THE UNIVERSALITY PHENOMENON IN RANDOM MATRIX THEORY
(AND BEYOND)

NICHOLAS A. COOK

Abstract. These are lecture notes for a graduate minicourse (9 lectures) given at Duke in the fall of 2022. No prior exposure to random matrix theory is assumed. Our main objective is to prove a version of the Tao–Vu four moment theorem for the correlation functions of eigenvalues of Wigner matrices at the local spectral scale. Stronger results for Wigner matrices due to Erdős, Yau and coauthors obtained using Dyson Brownian motion will be surveyed (though we won’t have time to go into the details in this minicourse). Finally, we review the Montgomery–Dyson pair correlation conjecture (and more general GUE hypothesis) for zeros of the Riemann zeta function.

Contents

1. Introduction and overview 1
   1.1. Notation 1
   1.2. Random matrix ensembles 2
   1.3. The Gaussian ensembles 4
   1.4. Universality 4
   2. Results on the spectrum of Wigner matrices – some highlights 5
      2.1. Scale of the spectrum? 5
      2.2. Limiting support and empirical distribution? (after rescaling) 6
      2.3. Outlier eigenvalues? 6
      2.4. Fluctuations of linear statistics? 7
      2.5. Correlation functions at local scale 8
   3. A first look at the Lindeberg swapping method: the CLT 10
   4. Lindeberg swapping for Wigner matrices 11
      4.1. The Stieltjes transform method and resolvents 11
      4.2. The Lindeberg method in RMT 13
      4.3. A two moment theorem at the global spectral scale 13
      4.4. Four moment theorems at the local spectral scale 16
   5. The local semicircle law and eigenvector delocalization 20
   6. The Montgomery–Dyson pair correlation conjecture 20

References 20

1. Introduction and overview

1.1. Notation. All random variables are supported on some background probability space \((\Omega, \mathcal{F}, \mathbb{P})\), and \(\mathbb{E}\) denotes expectation under \(\mathbb{P}\).

Date: November 22, 2022.
The Euclidean $\ell_p$-norms on $\mathbb{R}^n$ or $\mathbb{C}^n$ is denoted $\| \cdot \|_p$. The standard basis vectors for $\mathbb{C}^n$ are denoted $e_1, \ldots, e_n$. For an $n \times n$ matrix $M$ with complex entries we write
\[ \|M\|_{\text{op}} = \sup_{u \in S^{n-1}} \|Mu\|_2 \]
for the $\ell_2 \to \ell_2$ operator norm, and
\[ \|M\|_{\text{HS}} = \left( \sum_{i,j=1}^n |M_{ij}|^2 \right)^{1/2} \]
for the Hilbert–Schmidt (a.k.a. Frobenius) norm. When the entries, rows or columns of a matrix $M$ haven’t been assigned symbols we may write $M(i,j)$ or $M_{ij}$ for its entries, and $\text{row}_i(M), \text{col}_j(M)$ for its $i$th row and $j$th column.

The eigenvalues of an $n \times n$ Hermitian matrix $M$ are denoted $\lambda_1(M), \ldots, \lambda_n(M)$, and we write $u_1(M), \ldots, u_n(M)$ for an associated orthonormal basis of eigenvectors (while these are only determined up to arbitrary unit phase factors, we’ll mostly work with the uniquely determined spectral projections $u_i(M)u_i(M)^*$).

We write $\mathbb{H} = \{ z \in \mathbb{C} : \text{Im}(z) > 0 \}$ for the open upper half-plane. We use $|S|$ to denote the cardinality of a finite set $S$; we also write $|I|$ for the length of an interval $I \subset \mathbb{R}$ (there should be no risk of confusion between these two uses of $| \cdot |$).

We use standard asymptotic notation: for $f \in \mathbb{C}$ and $g > 0$ we write $f = O(g)$, $f \lesssim g$ to mean $|f| \leq Cg$ for some finite constant $C$, which is understood to be absolute unless dependence on parameters is indicated (we sometimes indicate dependence of the implicit constant on a parameter $p$ by writing $f = O_p(g), f \lesssim_p g$, etc.). We may write $f \asymp g$ to mean $g \lesssim f \lesssim g$. For $f, g$ depending on $n$ we write $f = o(g), f/g \to 0$ as $n \to \infty$, and $f \sim g$ for $f/g \to 1$, where the rates may depend on fixed parameters. Sometimes we informally write $f \ll g$ to mean $f$ is “much smaller” than $g$ (e.g. $f = o(g)$), but this will be in purely informal contexts (this is just to emphasize that, unlike in analytic number theory, it is not synonymous with $f = O(g)$).

1.2. Random matrix ensembles. Random matrix theory (RMT) is concerned with the distribution of eigenvalues and eigenvectors for large random matrices. In this course we will focus on square Hermitian random matrices, with the dimension $n$ either large or tending to infinity. In 1955 Wigner introduced a natural random matrix model of this type [Wig55]. His motivation was to understand the energy levels for heavy nuclei. In quantum mechanics these are the eigenvalues of the Hamiltonian operator, but for heavy nuclei an analytic formula for the energy levels is inaccessible. Wigner’s idea was to model the Hamiltonian by a large (or infinite) random matrix.

In this minicourse we will consider matrices of the following type:

**Definition 1.1** (Wigner random matrix). An $n \times n$ (complex) Wigner matrix $W$ is a Hermitian matrix with independent centered complex entries $(\xi_{ij})_{1 \leq i,j \leq n}$ on and above the diagonal, with $E[|\xi_{11}|^2] = 1$ and $E[\xi_{11}^2] < \infty$. A real Wigner matrix is a Wigner matrix with real entries. An iid (real or complex) Wigner matrix is a Wigner matrix with $\xi_{ij} \overset{d}{=} \xi_{12}$ for all $1 \leq i < j \leq n$, and $\xi_{ii} \overset{d}{=} 1$ for all $1 \leq i \leq n$ – in this case we refer to the variables $\xi_{11}, \xi_{12}$ as the atom variables.

\[ \overset{d}{=} \text{ denotes equality in distribution.} \]

1 Wigner’s work was preceded by a few decades by work of Wishart on random sample covariance matrices, motivated by problems in statistics. We will not consider sample covariance matrices in this course, though many of the results we’ll see have analogues for those matrices.
We’ll often work with the normalized Wigner matrix $H := \frac{1}{\sqrt{n}}W$. (As we’ll see below, under this scaling most of the spectrum is at scale 1 with high probability.)

We’ll sometimes consider a sequence $(W_n)_{n \geq 1}$ with $W_n$ an $n \times n$ Wigner matrix (though many of our main results will be stated for $n$ sufficiently large and fixed, with quantitative errors). In general we allow the distributions of entries of iid Wigner matrices to change with $n$, and so we may sometimes write $\xi_{ij}^{(n)}$ for the entries of $W_n$. However, we’ll often take the distributions of atom variables to be fixed independent of $n$. (Technically we should still write $\xi_{ij}^{(n)}$ as we make no assumption that the $W_n$ are coupled across different values of $n$ – e.g. if each $W_n$ is the top-left $n \times n$ submatrix of an infinite Wigner matrix – but we’ll tend to abusively drop the $n$ dependence from the notation and write $\xi_{ij}$.)

In this course we are interested in the distribution of the eigenvalues of $W_n$ at various scales. (There is also extensive work on the eigenvectors of Wigner matrices, but we will not have time to discuss those results in this course.)

To streamline the presentation we will mainly focus on real Wigner matrices in this course. It will sometimes be convenient to put additional tail hypotheses on the atom variable. A particularly convenient one is the following:

**Definition 1.2** (Sub-Gaussian variable). For $K > 0$, a real random variable $\xi$ is $K$-sub-Gaussian if

$$\mathbb{E} \exp(\xi^2/K^2) \leq 2.$$  

The sub-Gaussian constant $K_\xi$ of $\xi$ is the smallest $K$ for which the above holds.

**Exercise 1.3** (Equivalent characterizations of sub-Gaussian variables).

(a) Show that if a random variable $\xi$ is $K$-sub-Gaussian, then

$$\mathbb{P}(|\xi| \geq t) \leq 2 \exp(-t^2/K^2) \quad \forall \ t \geq 0$$  

and conversely, that if $\xi$ satisfies (1.1), then $\xi$ is $CK$-sub-Gaussian for a universal constant $C > 0$.

(b) Similarly show that the $K$-sub-Gaussian condition is equivalent to assuming the moments condition

$$\left(\mathbb{E} |\xi|^m\right)^{1/m} \leq K \sqrt{m}$$  

for all integers $m \geq 1$, up to modification of $K$ by an absolute constant factor.

Important examples of sub-Gaussian random variables are Gaussian variables and Rademacher variables (uniformly distributed in $\{-1, 1\}$).

The sub-Gaussian assumption for many of our results can be relaxed via truncation arguments, but we will not have time to go into such refinements in this minicourse – we refer to the texts [AGZ10, Tao12] for examples.

**1.2.1. Other types of random matrices.** There is also important work on singular values/vectors of rectangular $p \times n$ random matrices (such as the Wishart ensemble) with $p = p(n) \sim \alpha n$ for some fixed $\alpha \in \mathbb{R}$. Another restriction we will make is to focus on symmetric and Hermitian ensembles, in particular the Wigner matrices; we mention there is also a large body of work on non-Hermitian random matrices, such as random unitary matrices and permutation matrices, as well as random non-normal matrices, such as an $n \times n$ matrix with iid entries (perhaps the most natural sort of random matrix to consider). We refer to the texts [AGZ10, Tao12] for more on non-Hermitian and non-square random matrices.
1.3. The Gaussian ensembles. We have the following important types of Wigner matrices.

**Definition 1.4 (Gaussian Unitary/Orthogonal Ensemble).** A GOE matrix is an iid Wigner matrix with $\xi_{12} \sim N_{\mathbb{R}}(0,1)$ and $\xi_{11} \sim N_{\mathbb{R}}(0,2)$ (so $\xi_{11} \overset{d}{=} \sqrt{2}\xi_{12}$). A GUE matrix is an iid Wigner matrix with $\xi_{12} \sim N_{\mathbb{C}}(0,1)$ and $\xi_{11} \sim N_{\mathbb{R}}(0,2)$.

(Recall that the standard complex Gaussian distribution $N_{\mathbb{C}}(0,1)$ has independent real and imaginary parts with distribution $N_{\mathbb{R}}(0,1/2)$.)

The names come from the fact that the distribution of a GUE (resp. GOE) matrix is invariant under conjugation by fixed unitary (resp. orthogonal) matrices.

One can derive the explicit joint densities of eigenvalues. For the GUE the density at $\lambda = (\lambda_1, \ldots, \lambda_n)$ is given by

$$\rho_{n,2}(\lambda) = \frac{1}{Z_{n,2}} |\Delta(\lambda)|^2 e^{-||\lambda||^2/2} 1_{\lambda_1 \geq \cdots \geq \lambda_n} \quad (1.3)$$

and for the GOE by

$$\rho_{n,1}(\lambda) = \frac{1}{Z_{n,1}} |\Delta(\lambda)| e^{-||\lambda||^2/4} 1_{\lambda_1 \geq \cdots \geq \lambda_n} \quad (1.4)$$

where $Z_{n,\beta}$, $\beta \in \{1, 2\}$ are normalizing constants (partition functions) and we recall the Vandermonde determinant is defined

$$\Delta(x_1, \ldots, x_n) := \det[(x_i^{-1})_{1 \leq i,j \leq n}] = \prod_{i<j} (x_i - x_j). \quad (1.5)$$

We can alternatively write

$$\rho_{n,\beta}(\lambda) = \frac{1}{Z_{n,\beta}} e^{-\beta h_n(\lambda)} 1_{\lambda_1 \geq \cdots \geq \lambda_n} \quad (1.6)$$

where we define

$$h_n(\lambda) := \frac{1}{4} \sum_{i=1}^n \lambda_i^2 - \sum_{1 \leq i < j \leq n} \log |\lambda_i - \lambda_j|. \quad (1.7)$$

This takes the form of the Gibbs measure and Hamiltonian for a one-dimensional log-gas (or planar Coulomb gas restricted to a line) in a quadratic potential at inverse temperature $\beta$.

There is another Gaussian ensemble, the Gaussian Symplectic Ensemble, for which the joint density of eigenvalues is given by (1.6) with $\beta = 4$, but we will not discuss such matrices. We also mention that there is a one-parameter family of tridiagonal random matrix models $T_{n,\beta}, \beta > 0$ (random Hermitian matrices but not of Wigner type) whose joint density of eigenvalues is given by (1.6) – see [DE02].

One important feature to note in (1.3), (1.4) is that due to the Vandermonde factor, eigenvalues “repel” each other – that is, are less and less likely to be close to one another – and this repulsion is stronger for the GUE ensemble.

1.4. Universality. Using the explicit joint densities of eigenvalues (1.3), (1.4), the asymptotic distribution of various spectral statistics can be worked out explicitly (though this is often challenging, involving a range of beautiful ideas that we don’t have time to discuss in this course). But what can be said for general Wigner matrices?

It turns out that results for GUE/GOE matrices often carry over to complex/real Wigner matrices, as part of a general
Universality phenomenon: “Smooth” spectral statistics of iid Wigner matrices asymptotically only depend on the first few (usually 2 or 4) moments of the atom variables $\xi_{11}, \xi_{12}$.

We will spend the better part of the course proving a few instances of the universality phenomenon for Wigner matrices.

2. Results on the spectrum of Wigner matrices – some highlights

Here we present some natural questions and answers about the asymptotic behavior of the (random) spectrum of iid Wigner matrices. We note that many of these results were first established for the Gaussian ensembles by explicit methods.

We’ll be somewhat loose with the tail assumptions on the atom variables $\xi_{11}, \xi_{12}$.

2.1. Scale of the spectrum? We first note a couple of easy bounds on (parts of) the spectrum of a Wigner matrix.

The first is an easy bound for the operator norm:

$$\|W_n\|_{op} = \sup_{u \in \mathbb{S}^{n-1}} \|W_n u\|_2 = \max\{\lambda_1(W_n), |\lambda_n(W_n)|\}. \quad (2.1)$$

Indeed, by taking $u$ to be a standard basis vector, we see the operator norm is bounded below by the norm of any column. We easily compute

$$\mathbb{E}\|\text{col}_1(W_n)\|_2^2 = \sum_{i=1}^{n} \mathbb{E}|\xi_{i1}|^2 = n + O(1).$$

Moreover, since $\|\text{col}_1(W_n)\|_2^2$ is a sum of independent variables, it will concentrate around its expectation (at least if we make some higher moment assumptions on the entries). Thus we have

$$\max\{\lambda_1(W_n), |\lambda_n(W_n)|\} \geq \|\text{col}_1(W_n)\|_2 \gtrsim \sqrt{n} \quad (2.2)$$

with high probability.

It turns out $\sqrt{n}$ is the correct scale for most of the spectrum of $W_n$. To see an upper bound, at least for most eigenvalues, we can compute the second spectral moment. We have the following:

**Definition 2.1.** The empirical spectral distribution (ESD) of an $n \times n$ matrix $M$ with complex entries having eigenvalues $\lambda_1(M), \ldots, \lambda_n(M) \in \mathbb{C}$ (counted with multiplicity) is the probability measure on $\mathbb{C}$ given by

$$\mu_M := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(M)}.$$
We can compute the expected second moment of the ESD of a Wigner matrix exactly:

\[
\mathbb{E} \int t^2 d\mu_{W_n}(t) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i(W_n)^2
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \lambda_i(W_n^2)
\]

\[
= \frac{1}{n} \text{Tr} W_n^2
\]

\[
= \frac{1}{n} \sum_{i,j=1}^{n} E\xi_{ij}^2 = n + O(1).
\]

From Markov’s inequality we get that for any small \( \varepsilon > 0 \), the second moment of \( \mu_{H_n} \) is at most \( 1/\varepsilon \) with probability \( 1 - O(\varepsilon) \). From Chebyshev’s inequality we then conclude that with probability at least \( 1 - O(\varepsilon) \), for any (large) \( K > 0 \) we have that

\[
|\lambda_i(W_n)| \leq K \sqrt{n}
\]

for all but at most \( n/(\varepsilon K^2) \) values of \( i \in [n] \). In particular, with \( K = 1/\varepsilon \) we have that with probability \( 1 - O(\varepsilon) \) a proportion \( 1 - O(\varepsilon) \) of the spectrum is of size \( O(\varepsilon^{-1/2}) \).

2.2. Limiting support and empirical distribution? (after rescaling). We let \( \mu_{sc} \) denote the probability measure on \( \mathbb{R} \) with support \([−2, 2]\) having density

\[
\mu_{sc}(dx) := \rho_{sc}(x)dx := \frac{1}{2\pi} \sqrt{4 - x^2} dx, \quad x \in [-2, 2]
\]

with respect to Lebesgue measure. Recall that \( H_n := \frac{1}{\sqrt{n}} W_n \).

Theorem 2.2 (Wigner’s semicircle law). For each \( n \in \mathbb{N} \) let \( W_n \) be an iid Wigner matrix such that the distributions of the atom variables \( \xi_{12}, \xi_{11} \) are fixed independent of \( n \). Then the random ESDs \( \mu_{H_n} \) converge in probability to \( \mu_{sc} \). That is, for any fixed test function \( f \in C_c(\mathbb{R}) \) and \( \varepsilon > 0 \),

\[
\mathbb{P}(|\mu_{H_n}(f) - \mu_{sc}(f)| > \varepsilon) \to 0 \quad \text{as } n \to \infty.
\]

(We tend to write \( \mu(f) := \int_{\mathbb{R}} f d\mu \).) We’ll see more quantitative versions of the semicircle law later.

The assumption that the distribution of atom variables are fixed independent of \( n \) can be significantly relaxed, but some control on the growth of a higher moment is needed, as shown in the following exercise.

Exercise 2.3. Show that Theorem 2.2 is false in general if the distributions of atom variables are allowed to vary with \( n \) (while still satisfying the first and second moment conditions of Definition 1.1). (Hint: one may consider for instance an atom variable that is only nonzero with probability \( p(n) \) decaying rapidly with \( n \).)

2.3. Outlier eigenvalues? Note that Theorem 2.2 permits \( o(n) \) eigenvalues of \( H_n \) to stay bounded away from the limiting support, or even escape to infinity. In [FK81] it was shown that under higher moment assumptions the edge eigenvalues \( \lambda_1(H_n), \lambda_n(H_n) \) converge to the edges \( \pm 2 \) of the limiting support. The hypotheses were relaxed to the sharp finite fourth moment hypothesis in [BY88].
Theorem 2.4 (Füredi–Komlós ’81, Bai–Yin ’88). With $W_n$ as in Theorem 2.2, assume further that the atom variables have finite fourth moment. Then $\lambda_1(H_n) \to 2$ and $\lambda_n(H_n) \to -2$ in probability.

Remark 2.5. The above result can be combined with Talagrand’s concentration of measure inequality to show

$$P(|\lambda_1(H_n) - 2| > \varepsilon) \lesssim \varepsilon \exp(-c\varepsilon^2 n)$$

and similarly for $\lambda_n(H_n)$, at least when the entries of $W_n$ are a.s. bounded (such as Rademacher variables). Slightly degraded tail bounds can be established for matrices with entries having light enough tails by truncation arguments. See [Tao12].

Exercise 2.6. Show that the finite fourth moment hypothesis is sharp. (Hint: the operator norm $\|W_n\|_{op}$ is bounded below by the modulus $|\xi_{ij}|$ of any entry.)

2.4. Fluctuations of linear statistics? Theorem 2.2 gives a law of large numbers for linear statistics of the ESDs. What is the scale of fluctuations? And after appropriate rescaling, do we get a CLT?

One of the key surprises of RMT is that while generic, unnormalized linear statistics

$$\text{Tr } f(H_n) = \sum_{i=1}^n f(\lambda_i(H_n))$$

are asymptotically of size $n$, the variance of $\text{Tr } f(H_n)$ remains bounded, at least if $f$ is sufficiently smooth. Moreover, the recentered statistics

$$\text{Tr } f(H_n) - n \int f \, d\mu_{H_n}$$

satisfy a CLT. One should compare to the situation when $\lambda_i(H_n)$ are iid samples from a probability measure on $\mathbb{R}$ (such as the semicircle law $\mu_{sc}$, in which case the empirical measures would again converge weakly to $\mu_{sc}$) – in this case the order of fluctuations would be $\sqrt{n}$.

Theorem 2.7 (CLT for smooth linear statistics). Let $W_n$ be an iid Wigner matrix with uniformly (in $n$) sub-Gaussian entries, let $\varepsilon > 0$ be arbitrary, and fix $f \in H^{1/2+\varepsilon}(\mathbb{R})$. Then $\text{Tr } f(H_n) - n\mu_{sc}(f)$ converges to a centered Gaussian with an explicit (finite) variance depending on $f$ and the 4th moment of the atom variables.

The regularity assumption for $f$ is essentially sharp, as seen from the next theorem. The above result was shown very recently in [LS], building on earlier works making higher regularity assumptions; see [Shc11, SW13] and references therein.

On the other hand, when $f$ is the indicator of an interval (which just barely fails to be in $H^{1/2}$) the variance of $\text{Tr } f(H_n)$ grows logarithmically (which is still much smaller than the scale $\sqrt{n}$ for iid samples). The following was proved for the GUE case in [Gus05] and extended to complex Wigner matrices with 4 moments matching the GUE in [DV11].

Theorem 2.8 (CLT for eigenvalue counts). Let $W_n$ be a complex iid Wigner matrix such that the real and imaginary parts of the atom variables $\xi_{1,1}, \xi_{1,2}$ have matching moments up to order 4 with the the GUE matrix. Fix $E \in (-2, 2)$ and let $f = 1_{(-\infty, E]}$. Then

$$\frac{\text{Tr } f(H_n) - n\mu_{sc}(f)}{\sqrt{\frac{1}{2\pi^2} \log n}} \to N(0, 1)$$

in distribution.
The fact that the variance of linear statistics is so small indicates that the eigenvalues of Wigner matrices are very rigid down to fine scales. Spectral rigidity in turn is often viewed as a consequence of eigenvalue repulsion.

2.5. Correlation functions at local scale. The \( n \) (possibly repeated) eigenvalues \( \lambda_1(H_n) \geq \cdots \geq \lambda_n(H_n) \) of a normalized Wigner matrix can be viewed as a random point process on \( \mathbb{R} \). (Technically the point process is the measure \( n \mu_{H_n} = \sum_{i=1}^n \delta_{\lambda_i(H_n)} \).) From the semicircle law we expect that the typical order of spacings between consecutive eigenvalues will be of order \( 1/n \). Quantitative versions of the semicircle law (called local semicircle laws) show that the number of eigenvalues in intervals \( I \) of length \( \gg 1/n \) are asymptotically deterministic and given by \( n \mu_{sc}(I) \) – see [BGK17] for a survey.

Some terminology: we refer to the scale \( \sim 1 \) as the global scale of the entire spectrum of \( H_n \), the scale \( \sim 1/n \) of typical eigenvalue spacings as the local scale, and intermediate scales \( \ll 1 \) and \( \gg 1/n \) as mesoscopic scales. So the local semicircle law is really a quantitative semicircle law covering essentially all mesoscopic scales.

The \( k \)-point correlation function (also called the \( k \)-point joint intensity function) \( \rho_{H_n}^{(k)} \) on \( \mathbb{R}^k \) of the point process \( n \mu_{H_n} \) is defined via duality by requiring

\[
\mathbb{E} \sum_{1 \leq i_1, \ldots, i_k \leq n} \phi(\lambda_{i_1}(H_n), \ldots, \lambda_{i_k}(H_n)) = \int_{\mathbb{R}^k} \rho_{H_n}^{(k)}(x_1, \ldots, x_k) \phi(x_1, \ldots, x_k) dx_1 \cdots dx_k
\]

for all \( \phi \in C_c^\infty(\mathbb{R}^k \to \mathbb{C}) \). If the distribution of \( H_n \) is absolutely continuous (as is the case for the Gaussian ensembles) then the correlation functions are locally integrable functions, but in general they are only defined in the sense of distributions (in fact the above defines them as a Radon measure on \( \mathbb{R}^k \)). If \( H_n \) has a smooth density then one can alternatively define \( \rho_{H_n}^{(k)} \) as the smooth function such that the probability that there is an eigenvalue of \( H_n \) in each of the intervals \( (x_1 - \varepsilon, x_1 + \varepsilon), \cdots, (x_k - \varepsilon, x_k + \varepsilon) \) is \( \sim (2\varepsilon)^k \rho_{H_n}^{(k)}(x_1, \ldots, x_k) \) as \( \varepsilon \to 0 \) with \( n \) fixed.

The one-point correlation function is just a rescaled expected empirical spectral distribution:

\[
\mathbb{E} \mu_{H_n}(\phi) = \frac{1}{n} \mathbb{E} \sum_{i=1}^n \phi(\lambda_i(H_n)) = \frac{1}{n} \int \rho_{H_n}^{(1)}(x) \phi(x) dx.
\]

From the semicircle law we get that \( \rho_{H_n}^{(1)}(x) \sim n \rho_{sc}(x) \) for fixed \( x \in \mathbb{R} \). Since \( \rho_{sc} \) is continuous, by an approximation argument we can deduce that for a fixed interval \( I \subset \mathbb{R} \), defining the eigenvalue counting function

\[
N_{H_n}(I) := n \mu_{H_n}(I)
\]

we have

\[
\mathbb{E} N_{H_n}(I) \sim \int_I \rho_{n,1}(x).
\]

Similarly if \( \frac{1}{n^2} \rho_{H_n}^{(2)} \) converges to a continuous density, then we can show

\[
\text{Var} N_{H_n}(I) = \mathbb{E} \sum_{i \neq j} 1_I(\lambda_i(H_n)) 1_I(\lambda_j(H_n)) + \mathbb{E} \sum_i 1_I(\lambda_i(H_n)) - (\mathbb{E} \sum_i 1_I(\lambda_i(H_n)))^2
\]

\[\quad \sim \int_{I^2} \rho_{H_n}^{(2)}(x_1, x_2) dx_1 dx_2 - (\int_I \rho_{H_n}^{(1)}(x) dx) (\int_I \rho_{H_n}^{(1)}(x) dx - 1).
\]

Thus, asymptotics for \( k \)-point correlation functions yield asymptotics for moments of eigenvalue counting functions.
The analysis of correlation functions for the GUE ensemble $H_n^{\text{GUE}}$ is particularly nice, as it turns out that $n\mu_{H_n}$ is a determinantal point process, which means the correlation functions take the form
\[
\rho_{H_n}^{(k)}(x_1, \ldots, x_k) = \det[(K_n(x_i, x_j))_{1 \leq i, j \leq k}] \tag{2.6}
\]
for an explicit kernel function $K_n(x, y)$ involving the first $n$ Hermite polynomials (the family of orthogonal polynomials for the Gaussian measure). This can be shown by integrating out all but $k$ variables in the joint density (1.3) (after symmetrizing the density under relabelings of the eigenvalues) – see [Tao12, AGZ10] for the derivations. The determinantal structure ultimately stems from the presence of the squared Vandermonde factor in (1.3). By comparison, for a Poisson point process on $\mathbb{R}$ of constant intensity $\alpha$ the $k$-point correlation function is the constant $\alpha^k$.

At the local scale the semicircle law ceases to be valid, and the number of eigenvalues in an interval of length $O(1/n)$ will be a random variable that is not converging to a deterministic constant. We can consider the point process $n\mu_{H_n}$ in a $O(1/n)$ neighborhood of a fixed point $E \in (-2, 2)$ by considering the centered and rescaled $k$-point correlation functions
\[
\frac{1}{(n\rho_{sc}(E))^{k}} \rho_{H_n}^{(k)}(E + \frac{x_1}{\rho_{sc}(E)n}, \ldots, E + \frac{x_k}{\rho_{sc}(E)n}). \tag{2.7}
\]
The scalings by $\rho_{sc}(E)$ are needed to get a limit that is independent of $E$ (note that the typical size of eigenvalue spacings near $E$ will be inversely proportional to the local eigenvalue density $\rho_{sc}(E)$).

By a careful asymptotic analysis of the kernels in (2.6), one obtains that in the GUE case
\[
\frac{1}{(n\rho_{sc}(E))^{k}} \rho_{H_n}^{(k)}(E + \frac{x_1}{\rho_{sc}(E)n}, \ldots, E + \frac{x_k}{\rho_{sc}(E)n}) \rightarrow \det[(K(x_i, x_j))_{1 \leq i, j \leq k}] \tag{2.8}
\]
as $n \rightarrow \infty$, where $K$ is the Dyson sine kernel:
\[
K(x, y) := \frac{\sin \pi(x - y)}{\pi(x - y)}. \tag{2.9}
\]
Thus, at local scaling, the appropriately rescaled eigenvalue point processes $n\mu_{H_n}$ converge to a limiting determinantal point process – the sine(2) point process – with kernel given by (2.9). (There is a generalization to $\beta$-ensembles, but for $\beta \neq 2$ one does not get determinantal structure; for $\beta = 1, 4$ for instance one has Pfaffian point processes – see [AGZ10].)

The result (2.8)–(2.9) is for fixed $E \in (-2, 2)$ in the “bulk” of the spectrum (i.e. $|E \pm 2| \gtrsim 1$). For $E = \pm 2$ one gets a different limit involving the Airy kernel – again we refer to [AGZ10].

For the case $k = 2$ we have
\[
\frac{1}{(n\rho_{sc}(E))^{2}} \rho_{H_n}^{(2)}(E + \frac{x}{\rho_{sc}(E)n}, E + \frac{y}{\rho_{sc}(E)n}) \rightarrow 1 - \left(\frac{\sin \pi(x - y)}{\pi(x - y)}\right)^2. \tag{2.10}
\]
A key qualitative feature of the 2-point correlation function is that it vanishes to second order as $x - y \rightarrow 0$. Compare the situation for a Poisson point process, where the 2-point function is constant, and in particular doesn’t vanish as $x - y \rightarrow 0$.

The limiting Sine and Airy kernels have been shown to be universal for Wigner matrices, beginning around 2008–2009 with work of Erdős–Schlein–Yau using Dyson Brownian motion, and Tao–Vu using the Lindeberg swapping argument. In Section 4 we prove a four moment comparison theorem of Tao and Vu, showing that the $k$-point correlation functions asymptotically only depend on the first 4 moments of the atom distributions. Combining the comparison theorem with results for the GUE ensemble yields the Sine and Airy kernel asymptotics for a wide class of Wigner matrices. Note however that the 4 moments matching...
condition leaves out some important cases, such as the Rademacher distribution! In more recent years, strong two-moment universality results for eigenvalues as well as eigenvectors have been established by Erdős, Yau, Bourgade, Yin and coauthors using the Dyson Brownian motion method with deeper homogenization and coupling techniques. We will not have time to go into such advanced material in this short course, but refer to the text [EY17] and papers [EY15, BY17, BEYY16].

3. A first look at the Lindeberg swapping method: the CLT

Our approach to proving universality results for the spectrum of random matrices will be via a so-called Lindeberg swapping argument. As a warmup, in this section we introduce the method in the relatively simple (and original) context of the Central Limit Theorem (CLT). We roughly follow the treatment in [Tao12].

Let \((X_i)_{i \geq 1}\) be a sequence of iid real random variables with \(E X_i = 0\) and \(E X_i^2 = 1\). With \(Z_n := \frac{1}{\sqrt{n}} (X_1 + \cdots + X_n)\), the CLT says that \(Z_n \overset{d}{\to} G\) with \(G \sim N(0,1)\) a standard normal random variable. That is, we have \(E \phi(Z_n) \to E \phi(G)\) for any fixed \(\phi \in C_c(\mathbb{R})\).

The CLT is an example of what we call a two moment theorem: the distribution of a statistic of a large number of random variables (in this case \(\phi(\frac{1}{\sqrt{n}} \sum_{i \leq n} X_i)\)) is asymptotically insensitive to all details about the variables but the first and second moment.

The standard proof of the CLT (and stronger quantitative versions like the Berry–Esseen theorem) goes via Fourier analysis, which is well suited to the analysis of sums of independent random variables – at the level of distributions we are studying convolutions of measures.

We’ll use a different strategy – the Lindeberg swapping argument – to prove the following quantitative version of the CLT:

**Theorem 3.1** (Weak Berry–Esseen theorem). Let \(X\) be a random variable with \(E X = 0\), \(E X^2 = 1\) and \(E |X|^3 < \infty\). Let \(X_1, \ldots, X_n\) be iid copies of \(X\), put \(Z_n := \frac{1}{\sqrt{n}} (X_1 + \cdots + X_n)\), and let \(\phi \in C^\infty(\mathbb{R})\) with uniformly bounded derivatives up to order 3. Then

\[
E \phi(Z_n) = E \phi(G) + O(n^{-1/2}\|\phi'''\|_\infty E|X|^3).
\]

(3.1)

Most of the work will actually go into proving the following comparison theorem:

**Theorem 3.2** (Two moment theorem for sums). Let \(X, Y\) be random variables with moments matching to second order – i.e. \(E X^k = E Y^k\) for \(k = 1, 2\) – and finite absolute third moment. Let \(X = (X_i)_{i=1}^n\) and \(Y = (Y_i)_{i=1}^n\) be vectors of iid copies of \(X\) and \(Y\), respectively. Then defining \(f_n : \mathbb{R}^n \to \mathbb{R}\) by \(f_n(x) := \frac{1}{\sqrt{n}} (x_1 + \cdots + x_n)\), and letting \(\phi\) be as in Theorem 3.1, we have

\[
E \phi(f_n(X)) = E \phi(f_n(Y)) + O(n^{-1/2}\|\phi'''\|_\infty (E|X|^3 + E|Y|^3)).
\]

(3.2)

Theorem 3.1 follows immediately from Theorem 3.2 and the fact that \(G \overset{d}{=} \frac{1}{\sqrt{n}}(G_1 + \cdots + G_n)\) when \(G_1, \ldots, G_n\) are iid copies of \(G\).

**Remark** 3.3. Note that Theorem 3.2 makes no statement about the limiting or approximate distribution of \(f_n(X)\), only that it is essentially determined by the first 2 moments of the variables. 

**Proof of Theorem 3.2.** For each \(i \in [n]\) let

\[
V_i = (Y_1, \ldots, Y_i, X_{i+1}, \ldots, X_n)
\]

\[
V_i^0 = (Y_1, \ldots, Y_{i-1}, 0, X_{i+1}, \ldots, X_n)
\]
and \( V_0 = V_0^0 = X \). We have
\[
\mathbb{E}\phi(f_n(Y)) - \mathbb{E}\phi(f_n(X)) = \mathbb{E}\phi(f_n(V_n)) - \mathbb{E}\phi(f_n(V_0)) = \sum_{i=1}^{n} \mathbb{E}\phi(f_n(V_i)) - \mathbb{E}\phi(f_n(V_{i-1})).
\]

Now note that
\[
f_n(V_i) = f_n(V_i^0) + \frac{1}{\sqrt{n}} Y_i, \quad f_n(V_{i-1}) = f_n(V_i^0) + \frac{1}{\sqrt{n}} X_i.
\]

From Taylor expansion with remainder, we have
\[
\phi(f_n(V_i)) = \phi(f_n(V_i^0)) + \phi'(f_n(V_i^0)) \frac{Y_i}{\sqrt{n}} + \frac{1}{2} \phi''(f_n(V_i^0)) \frac{Y_i^2}{n} + O(\|\phi''\|_{\infty} |Y_i|^3 n^{-3/2})
\]
and similarly for \( \phi(f_n(V_{i-1})) \), with \( Y_i \)’s replaced by \( X_i \)’s. Subtracting and taking expectations, we obtain
\[
\mathbb{E}\phi(f_n(V_i)) - \mathbb{E}\phi(f_n(V_{i-1})) = \mathbb{E}\phi'(f_n(V_i^0)) \frac{Y_i - X_i}{\sqrt{n}} + \frac{1}{2} \mathbb{E}\phi''(f_n(V_i^0)) \frac{Y_i^2 - X_i^2}{n} + O(\|\phi''\|_{\infty} \mathbb{E}|X|^3 + \mathbb{E}|Y|^3 n^{-3/2}).
\]

Now for \( k = 1, 2 \), since \( X_i, Y_i \) are independent of \( V_i^0 \) and have matching \( k \)th moment, we get
\[
\mathbb{E}\phi^{(k)}(f_n(V_i^0)) \frac{Y_i^k - X_i^k}{n^{k/2}} = \mathbb{E}\phi^{(k)}(f_n(V_i^0)) \frac{\mathbb{E}(Y_i^k - X_i^k)}{n^{k/2}} = 0.
\]

Substituting back into (3.4), the first two terms on the right hand side are zero, and summing over \( i \) yields the claim. \( \square \)

**Remark 3.4.** A key feature of the functional \( f_n \) for the above argument is that it is linear, which allowed us to cleanly separate the dependence on a single coordinate \( X_i \) in (3.3). The situation will not be quite so clean when we work with spectral statistics of random matrices (although one motivation for working with resolvents will be that we can isolate the dependence on a single entry in a relatively clean way, thanks to the resolvent identity, though it will still not be quite so clean as in the linear case).

### 4. Lindeberg swapping for Wigner matrices

#### 4.1. The Stieltjes transform method and resolvents

In Section 4.3 we prove a comparison theorem for the spectrum of Wigner matrices that, when combined with the semicircle law for Gaussian matrices, yields the semicircle law for general Wigner matrices (Theorem 2.2). Thus we show an analogue of Theorem 3.2 from our proof of the CLT in Section 3; we note, however, that the Gaussian case for the semicircle law is not so easy as it was for the CLT.

Theorem 2.2 is a statement about convergence of (random) measures (in probability). Putting aside the randomness of the measures for a moment, we recall a few well-known methods for establishing convergence of deterministic sequences of probability measures.

A sequence \( (\mu_n)_{n \geq 1} \) of probability measures on \( \mathbb{R} \) converges weakly to another probability measure \( \mu \) (i.e. \( \mu_n \to \mu \)) if and only if \( \mu_n(\phi) \to \mu(\phi) \) for any \( \phi \in C_c(\mathbb{R}) \) (recall the notation \( \mu(\phi) := \int_{\mathbb{R}} \phi d\mu \)). We can specialize to various classes of test functions (or their restrictions to a sufficiently large compact interval \( I \) depending on \( \mu \)) whose linear span is dense in \( C_c(\mathbb{R}) \) (or \( C(I) \)). For instance, to show \( \mu_n \to \mu \) it suffices to show

1. (Method of moments). \( \int_{\mathbb{R}} \lambda^m d\mu_n(\lambda) \to \int_{\mathbb{R}} \lambda^m d\mu(\lambda) \) for all \( m \in \mathbb{N} \) (for this one also needs \( \mu \) to satisfy Carleman’s condition).
(2) (Fourier method). \( \int_{\mathbb{R}} e^{it\lambda} d\mu_n(\lambda) \to \int_{\mathbb{R}} e^{it\lambda} d\mu(\lambda) \) for all \( t \in \mathbb{R} \).

(3) (Stieltjes transform method). \( \int_{\mathbb{R}} \frac{d\mu_n(\lambda)}{\lambda - z} \to \int_{\mathbb{R}} \frac{d\mu(\lambda)}{\lambda - z} \) for all \( z \) in the upper half plane \( \mathbb{H} = \{ z \in \mathbb{C} : \text{Im} (z) > 0 \} \).

While the Fourier method is the main tool for analysis of sums of independent variables, (1) and (3) have been much more useful in the development of random matrix theory. In these notes we focus on (3), which has been shown to be particularly well suited to analyzing the spectrum at fine scales.

For a measure \( \mu \) on \( \mathbb{R} \) we denote its Stieltjes transform
\[
s(\mu,z) := \int_{\mathbb{R}} \frac{d\mu(\lambda)}{\lambda - z}, \quad z \in \mathbb{H}. \tag{4.1}
\]

How does \( (s(\mu,z))_{z \in \mathbb{H}} \) give information on \( \mu \)? Note that for \( z = E + i\eta \) with \( E \in \mathbb{R} \) and \( \eta > 0 \) we have
\[
\frac{1}{\pi} \text{Im} s(\mu,z) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\eta}{(\lambda - E)^2 + \eta^2} d\mu(\lambda) = \mu \ast \text{Cauchy}(\eta) \tag{4.2}
\]
where we recognized the Cauchy density with parameter \( \eta \). The sequence of densities \( \text{Cauchy}(\eta)_{\eta > 0} \) is a family of approximations to the identity, i.e. smooth densities converging weakly to the Dirac mass at 0. One deduces that
\[
\frac{1}{\pi} \text{Im} s(\mu,z) \to \mu \quad \text{as} \quad \eta \downarrow 0. \tag{4.3}
\]

For atomic measures such as the ESD \( \mu_M \) of an \( n \times n \) Hermitian matrix, the density \( \frac{1}{\pi} \text{Im} s(\mu,z) \) can be viewed as a smoothed version of \( \mu \) where the measure has smeared out at scale \( \eta \).

The Stieltjes transform is also convenient for studying ESDs of random matrices due to its relation to the matrix resolvent: for an \( n \times n \) matrix \( M \) and \( z \in \mathbb{C} \) we denote
\[
R(M,z) := (M - z)^{-1}. \]
If \( M \) is diagonalizable and \( z \) is not an eigenvalue of \( M \), then \( R(M,z) \) is diagonalizable with the same eigenbasis as \( M \) and eigenvalues \( 1/(\lambda_i(M) - z) \). For \( M \) Hermitian and \( z \in \mathbb{H} \) (so \( z \) cannot be an eigenvalue of \( M \) ), abusively writing
\[
s(M,z) := s(\mu_M, z),
\]
we have
\[
s(M,z) = \int_{\mathbb{R}} \frac{d\mu_M(\lambda)}{\lambda - z} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(M) - z} = \frac{1}{n} \text{Tr} R(M,z).
\]
Roughly speaking, estimates on \( s(M,E + i\eta) \) can be translated to estimates on the eigenvalues in a \( O(\eta) \)-neighborhood of \( E \), and vice versa. An example of this is shown in the following:

**Exercise 4.1** (Spectral anticoncentration from Stieltjes transform bounds). Let \( M \) be an \( n \times n \) Hermitian matrix. Show that if
\[
\sup_{z: \text{Im} (z) \geq \eta_0} |s(M,z)| \leq B_0 \tag{4.4}
\]
for some \( \eta_0 > 0 \), \( B_0 < \infty \), then for all intervals \( I \) of length at least \( \eta_0 \) we have
\[
\left| \{ i \in [n] : \lambda_i(M) \in I \} \right| \leq B_1 |I| n \tag{4.5}
\]
for some \( B_1 = O(B_0) \). Conversely, show that if \( \| M \|_{op} \leq B' \) for some \( B' < \infty \) and (4.5) holds all intervals \( I \) of length at least \( \eta_0 > 0 \), then (4.4) holds for some \( B_0 = O(B_1 \log(B'/\eta_0)) \).

(Hint: for one direction, note the Cauchy(\( \eta \)) density can be bounded pointwise from below by
\( c\eta^{-1}[-\eta,\eta] \) for a sufficiently small absolute constant \( c > 0 \). For the other direction, bound it above on a dyadic sequence of intervals \([\pm 2^k\eta, \pm 2^{k+1}\eta]\).

Resolvents are a useful tool for studying the spectrum of Hermitian matrices under perturbations, due to the resolvent identity. 

**Lemma 4.2 (Resolvent identity).** For \( n \times n \) Hermitian matrices \( M, A \) and \( z \in \mathbb{H} \), writing \( R = R(M, z) \) and \( R' = R(M + A, z) \), we have

\[
R' = R - RAR'.
\] (4.6)

Moreover, for any integer \( m \geq 1 \),

\[
R' = R + \sum_{k=1}^{m} (-RA)^k R + (-RA)^{m+1} R'.
\] (4.7)

**Proof.** One easily verifies (4.6) by multiplying both sides on the left by \( M + A - z \) and on the right by \( M - z \). The identity (4.7) follows from repeatedly substituting the identity (4.6) for the final factor of \( R' \) on the right hand side. \( \square \)

The error terms on the right hand side of (4.6), (4.7) take a particularly tractable form if \( A \) is supported on one or two entries, which will be the situation when we perform a Lindeberg swapping argument.

### 4.2. The Lindeberg method in RMT.

The Lindeberg swapping method was first introduced to random matrix theory by Chatterjee in [Cha], where (among other applications) he used it to prove a two moment comparison theorem for Stieltjes transforms of Wigner matrices, and, when combined with the result for the Gaussian case, give a new proof of Theorem 2.2. We give a proof of Chatterjee’s result in Section 4.3. We give a proof of Chatterjee’s result in Section 4.3.

Chatterjee’s argument was adapted to the empirical singular value distributions for random matrices of the form \( M - wI \) for \( M \) an \( n \times n \) matrix with iid entries and \( w \in \mathbb{C} \) by Krishnapur in the appendix of [TV10b]. The convergence of empirical singular value distributions for such matrices was an important step in the proof by of the Circular Law for the limit of ESDs of iid matrices, which was established under the optimal finite second moment hypothesis by Tao and Vu in [TV10b], building on many previous works – see the survey [BC12].

The Lindeberg swapping strategy was applied to the more challenging problem of universality of local spectral statistics of Wigner matrices by Tao and Vu in [TV11, TV10a]. In those first works they directly analyzed the effect of swapping on smooth statistics of the form \( \phi(\lambda_i(W_n), \ldots, \lambda_k(W_n)) \), which required technical a priori control on eigenvalue gaps due to the singular nature of their interaction. Another key component was the local semicircle law due to Erdős, Schlein and Yau [ESY09]. It was later realized that the swapping strategy is much cleaner (and less technically difficult) when considering smooth statistics of Stieltjes transforms rather than eigenvalues; a comparison theorem for eigenvalue statistics can then be deduced from a comparison theorem for Stieltjes transforms. This is the route we take in these notes. (The approach still relies on some consequences of the local semicircle law.)

### 4.3. A two moment theorem at the global spectral scale.

Our objective in this subsection is to establish the following:

**Theorem 4.3 (Two moment theorem for Stieltjes transforms).** Let \( H = (\frac{1}{\sqrt{n}}\xi_{ij}) \) and \( \tilde{H} = (\frac{1}{\sqrt{n}}\tilde{\xi}_{ij}) \) be normalized \( n \times n \) real Wigner matrices, such that for all \( 1 \leq i, j \leq n \) we have

\[
\phi(\lambda_i(H), \ldots, \lambda_k(H)) = \phi(\lambda_i(\tilde{H}), \ldots, \lambda_k(\tilde{H})),
\]
Proposition 4.5. Let $\xi_{ij}$ in (4.8). Thus, most of the work goes into showing the following:

\[ \text{Proposition 4.5. Let } \xi_{ij} \text{ and show the expected Stieltjes transforms only change by at most } n \text{ same at the slightly mesoscopic scale } \eta \text{ with respect to } \xi_{ij}. \]

Proof of Proposition 4.5.

From (4.8) as a telescoping sum:

\[ \mathbb{E}s(\tilde{H}, z) = \mathbb{E}s(H, z) + O(\frac{B}{n^{5/2} \eta^4}). \]  \hspace{1cm} (4.8)

Remark 4.4. While we only advertised the result as a statement about the global spectral scale, the quantitative error in fact shows the Stieltjes transforms of $H$ and $\tilde{H}$ are asymptotically the same at the slightly mesoscopic scale $\eta \gg n^{-1/8}$. This can be used (together with a result for the Gaussian case) to show that the semicircle law gives the asymptotic number of eigenvalues in intervals of length $\gg n^{-1/8}$.

As in the proof of the CLT from Section 3, well replace the variables $\xi_{ij}$ with $\tilde{\xi}_{ij}$ one by one and show the expected Stieltjes transforms only change by at most $n^{-2}$ times the error in (4.8). Thus, most of the work goes into showing the following:

Proposition 4.5. Let $H, \tilde{H}$ be as in Theorem 4.3 and assume further that $\xi_{ij} = \tilde{\xi}_{ij}$ a.s. for all $(i, j)$ except for $(i, j) = (p, q)$ for some $1 \leq p \leq q \leq n$. Then

\[ \mathbb{E}s(H, z) = \mathbb{E}s(\tilde{H}, z) + O(\frac{B}{n^{5/2} \eta^4}). \]  \hspace{1cm} (4.9)

Remark 4.6. We will actually show that (4.9) holds almost surely when expectation is taken with respect to $\xi_{pq}$ and $\tilde{\xi}_{pq}$ only.

Proof of Theorem 4.3 (granted Proposition 4.5). Let $H, \tilde{H}$ be as in Theorem 4.3. Fixing an arbitrary labelling $\alpha_1, \ldots, \alpha_{n+1}$ of $\{(i, j) : 1 \leq i \leq j \leq n\}$, we write the left hand side of (4.8) as a telescoping sum:

\[ \mathbb{E}s(\tilde{H}, z) - \mathbb{E}s(H, z) = \sum_{1 \leq k \leq \binom{n+1}{2}} \mathbb{E}s(H^{(k)}, z) - \mathbb{E}s(H^{(k-1)}, z) \]  \hspace{1cm} (4.10)

where $H^{(k)}$ has entries $\tilde{\xi}_{\alpha_i}/\sqrt{n}$ for $\ell \leq k$ and $\xi_{\alpha_i}/\sqrt{n}$ for $\ell > k$. Now note that for each $1 \leq k \leq \binom{n+1}{2}$, the pair of matrices $(H^{(k)}, H^{(k-1)})$ satisfies the hypotheses for $(H, \tilde{H})$ in Proposition 4.5 with $(p, q) = \alpha_k$. Hence, each summand in the right hand side of (4.10) is at most $O(B/(n^{5/2} \eta^4))$, and the claim follows upon summing over the $O(n^2)$ values of $k$.

To prove Proposition 4.5 we’ll use the resolvent expansion (4.7) with $m = 2$ and $A$ supported on the $(p, q)$ and $(q, p)$ entries. The terms $k = 1, 2$ in the expansion will cancel due to the independence of $\xi_{pq}, \tilde{\xi}_{pq}$ from the other entries and the matching moments hypothesis. To control the final error term we will use the following:

Lemma 4.7 (Crude bound for resolvents). Let $M$ be Hermitian and $z = E + i\eta \in \mathbb{H}$. Then $\|R(M, z)\|_{\text{op}} \leq 1/\eta$.


Proof of Proposition 4.5. Let

\[ E_{ij} = \begin{cases} e_i e_j^\top + e_j e_i^\top & i \neq j, \\ e_i e_i^\top & i = j \end{cases} \]  \hspace{1cm} (4.11)

and set

\[ H^0 = \frac{1}{\sqrt{n}} \sum_{i \leq j, (i, j) \notin (p, q)} \xi_{ij} E_{ij}. \]  \hspace{1cm} (4.12)
so that
\[ H = H^0 + \frac{1}{\sqrt{n}} \xi_{pq} E_{pq}, \quad \tilde{H} = H^0 + \frac{1}{\sqrt{n}} \tilde{\xi}_{pq} E_{pq}. \] (4.13)

The key observation is that \( \xi_{pq}, \tilde{\xi}_{pq} \) are independent from \( H^0 \). We abbreviate
\[ R^0 := R(H^0, z), \quad R^1 := R(H, z), \quad \tilde{R}^1 := R(\tilde{H}, z). \] (4.14)

Applying (4.7) with \( M = H^0, A = \frac{1}{\sqrt{n}} \xi_{pq} E_{pq} \) and \( m = 2 \), we have
\[ R^1 = R^0 - \frac{1}{n} \xi_{pq} R^0 E_{pq} R^0 + \frac{1}{n^2} \xi_{pq} R^0 E_{pq} R^0 R^0 - \frac{1}{n^{3/2}} \xi_{pq}^3 R^0 E_{pq} R^0 E_{pq} R^0 R^1. \] (4.15)

Similarly,
\[ \tilde{R}^1 = R^0 - \frac{1}{n} \tilde{\xi}_{pq} R^0 E_{pq} R^0 + \frac{1}{n^2} \tilde{\xi}_{pq} R^0 E_{pq} R^0 R^0 - \frac{1}{n^{3/2}} \tilde{\xi}_{pq}^3 R^0 E_{pq} R^0 E_{pq} \tilde{R}^1. \] (4.16)

Let
\[ E_{(p,q)}(\cdot) := \mathbb{E}(\cdot | (\xi_{ij})_{(i,j) \neq (p,q)}, (\tilde{\xi}_{ij})_{(i,j) \neq (p,q)}) \] (4.17)
de note expectation conditional on all entry random variables except \( \xi_{pq}, \tilde{\xi}_{pq} \). Subtracting (4.15) from (4.16), applying \( E_{(p,q)} \), and using that \( \xi_{pq}, \tilde{\xi}_{pq} \) are independent of \( R^0 \), we get that the zeroth, first and second order (in \( \xi_{pq}, \tilde{\xi}_{pq} \)) terms cancel, leaving
\[ E_{(p,q)}(\tilde{R}^1 - R^1) = \frac{1}{n^{3/2}} (\tilde{R}^0 E_{pq})^3 \mathcal{D} \quad \text{a.s.} \] (4.18)

where
\[ \mathcal{D} := E_{(p,q)} D, \quad D := \xi_{pq}^3 R^1 - \xi_{pq}^3 \tilde{R}^1. \] (4.19)

Let us consider only the case \( p < q \) (the case \( p = q \) being slightly simpler). We have
\[ E_{(p,q)}(s^1 - s^1) = \frac{1}{n} \text{Tr} E_{(p,q)}(\tilde{R}^1 - R^1) \]
\[ = \frac{1}{n^{3/2}} \sum_{i=1}^n \langle e_i, (\tilde{R}^0 E_{pq})^3 \mathcal{D} e_i \rangle \]
\[ = \frac{1}{n^{3/2}} \sum_{i=1}^n \langle e_i, (R^0 E_{pq})^2 R^0 e_p \rangle \langle e_q \mathcal{D} e_i \rangle + \langle e_i, (R^0 E_{pq})^2 R^0 e_q \rangle \langle e_p \mathcal{D} e_i \rangle \] (4.20)

almost surely. By the Cauchy–Schwarz inequality we have
\[ \left| \sum_{i=1}^n \langle e_i, (R^0 E_{pq})^2 R^0 e_p \rangle \langle e_q \mathcal{D} e_i \rangle \right| \leq \left( \sum_{i=1}^n |\langle e_i, (R^0 E_{pq})^2 R^0 e_p \rangle|^2 \right)^{1/2} \left( \sum_{i=1}^n |\langle e_q \mathcal{D} e_i \rangle|^2 \right)^{1/2} \]
\[ = \| \text{col}_p((R^0 E_{pq})^2 R^0)\|_2 \| \text{row}_q(\mathcal{D})\|_2 \]
\[ \leq \|(R^0 E_{pq})^2 R^0\|_{op} \|\mathcal{D}\|_{op} \quad \text{a.s.} \]

and we similarly get the same bound for the sum over the second terms in (4.20). Thus,
\[ |E_{(p,q)}(s^1 - s^1)| \leq \frac{2}{n^{3/2}} \|(R^0 E_{pq})^2 R^0\|_{op} \|\mathcal{D}\|_{op} \quad \text{a.s.} \] (4.21)

By sub-multiplicativity of the operator norm and Lemma 4.7 we have
\[ \|(R^0 E_{pq})^2 R^0\|_{op} \leq \|R^0\|_{op}^3 \|E_{pq}\|_{op}^2 \lesssim 1/n^3 \quad \text{a.s.} \] (4.22)
For $\|D\|_{\text{op}}$, from the triangle inequality and Lemma 4.7 we get
\[
\|D\|_{\text{op}} \leq |\xi_{pq}|^3 \|R_1\|_{\text{op}} + |\tilde{\xi}_{pq}|^3 \|\tilde{R}_1\|_{\text{op}} \leq \frac{1}{\eta} (|\xi_{pq}|^3 + |\tilde{\xi}_{pq}|^3) \quad \text{a.s.}
\]
and hence from Jensen’s inequality and the third moment hypothesis,
\[
\|D\|_{\text{op}} \leq \mathbb{E}_{(p,q)} \|D\|_{\text{op}} \leq 2B/\eta \quad \text{a.s.}
\] (4.23)
Substituting the bounds (4.22), (4.23) into (4.21), we have
\[
\|E_{(p,q)}(\tilde{s}^1 - s^1)\| \leq \frac{B}{n^{5/2} \eta^4} \quad \text{a.s.}
\]
From Jensen’s inequality we get
\[
|\mathbb{E}(\tilde{s}^1 - s^1)| = |\mathbb{E}\mathbb{E}_{(p,q)}(\tilde{s}^1 - s^1)| \leq \mathbb{E}|\mathbb{E}_{(p,q)}(\tilde{s}^1 - s^1)| \leq \frac{B}{n^{5/2} \eta^4}
\]
as desired. \qed

**Exercise 4.9.** Show that for any $\varepsilon > 0$ there exists $m(\varepsilon) \in \mathbb{N}$ such that if we assume $m$ matching moments and a bounded $(m+1)$st moment in Theorem 4.3, then we have \( \mathbb{E}s(H, E + i\eta) = \mathbb{E}\tilde{s}(\tilde{H}, E + i\eta) + o(1) \) uniformly for $\eta \geq n^{-1/2+\varepsilon}$.

4.4. **Four moment theorems at the local spectral scale.** The material in this subsection is loosely based on the notes [Tao19].

Now we establish a Stieltjes comparison result as in Theorem 4.3 allowing $\eta \asymp 1/n$, which is the local spectral scaling of typical eigenvalue spacings. In Exercise 4.9 we saw that the strategy of combining the resolvent bounds of Lemma 4.7 with matching moments hypotheses encounters a barrier at $\eta \gg n^{-1/2}$. To cross this barrier we need refined estimates on resolvent entries, which we obtain in Lemma 4.13 below as consequences of the local semicircle law. (For this we’ll need stronger tail hypotheses than a finite 3rd moment.) With these estimates in hand it will be enough to assume 4 matching moments, thus obtaining the following result. Towards an application to the universality of $k$-point correlation functions (Theorem 4.17) we establish universality for more general multilinear averages of Stieltjes transforms at several points.

**Theorem 4.10** (Four moment theorem for Stieltjes transforms at local scale). Let $H = (\frac{1}{\sqrt{n}} \xi_{ij})$ and $\tilde{H} = (\frac{1}{\sqrt{n}} \tilde{\xi}_{ij})$ be as in Theorem 4.3, only now we assume the entries $\xi_{ij}, \tilde{\xi}_{ij}$ have moments matching to 4th order and are $K$-sub-Gaussian for some fixed $K < \infty$ (see Definition 1.2). Fix $k \in \mathbb{N}$, let $E_1, \ldots, E_k \in \mathbb{R}$ (possibly depending on $n$) and $\eta \geq n^{-1-\frac{c_0}{100}}$ for a sufficiently small absolute constant $c_0 > 0$ ($\frac{1}{100}$ will do), and denote the function
\[
f(M) := \prod_{\ell=1}^k s(M, E_\ell + i\eta)
\] (4.24)
on $n \times n$ symmetric matrices $M$. We have
\[
\mathbb{E}f(H) = \mathbb{E}f(\tilde{H}) + O_{K,k}(n^{-c_k})
\] (4.25)
for a constant $c_k > 0$ depending only on $k$. Moreover, the same conclusion holds if we replace any number of the Stieltjes transforms in (4.24) with their complex conjugates.

In the remainder of this subsection $H$ and $\tilde{H}$ are as in Theorem 4.10.

As described above we need stronger resolvent bounds than what is provided by Lemma 4.7. Since that bound is sharp (Exercise!), any improvement could only hold with high probability.
Proposition 4.11 (A priori bounds for eigenvalues and eigenvectors). With $H'$ either $H$ or a matrix $H^0$ obtained as in (4.12) by setting a single variable $\xi_{pq}$ to zero, the following properties hold with probability $1 - O_{B,K}(n^{-B})$ for any fixed $B > 0$.

(a) (Spectral concentration). All eigenvalues of $H'$ have magnitude $O(1)$ (i.e. $\|H'\|_{\text{op}} = O(1)$).

(b) (Spectral anti-concentration). For any interval $I \subset \mathbb{R}$ (possibly depending on $n$), the number of eigenvalues of $H'$ in $I$ is at most $n^{o(1)}(1 + n|I|)$.

(c) (Eigenvector delocalization). The coefficients of all eigenvectors $u_1, \ldots, u_n$ of $H'$ are of magnitude at most $n^{o(1)} - 1/2$.

Property (a) can be established using Theorem 2.4 and following Remark 2.5; there are alternative approaches using techniques from high-dimensional geometry – see [Tao12, §2.3]. Properties (b) and (c) can be deduced from versions of the local semicircle law; for now we postpone discussion of the proof ideas.

Remark 4.12. We note that the bounds in (b), (c) are optimal up to the factors of $n^{o(1)}$. Indeed, for (c), the smallest $\|u\|_{\infty}$ can be for a unit vector $u$ is $1/\sqrt{n}$. For (b), we cannot have a bound better than 1 since one can take an arbitrarily small interval around a single eigenvalue. To see that we can’t do better than $O(n|I|)$, recall that from the semicircle law (Theorem 2.2), with probability $1 - o(1)$ (in particular at least $1/2$) there are at least $n - o(n)$ (in particular at least $n/2$) eigenvalues lying in the bounded interval $[-2, 2]$. Covering this interval with $O(1/|I|)$ translates of $I$, we see from the pigeonhole principle that some interval must contain $\gtrsim n|I|$ eigenvalues.

Using Proposition 4.11 and the spectral decomposition

$$R(H', z) = \sum_{i=1}^{n} \frac{u_i(H')u_i(H')^T}{\lambda_i(H') - z}$$

for the resolvent of $H'$, one can show the following:

Lemma 4.13 (Refined bounds for resolvent entries). With $H'$ as in Proposition 4.11, for any $B, \eta > 0$ and $E \in \mathbb{R}$, with probability $1 - O_{B,K}(n^{-B})$ we have that all entries of $R(H', E + i\eta)$ are of magnitude $O(n^{o(1)}(1 + \frac{1}{n\eta}))$.


Now we give the

Proof of Theorem 4.10. We give only the case $k = 1$, leaving the general case as an exercise. To simplify the bounds we focus on local regime and assume

$$n^{-1-c_0} \leq \eta \lesssim 1.$$  \hfill (4.27)

As in the proof of Theorem 4.3, by a telescoping sum argument it suffices to show

$$|E(s(H, E + i\eta) - s(\tilde{H}, E + i\eta))| \lesssim_{m,K} n^{-2-c_1}$$

under the additional hypothesis that $\xi_{ij} = \tilde{\xi}_{ij}$ almost surely for all $(i, j) \neq (p, q)$ for some $1 \leq p \leq q \leq n$.

2We stated the semicircle law for iid Wigner matrices with atom variables having distribution fixed independent of $n$, but it also holds for Wigner matrices satisfying the assumptions of Theorem 4.3 or Theorem 4.10.
We use the notation (4.11)–(4.14) from the proof of Theorem 4.3. Arguing similarly as in (4.15)–(4.20) but with (4.7) applied with general $m$ rather than $m = 2$ (we will later take $m = 4$), we find

$$|\mathbb{E}(s^1 - \tilde{s}^1)| \leq \mathbb{E}|\mathbb{E}_{(p,q)}(s^1 - \tilde{s}^1)| = \frac{1}{n^{1+(m+1)/2}} \mathbb{E} \left| \sum_{i=1}^{n} \langle e_i, (R^{0}E^{m+1}_p)D e_i \rangle \right|$$

where we now take

$$\overline{D} := \mathbb{E}_{(p,q)} D, \quad D := \xi^{m+1}_{pq} R^1 - \xi^{m+1}_{pq} \tilde{R}^1.$$  \hspace{1cm} (4.30)

We consider only the case $p < q$, the case $p = q$ being similar but slightly simpler. Expanding the product over $E^{pq} = e_p e_q^T + e_q e_p^T$ in (4.29) and applying the triangle inequality, we see the right hand side of (4.29) is bounded by a sum of $2^{m+1} = O(m)$ terms of the form

$$\frac{1}{n^{1+(m+1)/2}} \sum_{i=1}^{n} \langle e_i, R^0 e_{a_1} \rangle \langle e_{b_1}, R^0 e_{a_2} \rangle \cdots \langle e_{b_m}, R e_{a_{m+1}} \rangle \langle e_{b_{m+1}}, \overline{D} e_i \rangle$$

for $(a_\ell, b_\ell) \in \{(p, q), (q, p)\}$ for each $1 \leq \ell \leq m + 1$.

For $L > 0$ let $\mathcal{E}_L$ be the event that all entries of $R^0, R^1$ and $\tilde{R}^1$ have magnitude at most $L/n\eta$, and all entries of $H, \tilde{H}$ have magnitude at most $L/\sqrt{m}$. Applying Lemma 4.13, the sub-Gaussian hypothesis (specifically (1.1)) and the union bound we have

$$\mathbb{P}(\mathcal{E}_L) = 1 - O_{B,K}(n^{-B})$$

for any fixed $B > 0$ and some $L = n o(1)$. By using the crude bound of Lemma 4.7 on the complement of $\mathcal{E}_L$, we can bound

$$|\langle e_a, R e_b \rangle| \leq \frac{L}{n\eta} + \frac{1}{\eta} \mathbb{1}(\mathcal{E}_L^c) \quad \text{a.s.}$$

for any $a, b \in [n]$ and $R = R^0, R^1, \tilde{R}^1$. For any term in the sum (4.31) we can thus bound

$$|\langle e_i, R^0 e_{a_1} \rangle \langle e_{b_1}, R^0 e_{a_2} \rangle \cdots \langle e_{b_m}, R e_{a_{m+1}} \rangle \langle e_{b_{m+1}}, \overline{D} e_i \rangle|$$

$$\leq \left( \frac{L}{n\eta} \right)^{m+1} + \frac{1}{\eta^{m+1}} \mathbb{1}(\mathcal{E}_L^c) |\langle e_{b_{m+1}}, \overline{D} e_i \rangle| \quad \text{a.s.}$$

(4.33)

For the last term we have

$$|\langle e_j, \overline{D} e_i \rangle| \leq \mathbb{E}_{(p,q)} |\langle e_j, D e_i \rangle|$$

$$\leq \mathbb{E}_{(p,q)} (\xi_{pq}^{m+1} |\langle e_j, R^1 e_i \rangle| + |\xi_{pq}^{m+1} |\langle e_j, \tilde{R}^1 e_i \rangle|)$$

$$\leq \frac{L}{n\eta} \mathbb{E}_{(p,q)} (|\xi_{pq}^{m+1}| + |\xi_{pq}^{m+1}|) + \frac{1}{\eta} \mathbb{E}_{(p,q)} (|\xi_{pq}^{m+1}| + |\xi_{pq}^{m+1}|) \mathbb{1}(\mathcal{E}_L^c)$$

$$= O_{m,K}(\frac{L}{n\eta}) + \frac{1}{\eta} \mathbb{E}_{(p,q)} (|\xi_{pq}^{m+1}| + |\xi_{pq}^{m+1}|) \mathbb{1}(\mathcal{E}_L^c) \quad \text{a.s.}$$
where in the last line we used part (b) of Exercise 1.3. Thus,
\[
(\mathbb{E}|e_j, \overline{D}e_i|^2)^{1/2} \leq O_{m,K}(\frac{L}{n\eta}) + \frac{1}{\eta} \left( \mathbb{E}\left[ (|\xi_{pq}|^{2(m+1)} + |\tilde{\xi}_{pq}|^{2(m+1)}) 1(\mathcal{E}_L^c) \right] \right)^{1/2} \\
\leq O_{m,K}(\frac{L}{n\eta}) + \frac{1}{\eta} \left( (\mathbb{E}|\xi_{pq}|^{4(m+1)})^{1/2} + (\mathbb{E}|\tilde{\xi}_{pq}|^{4(m+1)})^{1/2} \right) \mathbb{P}(\mathcal{E}_L^c)^{1/4} \\
\leq O_{m,K}(\frac{L}{n\eta}) + O_{m,B,K}(\frac{1}{n} - B/4) \\
= O_{m,K}(\frac{L}{n\eta})
\]

taking \( B \) sufficiently large, where in the second line we used Cauchy–Schwarz, and in the third we used the sub-Gaussian assumption and (1.2). Taking expectations in (4.33), applying Cauchy–Schwarz and substituting the above bound and summing over \( i \), we get
\[
\mathbb{E}(|e_i, R^0 e_{a_1}\rangle\langle e_{b_1}, R^0 e_{a_2}\rangle \cdots \langle e_{b_m}, R^0 e_{a_{m+1}}\rangle\langle e_{b_{m+1}}, \overline{D}e_i|) \\
\leq O_{m,K}(\frac{L}{n\eta})^{m+2} + \frac{1}{\eta^{m+1}}O_{m,K}(\frac{L}{n\eta})\mathbb{P}(\mathcal{E}_L^c)^{1/2} \\
= O_{m,K}(\frac{L}{n\eta})^{m+2} = O_{m,K}(\frac{n^{o(1)}}{n\eta})^{m+2}
\]

where we applied (4.32) and took \( B \) sufficiently large. Taking expectation in (4.31), substituting the above bound and summing over \( i \) that (4.29) is bounded by
\[
\mathbb{E}|\mathbb{E}_{(p,q)}(s^1 - \bar{s}^1)| \lesssim_{m,K} \frac{n^{o(1)}}{n/(m+1/2)(\eta n)^{m+2}}. \tag{4.34}
\]

Taking \( m = 4 \), one verifies the right hand side above is of size at most \( n^{o(1)} - 2 - c \) for a constant \( c > 0 \) when \( \eta \geq n^{-1/4} \) (say). This yields (4.28) and hence the claim. \( \square \)

**Exercise 4.15.** Extend the above argument to establish Theorem 4.10 for general \( k \). \( \Diamond \)

**Exercise 4.16.** Show that we can relax the moment matching hypothesis to only assume two matching moments for diagonal entries. \( \Diamond \)

Now we use Theorem 4.10 to deduce a four moment theorem for smooth statistics of eigenvalues at the local spectral scaling.

**Theorem 4.17** (Four moment theorem for correlation functions at local scale). Let \( H, \tilde{H} \) be normalized real Wigner matrices as in Theorem 4.10. Fix \( k \in \mathbb{N} \) and \( \psi \in C_c^\infty(\mathbb{R}^k \to \mathbb{R}) \), let \( E \in \mathbb{R} \) (possibly depending on \( n \)), and denote the function
\[
g_k(M) := \sum_{i_1, \ldots, i_k \in [n] \text{ distinct}} \psi(n(\lambda_{i_1}(M) - E), \ldots, n(\lambda_{i_k}(M) - E))
\]
on \( n \times n \) symmetric matrices \( M \). We have
\[
\mathbb{E}g_k(H) = \mathbb{E}g_k(\tilde{H}) + O_{K,k,\psi}(n^{-c'_k}) \tag{4.35}
\]
for a constant \( c'_k > 0 \) depending only on \( k \).

Recalling the \( k \)-point correlation functions \( \rho^{(k)}_{H_n} \) defined in (2.5), from a change of variables we have
\[
\mathbb{E}g_k(H_n) = \frac{1}{n^k} \int_{\mathbb{R}^k} \rho^{(k)}_{H_n}(E + \frac{y_1}{n}, \ldots, E + \frac{y_k}{n})\psi(y_1, \ldots, y_k)dy_1 \cdots dy_k. \tag{4.36}
\]
Proof. We give only the case $k = 1$, leaving the general case as an exercise.

[...]

[... to be continued]

5. THE LOCAL SEMICIRCLE LAW AND EIGENVECTOR DELocalization

[...]

6. THE MONTGOMERY–DYSON PAIR CORRELATION CONJECTURE

[...]

REFERENCES


*Department of Mathematics, Duke University, 120 Science Dr, Durham, NC 27710

*E-mail address: nickcook@math.duke.edu*