19] T. Tao and V. Vu, Random matrices: The distribution of the smallest singular values, *Geom. Funct. Anal.* **20** (2010), no. 1, 260–297.

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

DEPARTMENT OF MATHEMATICS, RUTGERS, PISCATAWAY, NJ 08854  
*E-mail address:* vanvu@math.rutgers.edu