THE ASYMPTOTIC DISTRIBUTION OF A SINGLE EIGENVALUE GAP OF A WIGNER MATRIX

TERENCE TAO

ABSTRACT. We show that the distribution of (a suitable rescaling of) a single eigenvalue gap $\lambda_{i+1}(M_n) - \lambda_i(M_n)$ of a random Wigner matrix ensemble in the bulk is asymptotically given by the Gaudin-Mehta distribution, if the Wigner ensemble obeys a finite moment condition and matches moments with the GUE ensemble to fourth order. This is new even in the GUE case, as prior results establishing the Gaudin-Mehta law required either an averaging in the eigenvalue index parameter i, or fixing the energy level u instead of the eigenvalue index.

The extension from the GUE case to the Wigner case is a routine application of the Four Moment Theorem. The main difficulty is to establish the approximate independence of the eigenvalue counting function $N_{(-\infty,x)}(\tilde{M}_n)$ (where $\tilde{M}_n$ is a suitably rescaled version of M_n) with the event that there is no spectrum in an interval $[x, x+s]$, in the case of a GUE matrix. This will be done through some general considerations regarding determinantal processes given by a projection kernel.

1. INTRODUCTION

Given an $n \times n$ Hermitian matrix M_n, we let

$$\lambda_1(M_n) \leq \ldots \leq \lambda_n(M_n)$$

be the n eigenvalues of M_n in non-decreasing order, counting multiplicity. The purpose of this paper is to study the eigenvalue gaps $\lambda_{i+1}(M_n) - \lambda_i(M_n)$ of such matrices when M_n is drawn from the Gaussian Unitary Ensemble (GUE), or more generally from a Wigner random matrix ensemble, in the asymptotic limit $n \to \infty$ and for a single $i = i(n)$ in the bulk region $\varepsilon n \leq i \leq (1-\varepsilon)n$.

To begin with, let us set out our notational conventions for GUE and Wigner ensembles:

Definition 1 (Wigner and GUE). 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), and each ξ_{ij} has mean zero and variance one. We say that the Wigner

matrix ensemble *obeys condition* **C1** *with constant* C_0 if one has

$$\sup_{i,j} \mathbf{E}|\xi_{ij}|^{C_0} \leq C$$

for some constant C (independent of n).

A *GUE matrix* M_n is a Wigner Hermitian matrix in which ξ_{ij} is drawn from the complex gaussian distribution $N(0,1)_{\mathbb{C}}$ (thus the real and imaginary parts are independent copies of $N(0,1/2)_{\mathbb{R}}$) for $i \neq j$, and ξ_{ii} is drawn from the real gaussian distribution $N(0,1)_{\mathbb{R}}$.

A Wigner matrix $M_n = (\xi_{ij})_{1 \leq i,j \leq n}$ is said to *match moments to m^{th} order* with another Wigner matrix $M_n' = (\xi_{ij}')_{1 \leq i,j \leq n}$ for some $m \geq 1$ if one has

$$\mathbf{E}(\operatorname{Re}\xi_{ij})^a (\operatorname{Im}\xi_{ij})^b = \mathbf{E}(\operatorname{Re}\xi_{ij}')^a (\operatorname{Im}\xi_{ij}')^b$$

whenever $a,b \in \mathbf{N}$ with $a + b \leq m$.

The bulk distribution of the eigenvalues $\lambda_1(M_n), \ldots, \lambda_n(M_n)$ of a Wigner (and in particular, GUE) matrix is governed by the *Wigner semicircle law*. Indeed, if we let $N_I(M_n)$ denote the number of eigenvalues of M_n in an interval I, and we assume Condition **C1** for some $C_0 > 2$, then with probability[1] $1 - o(1)$, we have the asymptotic

$$N_{\sqrt{n}I}(M_n) = n \int_I \rho_{\text{sc}}(u) \, du + o(n)$$

uniformly in I, where ρ_{sc} is the Wigner semi-circular distribution

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

see e.g. [1]. Informally, this law indicates that the eigenvalues $\lambda_i(M_n)$ are mostly contained in the interval $[-2\sqrt{n}, 2\sqrt{n}]$, and for any energy level u in the bulk region $-2 + \varepsilon \leq u \leq 2 - \varepsilon$ for some fixed $\varepsilon > 0$, the average eigenvalue spacing should be $\frac{1}{\sqrt{n}\rho_{\text{sc}}(u)}$ near $\sqrt{n}u$.

Now let M_n be drawn from GUE. The distribution of the eigenvalues $\lambda_1(M_n), \ldots, \lambda_n(M_n)$ are then well-understood. If we define the k-point correlation functions $\rho_k^{(n)} : \mathbb{R}^k \to \mathbb{R}^+$ for $0 \leq k \leq n$ to be the unique symmetric continuous function for which

$$\mathbf{E} \sum_{1 \leq i_1 < \ldots < i_k \leq n} F(\lambda_{i_1}(M_n), \ldots, \lambda_{i_k}(M_n)) = \int_{\mathbb{R}^k} F(x_1, \ldots, x_k)\rho_k^{(n)}(x_1, \ldots, x_k) \, dx_1 \ldots dx_k$$

for any continuous function F which is compactly supported in the region $\{x_1 \leq \ldots \leq x_k\}$, then one has the well-known formula of Dyson [9]

$$\rho_n^{(n)}(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^n x_i^2/2} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2$$

and the *Gaudin-Mehta formula*

$$\rho_k^{(n)}(x_1, \ldots, x_k) = \det(K^{(n)}(x_i, x_j))_{1 \leq i,j \leq k}$$

---

[1]See Section 2 for the conventions for asymptotic notation such as $o(1)$ that are used in this paper.

where $K^{(n)}(x, y)$ is the kernel

$$(1) \qquad K^{(n)}(x, y) := \sum_{k=0}^{n-1} P_k(x) e^{-x^2/4} P_k(y) e^{-y^2/4}$$

and $P_0(x), P_1(x), \ldots$ are the L^2-normalised orthogonal polynomials with respect to the measure $e^{-x^2/2} \, dx$ (and are thus essentially Hermite polynomials); see e.g. [19] or [1]. In particular, the functions $P_k(x) e^{-x^2/4}$ for $i = 0, \ldots, n-1$ are an orthonormal basis to the subspace $V^{(n)}$ of $L^2(\mathbb{R}) = L^2(\mathbb{R}, dx)$ spanned by $x^i e^{-x^2/4}$ for $i = 0, \ldots, n-1$, thus the orthogonal projection $P^{(n)}$ to this subspace is given by the formula

$$P^{(n)} f(x) = \int_{\mathbb{R}} K^{(n)}(x, y) f(y) \, dy$$

for any $f \in L^2(\mathbb{R})$.

Applying the inclusion-exclusion formula, the Gaudin-Mehta formula implies that for any interval[2] I, the probability $\mathbf{P}(N_I(M_n) = 0)$ that M_n has no eigenvalues in I, where $N_I(M_n)$ is the number of eigenvalues of M_n in I, is equal to

$$\mathbf{P}(N_I(M_n) = 0) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \int_I \ldots \int_I \det(K^{(n)}(x_i, x_j))_{1 \leq i,j \leq k} \, dx_1 \ldots dx_k.$$

One can also express this probability as a Fredholm determinant

$$\mathbf{P}(N_I(M_n) = 0) = \det(1 - 1_I P^{(n)} 1_I),$$

where we view the indicator function 1_I as a multiplier operator on $L^2(\mathbb{R})$.

The asymptotics of $K^{(n)}$ as $n \to \infty$ are also well understood, especially in the bulk of the spectrum[3], which in this normalisation corresponds to the interval $[(-2 + \varepsilon)\sqrt{n}, (2 - \varepsilon)\sqrt{n}]$ for any fixed $\varepsilon > 0$. Indeed, if $-2 + \varepsilon < u < 2 + \varepsilon$ and x, y are bounded uniformly in n, then from the Plancherel-Rotarch asymptotics for Hermite polynomials one has

$$(2) \qquad \frac{1}{\rho_{\mathrm{sc}}(u)\sqrt{n}} K^{(n)}(u\sqrt{n} + \frac{x}{\rho_{\mathrm{sc}}(u)\sqrt{n}}, u\sqrt{n} + \frac{y}{\rho_{\mathrm{sc}}(u)\sqrt{n}}) = K_{\mathrm{Sine}}(x, y) + o(1)$$

where K_{Sine} is the Dyson sine kernel

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

with the usual convention that $K(x, x) = 1$; see e.g. [19], [1], or [8, Corollary 1]. Note that the normalisations in (2) are consistent with the heuristic, from the

---

[2]One can generalise this formula from intervals to arbitrary Borel measurable sets, but in our applications we will only need the interval case. Similarly for many of the other determinantal process identities used in this paper.

[3]For the edge of the spectrum, one can control individual eigenvalues instead by the Tracy-Widom law [29]. There is however a transitional regime between the bulk and the edge which is not covered by either our results or by the Tracy-Widom law, e.g. when $\min(i, n-i)$ is comparable to n^θ for some fixed $0 < \theta < 1$, and which may be worth further attention. Given that convergence to the sine kernel is also known in such regimes, one would expect the main results of this paper to extend to these settings, although to make this intuition rigorous would require some careful argument which we do not pursue here.

Wigner semi-circular law, that the mean eigenvalue spacing at $u\sqrt{n}$ is $\frac{1}{\rho_{\mathrm{sc}}(u)\sqrt{n}}$. We observe that the Dyson sine kernel is also the kernel to the orthogonal projection P_{Sine} to those functions $f \in L^2(\mathbb{R}, dx)$ whose Fourier transform

$$\hat{f}(\xi) := \int_{\mathbb{R}} e^{-2\pi i x \xi} f(x) \, dx$$

is supported on the interval $[-1/2, 1/2]$.

From (2) and some careful treatment of error terms (see e.g. [1, Chapter 3]) one obtains that

$$\mathbf{P}(N_{u\sqrt{n}+\frac{1}{\rho_{\mathrm{sc}}(u)\sqrt{n}}I}(M_n) = 0) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_I \cdots \int_I \det(K_{\mathrm{Sine}}(x_i, x_j))_{1 \le i, j \le k} \, dx_1 \ldots dx_k + o(1),$$

or in Fredholm determinant form,

$$\mathbf{P}(N_{u\sqrt{n}+\frac{1}{\rho_{\mathrm{sc}}(u)\sqrt{n}}I}(M_n) = 0) = \det(1 - 1_I P_{\mathrm{Sine}} 1_I) + o(1).$$

Note that the kernel $K_{\mathrm{Sine}}(x, y)1_I(y)$ of $P_{\mathrm{Sine}}1_I$ is square-integrable, and so $P_{\mathrm{Sine}}1_I$ is in the Hilbert-Schmidt class, and so $1_I P_{\mathrm{Sine}} 1_I = (P_{\mathrm{Sine}}1_I)^*(P_{\mathrm{Sine}}1_I)$ is trace class.

This asymptotic can in turn be used to control the distribution of the averaged gap spacing distribution. Indeed, if $1 < t_n < n$ is any sequence such that $1/t_n, t_n/n = o(1)$, then for any $-2 + \varepsilon < u < 2 - \varepsilon$ and $s > 0$ independent of n, the quantity

$$S(s, t_n, u, M_n) := \frac{\#\{1 \le i \le n - 1 : \lambda_{i+1}(M_n) - \lambda_i(M_n) \le \frac{s}{\sqrt{n}\rho_{\mathrm{sc}}(u)}; |\lambda_i(M_n) - u\sqrt{n}| \le \frac{t_n}{\sqrt{n}}\}}{2t_n}$$

has the asymptotic

$$(3) \qquad\qquad S(s, t_n, u, M_n) = \int_0^s p(y) \, dy + o(1)$$

where p is the *Gaudin distribution* (or *Gaudin-Mehta distribution*)

$$p(y) := \frac{d^2}{dy^2} \det(1 - 1_{[0,y]} P_{\mathrm{Sine}} 1_{[0,y]}),$$

or equivalently

$$(4) \qquad\qquad \det(1 - 1_{[0,y]} P_{\mathrm{Sine}} 1_{[0,y]}) = \int_y^\infty p(z)(z - y) \, dz;$$

see [7] for details. The quantity $\det(1 - 1_{[0,y]} P_{\mathrm{Sine}} 1_{[0,y]})$ (and hence the Gaudin distribution $p(y)$) can also be expressed in terms of a solution to a Painlevé V ordinary differential equation. More precisely, one has

$$\det(1 - 1_{[0,y]} P_{\mathrm{Sine}} 1_{[0,y]}) = \exp\left(\int_0^{\pi y} \frac{\sigma(x)}{x} \, dx\right)$$

where σ solves the ODE

$$(x\sigma'')^2 + 4(x\sigma' - \sigma)(x\sigma' - \sigma + (\sigma')^2) = 0$$

with boundary condition $\sigma(x) \sim -\frac{x}{\pi}$ as $x \to 0$; see [15] (or the later treatment in [30]). Among other things, this implies that the Gaudin distribution p and all of its derivatives are[4] smooth, bounded, and rapidly decreasing on $(0, +\infty)$.

We also remark that the extreme values of the gaps $\lambda_{i+1}(M_n) - \lambda_i(M_n)$ are also well understood; see [2]. However, our focus here will be on the bulk distribution of these gaps rather than on the tail behaviour.

In [25], a *Four Moment Theorem* for the eigenvalues of Wigner matrices was established, which roughly speaking asserts that the fine scale statistics of these eigenvalues depend only on the first four moments of the coefficients of the Wigner matrix, so long as some decay condition (such as Condition **C1**) is obeyed. In particular, by applying this theorem to the asymptotic (3) for GUE matrices, one obtains

Corollary 2. *The asymptotic* (3) *is also valid for Wigner matrices M_n which obey Condition* **C1** *for some sufficiently large absolute constant C_0, and which match moments with GUE to fourth order.*

Proof. See [25, Theorem 9]. Strictly speaking, the arguments in that paper require an exponential decay hypothesis on the coefficients on M_n rather than a finite moment condition, because the four moment theorem in that paper also has a similar requirement. However, the refinement to the four moment theorem established in the subsequent paper [26] (or in the later papers [27], [17]) relaxes that exponential decay condition to a finite moment condition. $\square$

We remark that the moment matching hypothesis in this corollary can in fact be removed by combining the above argument with some similar results obtained (by a different method) in [16], [10]; see [11].

The Wigner semi-circle law predicts that the location of an individual eigenvalue $\lambda_i(M_n)$ of a Wigner or GUE matrix M_n for i in the bulk region $\varepsilon n \leq i \leq (1 - \varepsilon)n$ should be approximately $\sqrt{n}u$, where $u = u_{i/n}$ is the *classical location* of the eigenvalue, given by the formula

$$(5) \qquad \int_{-\infty}^{u} \rho_{\mathrm{sc}}(y) \ dy = \frac{i}{n}.$$

Indeed, it is a result of Gustavsson [14] that $\frac{\lambda_i(M_n) - \sqrt{n}u}{\sqrt{\log n/2\pi^2}/\sqrt{n}\rho_{\mathrm{sc}}(u)}$ converges in distribution to the standard real Gaussian distribution $N(0,1)_{\mathbb{R}}$, or more informally that

$$(6) \qquad \lambda_i(M_n) \approx N(\sqrt{n}u, \frac{\log n/2\pi^2}{(\sqrt{n}\rho_{\mathrm{sc}}(u))^2}).$$

---

[4]Indeed, the famous *Wigner surmise* predicts the reasonably accurate approximation $p(x) \approx \frac{1}{2}\pi x e^{-\pi x^2/4}$; see e.g. [19].

Note that the standard deviation $\frac{\sqrt{\log n/2\pi^2}}{\sqrt{n}\rho_{\mathrm{sc}}(u)}$ here exceeds the mean eigenvalue spacing $\frac{1}{\sqrt{n}\rho_{\mathrm{sc}}(u)}$ by a factor comparable to $\sqrt{\log n}$. If one heuristically applies this approximation (6) to the gap distribution law (3), one is led to the conjecture that the normalised eigenvalue gap

$$\frac{\lambda_{i+1}(M_n) - \lambda_i(M_n)}{1/(\sqrt{n}\rho_{\mathrm{sc}}(u))}$$

should converge in distribution to the Gaudin distribution, in the sense that

$$(7) \qquad \mathbf{P}\Big(\frac{\lambda_{i+1}(M_n) - \lambda_i(M_n)}{1/(\sqrt{n}\rho_{\mathrm{sc}}(u))} \le s\Big) = \int_0^s p(y)\ dy + o(1)$$

for any fixed $s > 0$.

Unfortunately, this is not quite a rigorous proof of (7). The problem is that the asymptotic (3) involves not just a single eigenvalue gap $\lambda_{i+1} - \lambda_i$, but is instead an average over all eigenvalue gaps near the energy level $\sqrt{n}u$. By (6), one is then forced to consider the contributions of at least $\gg \sqrt{\log n}$ different values of i that could contribute to (3). One would of course expect the behaviour of $\lambda_{i+1} - \lambda_i$ for adjacent values of i to be essentially identical, in which case one could pass from the averaged gap distribution (3) to the individual gap distribution (7). However, it is *a priori* conceivable (though admittedly quite strange) that there is non-trivial dependence on i, for instance that $\lambda_{i+1} - \lambda_i$ might tend to be larger than predicted by the Gaudin distribution for even i, and smaller than predicted for odd i, with the two effects canceling out in averaged statistics such as (3), but not in non-averaged statistics such as (7).

Our main result rules out such a pathological possibility:

Theorem 3 (Individual gap spacing). *Let M_n be drawn from GUE, and let $\varepsilon n \le i \le (1 - \varepsilon)n$ for some fixed $\varepsilon > 0$. Then one has the asymptotic (7) for any fixed $s > 0$, where $u = u_{i/n}$ is given by (5).*

Applying the four moment theorem from [25] (with the extension to the finite moment setting in [26]), one obtains an immediate corollary:

Corollary 4. *The conclusion of Theorem 3 is also valid for Wigner matrices M_n which obey Condition* **C1** *for some sufficiently large absolute constant C_0, and which match moments with GUE to fourth order.*

Proof. This can be established by repeating the proof of [25, Theorem 9] (in fact the argument is even simpler than this, because one is working with a single eigenvalue gap rather than with an average, and can proceed more analogously to the proof of [25, Corollary 21]). We omit the details. $\square$

In view of the results in [11], it is natural to conjecture that the moment matching condition can be removed. Following [11], it would be natural to use heat flow methods to do so, in particular by trying to extend Theorem 3 to the *gauss divisible* ensembles studied in [16]. However, the methods in this paper rely very heavily

on the determinant form of the joint eigenvalue distribution of GUE (and not just on control of the k-point correlation functions); the formulae in [16] also have some determinantal structure, but it is unclear to us whether this similarity of structure is sufficient to replicate the arguments[5]. On the other hand, we expect analogues Theorem 3 to be establishable for other ensembles with a determinantal form, such as GOE and GSE, or to more general β ensembles involving a non-quadratic potential for the classical values $1, 2, 4$ of β. We will not pursue these matters here.

The key to proving Theorem 3 lies in establishing the approximate independence[6] of the eigenvalue counting function $N_{(-\infty, x)}(\tilde{M}_n)$ from the event that $\tilde{M}_n$ has no eigenvalues in a short interval $[x, x+s]$ (i.e. that $N_{[x,x+s]}(\tilde{M}_n) = 0$), where $\tilde{M}_n$ is a suitably rescaled version of M_n. Roughly speaking, this independence, coupled with a central limit theorem for $N_{(-\infty, x)}(\tilde{M}_n)$, will imply that the distribution of a gap $\lambda_{i+1}(M_n) - \lambda_i(M_n)$ is essentially invariant with respect to small changes in the i parameter. To obtain this approximate independence, we use the properties of determinantal processes, and in particular the fact that a determinantal point process Σ, when conditioned on the event that a given interval such as $[x, x+s]$ contains no elements of Σ, remains a determinantal point process (though with a slightly different kernel). The main difficulty is then to ensure that the new kernel is close to the old kernel in a suitable sense (more specifically, we will compare the two kernels in the nuclear norm S^1).

We thank Peter Forrester, Van Vu, and the anonymous referee for corrections.

2. Notation

In this paper, n will be an asymptotic parameter going to infinity. A quantity is said to be *fixed* if it does not depend on n; if a quantity is not specified as fixed, then it is permitted to vary with n. Given two quantities X, Y, we write $X = O(Y)$, $X \ll Y$, or $Y \gg X$ if we have $|X| \le CY$ for some fixed C, and $X = o(Y)$ if X/Y goes to zero as $n \to \infty$.

An *interval* will be a connected subset of the real line, which may possibly be half-infinite or infinite. If I is an interval, we use $I^c := \mathbb{R} \backslash I$ to denote its complement.

---

[5]By using heat flow methods such as the method of local relaxation flow [12], one can obtain control on energy-averaged correlation functions in this setting, and similarly for non-classical β-ensembles $\beta \neq 1, 2, 4$ as was done recently in [5]. Such bounds are sufficient to obtain averaged gap information of the form (3) (at least for values of t_n that grow faster than logarithmic), but it is not obvious how to isolate a single eigenvalue gap to then obtain (7).

[6]It may be surprising to the experts that the counting functions on $(-\infty, x)$ and $[x, x+s]$ are approximately independent, as the intervals are adjacent. The point is that while there is a correlation between the two counting functions, the covariance between them is essentially of order $O(1)$, whilst the variance of $N_{(-\infty, x)}$ is of order $\log n$, and so the correlation between the two random variables ends up being asymptotically negligible. To put it another way, most of the random fluctuation of $N_{(-\infty, x)}$ comes from the portion of the spectrum that is far away from x, and this contribution will be almost completely decoupled from the spectrum at $[x, x+s]$.

We use $\sqrt{-1}$ to denote the imaginary unit, in order to free up the symbol i for other purposes, such as indexing eigenvalues.

Given a bounded operator A on a Hilbert space H, we denote the operator norm of A as $\|A\|_{\mathrm{op}}$. We will also need the *Hilbert-Schmidt norm* (or *Frobenius norm*)

$$\|A\|_{HS} := (\mathrm{trace}(A^*A))^{1/2} = (\mathrm{trace}(AA^*))^{1/2},$$

with the convention that this norm is infinite if A^*A or AA^* is not trace class. Similarly, we will need the *Schatten 1-norm* (or *nuclear norm*)

$$\|A\|_{S^1} := \mathrm{trace}((A^*A)^{1/2}) = \mathrm{trace}((AA^*)^{1/2}),$$

which is finite when A is trace class. Note that if A is compact with non-zero singular values $\sigma_1, \sigma_2, \ldots$ then we have

$$\|A\|_{\mathrm{op}} = \sup_i |\sigma_i|$$

$$\|A\|_{HS} = (\sum_i |\sigma_i|^2)^{1/2}$$

$$\|A\|_{S^1} = \sum_i |\sigma_i|.$$

Indeed, one should view the operator, Hilbert-Schmidt, and nuclear norms as non-commutative versions of the ℓ^∞, ℓ^2, and ℓ^1 norms respectively.

For us, the reason for introducing the nuclear norm S^1 is that it controls the trace:

$$|\mathrm{trace}\, A| \leq \|A\|_{S^1}.$$

On the other hand, the Hilbert-Schmidt and operator norms are significantly easier to estimate than the nuclear norm. To bridge the gap, we will rely heavily on the non-commutative Hölder inequalities

$$\|AB\|_{\mathrm{op}} \leq \|A\|_{\mathrm{op}}\|B\|_{\mathrm{op}}$$
$$\|AB\|_{HS} \leq \|A\|_{\mathrm{op}}\|B\|_{HS}$$
$$\|AB\|_{HS} \leq \|A\|_{HS}\|B\|_{\mathrm{op}}$$
$$\|AB\|_{S^1} \leq \|A\|_{\mathrm{op}}\|B\|_{S^1}$$
$$\|AB\|_{S^1} \leq \|A\|_{S^1}\|B\|_{\mathrm{op}}$$
$$\|AB\|_{S^1} \leq \|A\|_{HS}\|B\|_{HS};$$

see e.g. [4]. We will use these inequalities in this paper without further comment.

We remark that for integral operators

$$Tf(x) := \int_{\mathbb{R}} K(x,y)f(y)\, dy$$

on $L^2(\mathbb{R})$ for locally integrable K, the Hilbert-Schmidt norm of T is given by

$$\|T\|_{HS} = (\int_{\mathbb{R}} \int_{\mathbb{R}} |K(x,y)|^2\, dxdy)^{1/2}$$

when the right-hand side is finite.

3. Some general theory of determinantal processes

In this section we record some of the theory of determinantal processes which we will need. We will not attempt to exhaustively describe this theory here, referring the interested reader to the surveys [21], [23] or [20] instead. We will also not aim for maximum generality in this section, restricting attention to determinantal processes on $\mathbb{R}$, whose associated locally trace class operator P will usually be an orthogonal projection, and often of finite rank.

Define a *good kernel* to be a locally integrable function $K : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$, such that the associated integral operator

$$Pf(x) := \int_{\mathbb{R}} K(x,y)f(y) \ dy$$

can be extended from $C_c(\mathbb{R})$ to a self-adjoint bounded operator on $L^2(\mathbb{R})$, with spectrum in $[0,1]$. Furthermore, we require that P be locally trace class in the sense that for every compact interval I, the operator $1_I P 1_I$ is trace class; this will for instance be the case if K is smooth. If K is a good kernel, then (as was shown in [18], [23]; see also [20] or [1]), K defines a point process $\Sigma \subset \mathbb{R}$, i.e. a random subset[7] of $\mathbb{R}$ that is almost surely locally finite, with the k-point correlation functions

(8) $$\rho_k(x_1, \ldots, x_k) := \det(K(x_i, x_j))_{1 \leq i,j \leq k}$$

for any $k \geq 0$, thus

$$\mathbf{E} \prod_{i=1}^{k} \#(\Sigma \cap I_i) = \int_{I_1 \times \ldots \times I_k} \rho_k(x_1, \ldots, x_k) \ dx_1 \ldots dx_k$$

for any disjoint intervals $I_1, \ldots, I_k$. This process is known as the *determinantal point process* with kernel K.

The distribution of a determinantal point process in an interval I is described by the following lemma:

Lemma 5. *Let Σ be a determinantal point process on $\mathbb{R}$ associated to a good kernel K and associated operator P. Let I be a compact interval, and suppose that the operator $1_I P 1_I$ has non-zero eigenvalues $\lambda_1, \lambda_2, \ldots \in (0,1]$. Then $\#(\Sigma \cap I)$ has the same distribution as $\sum_i \xi_i$, where the ξ_i are jointly independent Bernoulli random variables, with each ξ_i equalling 1 with probability λ_i and 0 with probability $1 - \lambda_i$. In particular, one has*

$$\mathbf{E} \#(\Sigma \cap I) = \sum_i \lambda_i = \mathrm{trace}(1_I P 1_I)$$

and

$$\mathbf{Var} \#(\Sigma \cap I) = \sum_i (1 - \lambda_i)\lambda_i = \mathrm{trace}((1 - 1_I P 1_I)1_I P 1_I),$$

---

[7]Strictly speaking, a point process is permitted to have multiplicity, so that it becomes a multiset rather than a set. However, as we are restricting attention to kernels K which are locally integrable, the determinantal point processes we consider will be almost surely simple, in the sense that no multiplicity occurs.

and

$$\mathbf{P}(\#(\Sigma \cap I) = 0) = \prod_i (1 - \lambda_i) = \det(1 - 1_I P 1_I).$$

Proof. See e.g. [1, Corollary 4.2.24]. $\qquad\qquad\square$

As a corollary of Lemma 5, we see that $\mathbf{P}(\#(\Sigma \cap I) = 0) > 0$ unless P has an eigenfunction of eigenvalue 1 that is supported on I.

An important special case of determinantal point processes arises when the operator P is an orthogonal projection of some finite rank n, which is the situation with the GUE point process $\{\lambda_1(M_n), \ldots, \lambda_n(M_n)\}$, which as discussed in the introduction is a determinantal point process with kernel $K^{(n)}$ given by (1). In this case, the hypotheses on P (i.e. self-adjoint trace class with eigenvalues in $[0, 1]$) are automatically satisfied, and the determinantal point process Σ is almost surely a set of cardinality n; see e.g. [23], [20] or [1]. In this situation, the k-point correlation functions ρ_k vanish for $k > n$, and for $k < n$ we have the *Gaudin lemma*

$$(9) \qquad \rho_k(x_1, \ldots, x_k) = \frac{1}{n-k} \int_{\mathbb{R}} \rho_{k+1}(x_1, \ldots, x_{k+1}) \, dx_{k+1}$$

which allows one to recursively obtain the correlation functions from the n-point correlation function ρ_n (which is essentially the joint density function of the n elements of Σ). Note that (9) in fact holds for any point process whose cardinality is almost surely n, if the process is almost surely simple with locally integrable correlation functions.

If V is the n-dimensional range of P, and $\phi_1, \ldots, \phi_n$ is an orthonormal basis for V, then the kernel K of the orthogonal projection P can be expressed explicitly as

$$K(x, y) = \sum_{i=1}^{n} \phi_i(x) \overline{\phi_i(y)}$$

and thus (by the basic formula $\det(A^* A) = |\det(A)|^2$)

$$(10) \qquad \rho_n(x_1, \ldots, x_n) = |\det(\phi_i(x_j))_{1 \le i, j \le n}|^2.$$

This leads to the following consequence:

Proposition 6 (Exclusion of an interval). *Let Σ be a determinantal process associated to the orthogonal projection P_V to an n-dimensional subspace V of $L^2(\mathbb{R})$. Let I be a compact interval, and suppose that no non-trivial element of V is supported in I. Then the event $E := (\#(\Sigma \cap I) = 0)$ occurs with non-zero probability, and upon conditioning to this event E, the resulting random variable $(\Sigma|E)$ is a determinantal point process associated to the orthogonal projection $P_{1_{I^c} V}$ to the n-dimensional subspace $1_{I^c} V$ of $L^2(\mathbb{R})$.*

Proof. This is a continuous variant of [21, Proposition 6.3], and can be proven as follows. By construction, P_V has no eigenvector of eigenvalue 1 supported in I, and so $\mathbf{P}(E) = \det(1 - 1_I P_V)$ is non-zero. The point process $(\Sigma|E)$ clearly has

cardinality n almost surely, and is thus described by its n-point correlation function, which is a constant multiple of

$$\rho_n(x_1, \ldots, x_n)1_{I^c}(x_1) \ldots 1_{I^c}(x_n),$$

which by (10) can be written as

$$(11) \qquad\qquad |\det(\phi_i 1_{I^c}(x_j))_{1 \leq i,j \leq n}|^2,$$

where $\phi_1, \ldots, \phi_n$ is an orthonormal basis for V.

By hypothesis on V, $\phi_1 1_{I^c}, \ldots, \phi_n 1_{I^c}$ is a (not necessarily orthonormal) basis for $1_{I^c}V$. By row operations, we can thus write (11) as a constant multiple of

$$|\det(\phi_i'(x_j))_{1 \leq i,j \leq n}|^2,$$

where $\phi_1', \ldots, \phi_n'$ is an orthonormal basis for $1_{I^c}V$. But this is the n-point correlation function for the determinantal point process of $P_{1_{I^c}V}$. As the n-point correlation function of an n-point process integrates to $n!$ (cf. (9)), we see that the n-point correlation function of $(\Sigma|E)$ must be exactly equal to that of the determinantal point process of $P_{1_{I^c}V}$, as claimed. $\qquad\square$

It is likely that the above proposition can be extended to infinite-dimensional projections (possibly after imposing some additional regularity hypotheses), but we will not pursue this matter here.

We saw in Lemma 5 that if Σ is a determinantal point process and I is a compact interval, then the random variable $\#(\Sigma \cap I)$ is the sum of independent Bernoulli random variables. If the variance of this sum is large, then such a sum should converge to a gaussian, by the central limit theorem. Examples of such central limit theorems for $\#(\Sigma \cap I)$ were formalised in [6], [24]. We will need a slight variant of these theorems, which gives uniform convergence on the probability density function of $\#(\Sigma \cap I)$ as opposed to the probability distribution function.

Lemma 7 (Discrete density version of central limit theorem). *Let $X = \xi_1 + \ldots + \xi_n$ be the sum of independent Bernoulli random variables $\xi_1, \ldots, \xi_n$, with mean $\mathbf{E}X = \mu$ and variance $\mathbf{Var}X = \sigma^2$ for some $\sigma > 0$. Then for any integer m, one has*

$$\mathbf{P}(X = m) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(m-\mu)^2/2\sigma^2} + O(\sigma^{-1.7}).$$

One can improve the error term here to $O(\sigma^{-2})$ with a little more effort, but any error better than $1/\sigma$ will suffice for our purposes.

Proof. We may assume that σ is larger than any given absolute constant, as the claim is trivial otherwise. We use the Fourier-analytic method. Write $p_i := \mathbf{E}\xi_i$, then

$$\mu = \sum_{i=1}^{n} p_i$$

and

$$(12) \qquad \sigma^2 = \sum_{i=1}^{n} p_i(1 - p_i).$$

Observe that X has characteristic function

$$\mathbf{E}e^{2\pi\sqrt{-1}tX} = \prod_{i=1}^{n}((1 - p_i) + p_i e^{2\pi\sqrt{-1}t})$$

and so

$$\mathbf{P}(X = m) = \int_{-1/2}^{1/2} \prod_{i=1}^{n}((1 - p_i)e^{-2\pi\sqrt{-1}p_it} + p_i e^{2\pi(1-p_i)\sqrt{-1}t})e^{-2\pi\sqrt{-1}(m-\mu)t} \ dt.$$

We can rewrite this integral slightly as $A + B$, where

$$A := \int_{|t|\leq\sigma^{-0.9}} \prod_{i=1}^{n}((1 - p_i)e^{-2\pi\sqrt{-1}p_it} + p_i e^{2\pi\sqrt{-1}(1-p_i)t})e^{-2\pi\sqrt{-1}(m-\mu)t} \ dt$$

and

$$B := \int_{\sigma^{-0.9}<t\leq 1/2} \prod_{i=1}^{n}((1 - p_i)e^{-2\pi\sqrt{-1}p_it} + p_i e^{2\pi\sqrt{-1}(1-p_i)t})e^{-2\pi\sqrt{-1}(m-\mu)t} \ dt.$$

We first control A. From Taylor expansion one has

$$(1 - p_i)e^{-2\pi\sqrt{-1}p_it} + p_i e^{2\pi\sqrt{-1}(1-p_i)t} = \exp(-2\pi^2 p_i(1 - p_i)t^2 + O(p_i(1 - p_i)|t|^3))$$

in this regime, and so by (12)

$$\prod_{i=1}^{n}(1 - p_i)e^{-2\pi\sqrt{-1}p_it} + p_i e^{2\pi\sqrt{-1}(1-p_i)t} = \exp(-2\pi^2\sigma^2 t^2 + O(\sigma^{-0.7})).$$

We therefore have

$$A = \int_{-\sigma^{-0.9}}^{\sigma^{-0.9}} (1 + O(\sigma^{-0.7}))e^{-2\pi^2\sigma^2 t^2}e^{-2\pi\sqrt{-1}(m-\mu)t} \ dt.$$

Since

$$\int_{\mathbb{R}} e^{-2\pi^2\sigma^2 t^2} \ dt = \frac{1}{\sqrt{2\pi}\sigma}e^{-(m-\mu)^2/2\sigma^2}$$

and

$$\int_{\mathbb{R}} e^{-2\pi^2\sigma^2 t^2}e^{-2\pi\sqrt{-1}(m-\mu)t} \ dt = \frac{1}{\sqrt{2\pi}\sigma}e^{-(m-\mu)^2/2\sigma^2}$$

and

$$\int_{|t|\geq\sigma^{-0.9}} e^{-2\pi^2\sigma^2 t^2} \ dt = O(\sigma^{-100})$$

(say), we conclude that

$$(13) \qquad A = \frac{1}{\sqrt{2\pi}\sigma}e^{-(m-\mu)^2/2\sigma^2} + O(\sigma^{-1.7}).$$

Now we control B. Elementary computation shows that

$$|(1 - p_i)e^{-2\pi\sqrt{-1}p_it} + p_i e^{2\pi\sqrt{-1}(1-p_i)t}| \leq \exp(-cp_i(1 - p_i)t^2)$$

in this regime for some absolute constant $c > 0$. By (12), we may thus bound

$$|B| \le \int_{|t| \ge \sigma^{-0.9}} e^{-c\sigma^2 t^2} \, dt$$

and so $B = O(\sigma^{-1.7})$. Combining this with (13), the claim follows. $\qquad \square$

Combining Lemma 5, Proposition 6, and Lemma 7 we immediately obtain

Corollary 8. *Let Σ be a determinantal point process on $\mathbb{R}$ whose kernel P is an orthogonal projection to an n-dimensional subspave V of $L^2(\mathbb{R})$. Let I be a compact interval, and m be an integer. Then*

$$\mathbf{P}(\#(\Sigma \cap I) = m) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(m-\mu)^2/2\sigma^2} + O(\sigma^{-1.7}).$$

where

$$\mu := \text{trace}(1_I P 1_I)$$

and

$$\sigma^2 := \text{trace}((1 - 1_I P 1_I)(1_I P 1_I)).$$

Furthermore, if J is another compact interval disjoint from J, such that no non-trivial element of V is supported on J, then

$$\mathbf{P}(\#(\Sigma \cap I) = m | \#(\Sigma \cap J) = 0) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} e^{-(m-\tilde{\mu})^2/2\tilde{\sigma}^2} + O(\tilde{\sigma}^{-1.7}),$$

where

$$\tilde{\mu} := \text{trace}(\tilde{P} 1_I)$$

and

$$\tilde{\sigma}^2 := \text{trace}(\tilde{P} 1_{I^c} \tilde{P} 1_I),$$

and $\tilde{P}$ is the orthogonal projection to $1_{J^c} V$.

Let the notation be as in the above corollary. In our application, we will need to determine the extent to which events such as $\#(\Sigma \cap I) = m$ and $\#(\Sigma \cap J) = 0$ are independent. In view of Corollary 8, it is then natural to determine the extent to which the projection $\tilde{P}$ differs from that of P.

Observe that as no non-trivial element of V is supported on J, the operator $P 1_{J^c} P$, viewed as a map from V to V, is invertible. Denoting its inverse on V by $(P 1_{J^c} P)_V^{-1}$, we then see that the operator $1_{J^c} P (P 1_{J^c} P)_V^{-1} P 1_{J^c}$ is self-adjoint, idempotent, and has $1_{J^c} V$ as its range, and so must be equal to $\tilde{P}$:

$$(14) \qquad \tilde{P} := 1_{J^c} P (P 1_{J^c} P)_V^{-1} P 1_{J^c}.$$

We can also write $(P 1_{J^c} P)_V^{-1}$ as $(1 - P 1_J P)_V^{-1}$ (since P is the identity on V). Thus, by Neumann series, we formally have the expansion

$$\tilde{P} = 1_{J^c} P 1_{J^c} + 1_{J^c} P 1_J P 1_{J^c} + 1_{J^c} P 1_J P 1_J P 1_{J^c} + \dots$$

This expansion is convergent for sufficiently small J, but does not necessarily converge for J large. However, in practice we will be able to invert $1 - P 1_J P$ by a perturbation argument involving the Fredholm alternative. More precisely, in our application, the finite rank projection P will be "close" in some weak sense to an infinite rank projection P_0 (in our application, P_0 will be the Dyson projection

P_{Sine}), projecting to some infinite-dimensional Hilbert space V_0. We will assume P_0 to be locally trace class, so that $P_0 1_J P_0$ is compact. If we assume that no non-trivial element of V_0 is supported on J, then the Fredholm alternative (see e.g. [22, Theorem VI.14]) then implies that $1 - P_0 1_J P_0 : V_0 \to V_0$ is invertible, with inverse $(1 - P_0 1_J P_0)_{V_0}^{-1} = 1 + K_0$ for some compact operator K_0, thus

$$(15) \qquad\qquad (1 + K_0)(1 - P_0 1_J P_0) = 1$$

on $L^2(\mathbb{R})$. As $1 - P_0 1_J P_0$ is self-adjoint and is the identity on V_0, we see that K_0 has range in V_0 and cokernel in $V_0^\perp$, thus

$$K_0 = P_0 K_0 = K_0 P_0 = P_0 K_0 P_0.$$

One then expects $1 + P K_0 P : V \to V$ to be an approximate inverse to $1 - P 1_J P$. Indeed, we have

$$(16) \qquad\qquad (1 + P K_0 P)(1 - P 1_J P) = 1 + E$$

where

$$E := P K_0 P - P(1 + K_0) P 1_J P.$$

Meanwhile, from (15) we have

$$(17) \qquad\qquad K_0 = (1 + K_0) P_0 1_J P_0$$

and thus

$$E = P(1 + K_0)(P_0 1_J P_0 - P 1_J P) P.$$

Let us now bound some norms of E. As the projection operator P has an operator norm of at most 1, one has

$$\|E\|_{\text{op}} \le (1 + \|K_0\|_{\text{op}})\|P_0 1_J P_0 - P 1_J P\|_{\text{op}};$$

splitting $P_0 1_J P_0 - P 1_J P$ as $(P_0 - P) 1_J P_0 - P 1_J (P_0 - P)$ we conclude that

$$\|E\|_{\text{op}} \le (1 + \|K_0\|_{\text{op}})(\|(P_0 - P) 1_J P_0\|_{\text{op}} + \|(P_0 - P) 1_J P\|_{\text{op}}).$$

If we now make the hypothesis that

$$(18) \qquad\qquad \|(P_0 - P) 1_J\|_{\text{op}} \le \frac{1}{4(1 + \|K_0\|_{\text{op}})}$$

then we have $\|E\|_{\text{op}} \le 1/2$, and so we have the Neumann series

$$(1 + E)^{-1} = 1 - E + E^2 - \dots.$$

In particular,

$$\|(1 + E)^{-1} - 1\|_{S^1} \le 2\|E\|_{S^1}.$$

To bound the right-hand side, we use the triangle inequality to obtain

$$\|E\|_{S^1} \le (1 + \|K_0\|_{\text{op}})(\|P_0 1_J P_0\|_{S^1} + \|P 1_J P\|_{S^1}).$$

Factorising $P_0 1_J P_0 = (1_J P_0)^*(1_J P_0)$ and similarly for $P 1_J P$, we conclude that

$$\|(1 + E)^{-1} - 1\|_{S^1} \le 2(1 + \|K_0\|_{\text{op}})(\|1_J P_0\|_{HS}^2 + \|1_J P\|_{HS}^2).$$

Note that E maps V to itself, and so $(1 + E)^{-1}$ can also be viewed as an operator from V to itself (being the identity on $V^\perp$). From (16) one then has

$$(P 1_{J^c} P)_V^{-1} = (1 + E)^{-1}(1 + P K_0 P)$$

and thus

$$\|(P 1_{J^c} P)_V^{-1} - (1 + P K_0 P)\|_{S^1} \le 2(1 + \|K_0\|_{\text{op}})^2(\|1_J P_0\|_{HS}^2 + \|1_J P\|_{HS}^2).$$

Applying (14), we conclude that

$$\|\tilde{P} - 1_{J^c}P(1 + PK_0P)P1_{J^c}\|_{S^1} \le 2(1 + \|K_0\|_{\mathrm{op}})^2(\|1_JP_0\|_{HS}^2 + \|1_JP\|_{HS}^2).$$

To deal with the PK_0P term we observe from (17) and the factorisation $P_0 1_J P_0 = (1_J P_0)^*(1_J P_0)$ that

$$\|K_0\|_{S^1} \le (1 + \|K_0\|_{\mathrm{op}})\|1_J P_0\|_{HS}^2,$$

and so

$$\|\tilde{P} - 1_{J^c}P1_{J^c}\|_{S^1} \le 3(1 + \|K_0\|_{\mathrm{op}})^2(\|1_JP_0\|_{HS}^2 + \|1_JP\|_{HS}^2).$$

We summarise the above discussion as a proposition:

Proposition 9 (Approximate description of $\tilde{P}$). *Let P be a projection to an n-dimensional subspace V of $L^2(\mathbb{R})$, and let J be a compact interval such that no non-trivial element of V is supported on J. Let P_0 be a projection to a (possibly infinite-dimensional) subspace V_0 of $L^2(\mathbb{R})$ which is locally trace class, and such that no non-trivial element of V_0 is supported on J. Let $K_0 : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be the compact operator solving* (15) *that is provided by the Fredholm alternative. Suppose that*

$$(19) \qquad \|(P_0 - P)1_J\|_{\mathrm{op}} \le \frac{1}{4(1 + \|K_0\|_{\mathrm{op}})}.$$

Let $\tilde{P}$ be the orthogonal projection to $1_{J^c}V$. Then

$$\|\tilde{P} - 1_{J^c}P1_{J^c}\|_{S^1} \le 3(1 + \|K_0\|_{\mathrm{op}})^2(\|1_JP_0\|_{HS}^2 + \|1_JP\|_{HS}^2).$$

Because the S^1 norm controls the trace, this proposition allows us to compare the quantities $\tilde{\mu}, \tilde{\sigma}^2$ from Corollary 8 with their counterparts μ, σ^2:

Corollary 10. *Let n, P, V, J, P_0, K_0 be as in Proposition 9 (in particular, we make the hypothesis* (19)*). Let I be a compact interval disjoint from J, and let $\mu, \sigma^2, \tilde{\mu}, \tilde{\sigma}^2$ be as in Corollary 8. Then we have*

$$\tilde{\mu} = \mu + O(M)$$

and

$$\tilde{\sigma}^2 = \sigma^2 + O(M)$$

where M is the quantity

$$M := (1 + \|K_0\|_{\mathrm{op}})^2(\|1_JP_0\|_{HS}^2 + \|1_JP\|_{HS}^2).$$

In practice, this corollary will allow us to show that the random variable $\#(\Sigma \cap I)$ is essentially independent of the event $\#(\Sigma \cap J) = 0$ for certain determinantal point processes Σ and disjoint intervals I, J.

4. Proof of main theorem

We are now ready to prove Theorem 3. We may of course assume that n is larger than any given absolute constant.

Let $n, M_n, \varepsilon, i, u$ be as in Theorem 3, and let X be the random variable

$$X := \frac{\lambda_{i+1}(M_n) - \lambda_i(M_n)}{1/(\sqrt{n}\rho_{\text{sc}}(u))}.$$

Clearly X takes values in $\mathbb{R}^+$ almost surely. Our task is to show that

$$\mathbf{P}(X \leq s) = \int_0^s p(y) \, dy + o(1)$$

for all fixed $s > 0$, or equivalently that

$$(20) \qquad\qquad \mathbf{P}(X > s) = \int_s^\infty p(y) \, dy + o(1)$$

for all fixed $s > 0$.

It will suffice to show that

$$(21) \qquad\qquad \mathbf{E}(X - s)_+ = \int_s^\infty (y - s)p(y) \, dy + o(1)$$

for all fixed $s > 0$, since on applying this with two choices $0 < s_1 < s_2$ of s, subtracting, and then dividing by $s_2 - s_1$ we see that

$$\mathbf{E}\min(\frac{(X - s_1)_+}{s_2 - s_1}, 1) = \int_{s_1}^\infty \min(\frac{y - s_1}{s_2 - s_1}, 1)p(y) \, dy + o(1);$$

letting s_1, s_2 approach a given value s from the left or right, we then conclude the bounds

$$\int_{s+\delta}^\infty p(y) \, dy - o(1) \leq \mathbf{P}(X > s) \leq \int_{s-\delta}^\infty p(y) \, dy + o(1)$$

for any fixed $\delta > 0$, and (20) follows from the monotone convergence theorem.

It remains to prove (21). By (4), the left-hand side of (21) can be written as

$$\det(1 - 1_{[0,s]}P_{\text{Sine}}1_{[0,s]}) + o(1).$$

Meanwhile, if we introduce the normalised random matrix

$$\tilde{M}_n := \frac{M_n - u\sqrt{n}}{1/\sqrt{n}\rho_{\text{sc}}(u)}$$

then we have

$$X = \lambda_{i+1}(\tilde{M}_n) - \lambda_i(\tilde{M}_n).$$

For any fixed choice of $\tilde{M}_n$, we observe the identity

$$(X - s)_+ = \int_\mathbb{R} 1_{N_{(-\infty,x)}(\tilde{M}_n)=i \wedge N_{[x,x+s]}(\tilde{M}_n)=0} \, dx,$$

since the set of real numbers x for which $N_{(-\infty,x)}(\tilde{M}_n) = i \wedge N_{[x,x+s]}(\tilde{M}_n) = 0$ holds is an interval of length $X - s$ when $X > s$, and empty otherwise. Taking expectations and using the Fubini-Tonelli theorem, we conclude that

$$\mathbf{E}(X - s)_+ = \int_{\mathbb{R}} \mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i \wedge N_{[x,x+s]}(\tilde{M}_n) = 0) \, dx.$$

Our task is thus to show that
(22)
$$\int_{\mathbb{R}} \mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i \wedge N_{[x,x+s]}(\tilde{M}_n) = 0) \, dx = \det(1 - 1_{[0,s]} P_{\mathrm{Sine}} 1_{[0,s]}) + o(1).$$

Let $t_n := \log^{0.6} n$ (say). We will shortly establish the following claims:

(i) (Tail estimate) We have

(23)
$$\int_{|x| \geq t_n} \mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i) \, dx = o(1).$$

(ii) (Approximate independence) For $|x| < t_n$, one has
(24)
$$\mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i \wedge N_{[x,x+s]}(\tilde{M}_n) = 0) = \mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i)\mathbf{P}(N_{[x,x+s]}(\tilde{M}_n) = 0) + O(\log^{-0.85} n).$$

(iii) (Gap probability at fixed energy) For $|x| < t_n$, one has

(25)
$$\mathbf{P}(N_{[x,x+s](\tilde{M}_n)=0}) = \det(1 - 1_{[0,s]} P_{\mathrm{Sine}} 1_{[0,s]}) + o(1).$$

(iv) (Central limit theorem) For $|x| < t_n$, one has

(26)
$$\mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} + O(\log^{-0.85} n)$$

where $\sigma := \sqrt{\log n / 2\pi^2}$.

Let us assume these estimates for the moment. From (24), (25), (26) one has

$$\mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i \wedge N_{[x,x+s]}(\tilde{M}_n) = 0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} (\det(1 - 1_{[0,s]} P_{\mathrm{Sine}} 1_{[0,s]}) + o(1)) + O(\log^{-0.85} n)$$

for $|x| \leq t_n$. Since

$$\int_{|x| \leq t_n} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} = 1 - o(1)$$

we conclude from the choice of t_n that

$$\int_{|x| \leq t_n} \mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i \wedge N_{[x,x+s]}(\tilde{M}_n) = 0) = \det(1 - 1_{[0,s]} P_{\mathrm{Sine}} 1_{[0,s]}) + o(1)$$

and the claim (22) then follows from (23).

It remains to establish the estimates (23), (24), (25), (26). We begin with (23). We can rewrite

$$N_{(-\infty,x)}(\tilde{M}_n) = N_{(-\infty,\sqrt{n}u + \frac{x}{\sqrt{n}\rho_{\mathrm{sc}}(u)})}(M_n).$$

From the rigidity of eigenvalues of GUE (see[8] e.g. [28, Corollary 5]) we know that

$$\mathbf{P}(N_{(-\infty,y)}(M_n) = i) \ll n^{-100}$$

(say) unless

$$y = \sqrt{n}u + O(\log^{O(1)} n/\sqrt{n}).$$

Because of this, to prove (23) we may restrict to the regime where $x = O(\log^{O(1)} n)$.

By[9] Lemma 5, for any real number y, $N_{(-\infty,y)}(M_n)$ is the sum of n independent Bernoulli variables. The mean and variance of such random variables was computed in [14]. Indeed, from [14, Lemma 2.1] one has (after adjusting the normalisation)

$$\mathbf{E}N_{(-\infty,y)}(M_n) = \int_{-\infty}^{y/\sqrt{n}} \rho_{\mathrm{sc}}(t) \, dt + O(\frac{\log n}{n})$$

while from [14, Lemma 2.3] one has

$$\mathbf{Var}N_{(-\infty,y)}(M_n) = (\frac{1}{2\pi^2} + o(1))\log n.$$

Renormalising (and using the hypothesis $x = O(\log^{O(1)} n)$), we conclude that

$$\mathbf{E}N_{(-\infty,x)}(\tilde{M}_n) = i + x + O(1)$$

and

$$\mathbf{Var}N_{(-\infty,x)}(\tilde{M}_n) = (\frac{1}{2\pi^2} + o(1))\log n.$$

Applying Bennet's inequality (see [3]), we conclude that

$$\mathbf{P}(\mathbf{E}N_{(-\infty,x)}(\tilde{M}_n) = i) \ll \exp(-cx/\sqrt{\log n})$$

for some absolute constant $c > 0$, which gives (23). The bound (26) follows from the same computations, using Lemma 7 (or Corollary 8) in place of Bennet's inequality.

The estimate (25) is well known (see[10] e.g. [1, Theorem 3.1.1]); for future reference we remark that this estimate also implies the crude lower bound

$$(27) \qquad\qquad\qquad \mathbf{P}(N_{[x,x+s](\tilde{M}_n)=0}) \gg 1$$

for n sufficiently large. We therefore turn to (24). By (27) and (26), it suffices to establish the conditional probability estimate

$$(28) \quad \mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i|N_{[x,x+s]}(\tilde{M}_n) = 0) = \frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2} + O(\log^{-0.85} n).$$

---

[8]One can also derive this rigidity from the Bennett's inequality argument given below. One could also use the rigidity results for more general Wigner matrices here, see [13] or [28], though this would be overkill.

[9]Strictly speaking, Lemma 5 is not applicable as stated because $(-\infty, y)$ is not a compact interval, but this can be addressed by the usual truncation argument, replacing $(-\infty, y)$ with $(-M, y)$ and then letting M go to infinity, exploiting the exponential decay of M_n. We omit the routine details.

[10]Strictly speaking, Theorem 3.1.1 of [1] only treats the case $u = 0$, but the general case $-2 + \varepsilon < u < 2 - \varepsilon$ follows from the same methods; see [1, Exercise 3.7.5].

We now turn to (28). Recall that the eigenvalues of M_n form a determinantal point process with kernel $K^{(n)}$ given by (1). Rescaling this, we see that the eigenvalues of $\tilde{M}_n$ form a determinantal point process with kernel $\tilde{K}^{(n)}$ given by the formula

$$\tilde{K}^{(n)}(x,y) := \frac{1}{\rho_{\rm sc}(u)\sqrt{n}} K^{(n)}\left(u\sqrt{n} + \frac{x}{\rho_{\rm sc}(u)\sqrt{n}}, u\sqrt{n} + \frac{y}{\rho_{\rm sc}(u)\sqrt{n}}\right).$$

This is the kernel of an orthogonal projection $\tilde{P}^{(n)}$ to some n-dimensional subspace $\tilde{V}^{(n)}$ in $L^2(\mathbb{R})$. The elements of this subspace consist of polynomial multiples of a gaussian function, and in particular there is no non-trivial element of $\tilde{V}^{(n)}$ that vanishes on $[x, x+s]$. Applying Corollary 8 (and a truncation argument to deal with the non-compact nature of $(-\infty, x)$), one has

$$\mathbf{P}(N_{(-\infty,x)}(\tilde{M}_n) = i | N_{[x,x+s]}(\tilde{M}_n) = 0) = \frac{1}{\sqrt{2\pi}\sigma'} e^{-(i-\mu')^2/2(\sigma')^2} + O((\sigma')^{-1.7})$$

where

$$\mu' := \operatorname{trace}(P'1_{(-\infty,x)})$$

and

$$(\sigma')^2 := \operatorname{trace}(P'1_{(-\infty,x)^c}P'1_{(-\infty,x)}),$$

and P' is the orthogonal projection to $1_{[x,x+s]^c}\tilde{V}^{(n)}$. To establish (28), it will thus suffice to establish the bounds

$$\mu' = O(1)$$

and

$$(\sigma')^2 = \sigma^2 + O(1).$$

To do this, we will use Corollary 10, with $J := [x, x+s]$, and the role of P_0 being played by the Dyson projection $P_{\rm Sine}$. From the well-known fact that a non-trivial function and its Fourier transform cannot both be compactly supported, we see that there is no non-trivial function in the range of $P_{\rm Sine}$ supported in J. As $P_{\rm Sine}$ is locally trace class, we conclude from the Fredholm alternative (see e.g. [22, Theorem VI.14]) that the compact operator K_0 defined by (15) exists. As K_0 is independent of n, we certainly have[11]

$$\|K_0\|_{op} \ll 1$$

and similarly

(29) $$\|1_J P_{\rm Sine}\|_{HS} \ll 1.$$

By Corollary 10 (once again using a truncation argument to deal with the half-infinite nature of $(-\infty, x)$), it will thus suffice to show that

(30) $$\|1_J \tilde{P}^{(n)}\|_{HS} \ll 1$$

and

(31) $$\|(P_{\rm Sine} - \tilde{P}^{(n)})1_J\|_{\rm op} = o(1).$$

---

[11]Note that our bound here on $\|K_0\|_{op}$ is *ineffective*, as it relies on the Fredholm alternative. However, it is quite probable that one can obtain an effective bound on K_0 here by using a quantitative versions of Hardy's uncertainty principle to give a more robust version of the assertion that a non-trivial function and its Fourier transform cannot both be compactly supported. We will not pursue this issue here.

Since the Hilbert-Schmidt norm controls the operator norm, we see from (29) that (30), (31) will both follow from the bound

$$\|(P_{\text{Sine}} - \tilde{P}^{(n)})1_J\|_{HS} = o(1).$$

Using the integral kernels $K_{\text{Sine}}, \tilde{K}^{(n)}$ of $P_{\text{Sine}}, \tilde{P}^{(n)}$ and the compact nature of J, it suffices to show that

$$(32) \qquad \int_{\mathbb{R}} |K_{\text{Sine}}(x,y) - \tilde{K}^{(n)}(x,y)|^2 \ dx = o(1)$$

uniformly for all $y \in J$. In principle one could establish this bound from a sufficiently precise analysis of the asymptotics of Hermite polynomials (such as those given in [8]), but one can actually derive this bound from the standard convergence result (2) as follows. From (2) we know that $\tilde{K}^{(n)}(x,y)$ converges locally uniformly in x, y to $K_{\text{Sine}}(x,y)$ as $n \to \infty$, and so

$$(33) \qquad \int_{-L}^{L} |K_{\text{Sine}}(x,y) - \tilde{K}^{(n)}(x,y)|^2 \ dx = o(1)$$

for any fixed L. Also, as $P_{\text{Sine}}, \tilde{P}^{(n)}$ are both projections, one has

$$\int_{\mathbb{R}} |K_{\text{Sine}}(x,y)|^2 \ dx = K_{\text{Sine}}(y,y)$$

and

$$\int_{\mathbb{R}} |\tilde{K}^{(n)}(x,y)|^2 \ dx = \tilde{K}^{(n)}(y,y).$$

From (2), one has

$$\tilde{K}^{(n)}(y,y) = K_{\text{Sine}}(y,y) + o(1).$$

For any given $\varepsilon > 0$, one can find an L such that

$$(34) \qquad \int_{|x|>L} |K_{\text{Sine}}(x,y)|^2 \ dx = O(\varepsilon)$$

and thus

$$\int_{\mathbb{R}} |\tilde{K}^{(n)}(x,y)|^2 \ dx = \int_{-L}^{L} |K_{\text{Sine}}(x,y)|^2 \ dx + O(\varepsilon) + o(1).$$

But from (33) and the triangle inequality we have

$$\int_{-L}^{L} |\tilde{K}^{(n)}(x,y)|^2 \ dx = \int_{-L}^{L} |K_{\text{Sine}}(x,y)|^2 \ dx + o(1)$$

and so

$$\int_{|x|>L} |\tilde{K}^{(n)}(x,y)|^2 \ dx = O(\varepsilon) + o(1).$$

From this, (33), and (34) we conclude that

$$\int_{\mathbb{R}} |K_{\text{Sine}}(x,y) - \tilde{K}^{(n)}(x,y)|^2 \ dx = O(\varepsilon) + o(1)$$

and the claim (32) follows by sending ε to zero. The proof of Theorem 3 (and thus also Corollary 4) is now complete.

References

[1] G. Anderson, A. Guionnet and O. Zeitouni, An introduction to random matrices, Cambridge Studies in Advanced Mathematics, 118. Cambridge University Press, Cambridge, 2010.

[2] G. Ben Arous, P. Bourgade, Extreme gaps between eigenvalues of random matrices, *arXiv:1010.1294*

[3] G. Bennett, Probability Inequalities for the Sum of Independent Random Variables, *Journal of the American Statistical Association* **57** (1962), 33-45.

[4] R. Bhatia, Matrix Analysis, Springer-Verlag, New York 1997.

[5] P. Bourgade, L. Erdős, H. T. Yau, Universality of General β-Ensembles, *arXiv:1104.2272*

[6] O. Costin and J. Lebowitz, Gaussian fluctuations in random matrices, *Phys. Rev. Lett.* 75 (1) (1995) 69–72.

[7] P. Deift, T. Kriecherbauer, K.T.-R. McLaughlin, S. Venakides and X. Zhou, Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory, *Comm. Pure Appl. Math.* **52** (1999), no. 11, 1335–1425.

[8] B. Delyon, J. Yao, On the spectral distribution of Gaussian random matrices, Acta Mathematicae Sinica, English Series Vol. 22, No. 2 (2006) 297–312.

[9] F. Dyson, The Threefold Way. Algebraic Structure of Symmetry Groups and Ensembles in Quantum Mechanics, *J. Math. Phys.* **3** (1962), 1199–1215.

[10] L. Erdős, S. Peche, J. Ramirez, B. Schlein and H.-T. Yau, Bulk universality for Wigner matrices, *arXiv:0905.4176*

[11] L. Erdős, J. Ramirez, B. Schlein, T. Tao, V. Vu, and H.-T. Yau, Bulk universality for Wigner hermitian matrices with subexponential decay, *arxiv:0906.4400, To appear in Math. Research Letters.*

[12] L. Erdős, B. Schlein and H.-T. Yau, Universality of Random Matrices and Local Relaxation Flow, *arXiv:0907.5605*

[13] L. Erdős, H.-T.Yau, and J. Yin, Rigidity of Eigenvalues of Generalized Wigner Matrices. *arXiv:1007.4652*

[14] J. Gustavsson, Gaussian fluctuations of eigenvalues in the GUE, *Ann. Inst. H. Poincaré Probab. Statist.* 41 (2005), no. 2, 151–178.

[15] M. Jimbo, T. Miwa, Tetsuji, Y. Mori and M. Sato, Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent, *Phys. D 1.* (1980), no. 1, 80–158.

[16] K. Johansson, Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices, *Comm. Math. Phys.* 215 (2001), no. 3, 683–705.

[17] A. Knowles, J. Yin, Eigenvector Distribution of Wigner Matrices, *arXiv:1102.0057*

[18] O. Macchi, The coincidence approach to stochastic point processes, *Adv. Appl. Probab.*, **7** (1975), 83-122.

[19] M.L. Mehta, Random Matrices and the Statistical Theory of Energy Levels, Academic Press, New York, NY, 1967.

[20] J. Hough, M. Krishnapur, Y. Peres, B. Virág, Determinantal processes and independence, *Probab. Surv.* **3** (2006), 206-229.

[21] R. Lyons, Determinantal probability measures, *Publ. Math. Inst. Hautes Études Sci.* **98**, 167-212.

[22] M. Reed, B. Simon, Methods of modern mathematical physics. I. Functional analysis. Second edition. Academic Press, Inc. (Harcourt Brace Jovanovich, Publishers), New York, 1980.

[23] A. Soshnikov, Determinantal random point fields, *Russian Math. Surveys* **55** (2000), 923975.

[24] A. Soshnikov, Gaussian limit for determinantal random point fields, *Ann. Probab.* 30 (2002), no. 1, 171187.

[25] T. Tao and V. Vu, Random matrices: Universality of the local eigenvalue statistics, *Acta Mathematica* 206 (2011), 127–204.

[26] T. Tao, V. Vu, Random covariance matrices: university of local statistics of eigenvalues, *to appear in Annals of Probability.*

[27] T. Tao and V. Vu, Random matrices: Universal properties of eigenvectors, *to appear in Random Matrices: Theory and Applications.*

[28] T. Tao and V. Vu, Random matrices: Sharp concentration of eigenvalues, *preprint* http://arxiv.org/pdf/1201.4789.pdf.

[29] C. Tracy and H. Widom, On orthogonal and symplectic matrix ensembles, *Commun. Math. Phys.* 177 (1996) 727–754.

[30] C. A. Tracy and H. Widom, Introduction to random matrices, in Geometric and Quantum Aspects of Integrable Systems, G. F. Helminck, ed., Lecture Notes in Physics, Vol. 424, Springer-Verlag (Berlin), 1993, pgs. 103-130.

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

E-mail address: tao@math.ucla.edu