Universality of Wigner Random Matrices

László Erdős

Institute of Mathematics, University of Munich, Theresienstr. 39, D-80333 Munich, Germany

Sep 15, 2009

Abstract

We consider $N \times N$ symmetric or hermitian random matrices with independent, identically distributed entries where the probability distribution for each matrix element is given by a measure $\nu$ with a subexponential decay. We prove that the local eigenvalue statistics in the bulk of the spectrum for these matrices coincide with those of the Gaussian Orthogonal Ensemble (GOE) and the Gaussian Unitary Ensemble (GUE), respectively, in the limit $N \to \infty$. Our approach is based on the study of the Dyson Brownian motion via a related new dynamics, the local relaxation flow. We also show that the Wigner semicircle law holds locally on the smallest possible scales and we prove that eigenvectors are fully delocalized and eigenvalues repel each other on arbitrarily small scales.

Keywords. Wigner random matrix, Dyson Brownian Motion.

AMS Subject Classification: 15B52, 82B44

1 Introduction

A central question concerning random matrices is the universality conjecture which states that local statistics of eigenvalues of large $N \times N$ square matrices $H$ are determined by the symmetry type of the ensembles but are otherwise independent of the details of the distributions.

There are two types of universalities: the edge universality and the bulk universality concerning the interior of the spectrum. The edge universality is commonly approached via the fairly robust moment method [33, 34]; very recently an alternative proof was given [36]. The bulk universality is a subtler problem. In the hermitian case, it states that the local $k$-point correlation functions of the eigenvalues, after appropriate rescaling, are given by the determinant of the sine kernel

$$\det (K(x_\ell - x_j))_{\ell,j=1}^{k}, \quad K(x) = \frac{\sin \pi x}{\pi x}, \quad (1.1)$$

independently of the distribution of the entries. Similar statement holds for the symmetric matrices but the explicit formulae are somewhat more complicated.

*Partially supported by SFB-TR 12 Grant of the German Research Council
For ensembles that remain invariant under the transformations $H \rightarrow U^* H U$ for any unitary matrix $U$, the joint probability density function of all the $N$ eigenvalues can be explicitly computed. These ensembles are typically given by the probability density

$$P(H)dH \sim \exp(-N \text{Tr} V(H))dH$$

where $V$ is a real function with sufficient growth at infinity and $dH$ is the flat measure. The eigenvalues are strongly correlated and they are distributed according to a Gibbs measure with a long range logarithmic interaction potential. The joint probability density of the eigenvalues of $H$ can be computed explicitly:

$$f(\lambda_1, \lambda_2, \ldots, \lambda_N) = \text{const.} \prod_{i<j} (\lambda_i - \lambda_j)^{\beta} \prod_{j=1}^{N} e^{-N \sum_{j=1}^{N} V(\lambda_j)},$$  \hspace{1cm} (1.2)

where $\beta = 1$ for symmetric and $\beta = 2$ for hermitian ensembles. The local statistics can be obtained via a detailed analysis of orthogonal polynomials on the real line with respect to the weight function $\exp(-V(x))$. Quadratic $V$ corresponds to the Gaussian ensembles. This approach was originally applied [26] for ensembles that lead to classical orthogonal polynomials (e.g. GUE leads to Hermite polynomials). Later a general method using orthogonal polynomials has been developed to tackle a very general class of unitary ensembles (see, e.g. [5, 9, 10, 26, 29] and references therein).

Many natural matrix ensembles are typically not unitarily invariant; the most prominent example is the Wigner matrices. These are symmetric or hermitian matrices whose entries above the diagonal are independent, identically distributed random variables. The only unitarily invariant Wigner ensembles are the Gaussian ensembles. For general Wigner matrices, no explicit formula is available for the joint eigenvalue distribution. Thus the basic algebraic connection between eigenvalue ensembles and orthogonal polynomials is missing and completely new methods needed to be developed.

The bulk universality for hermitian Wigner ensembles has been established jointly with S. Péché, J. Ramirez, B. Schlein and H.T. Yau, and independently by Tao-Vu [20, 35, 21]. These works rely on the Wigner matrices with Gaussian divisible distribution, i.e. ensembles of the form

$$\hat{H} + sV,$$  \hspace{1cm} (1.3)

where $\hat{H}$ is a Wigner matrix, $V$ is an independent standard GUE matrix and $s$ is a positive constant. Johansson [24] (see also [4]) proved bulk universality for the eigenvalues of such matrices using an explicit formula by Brézin-Hikami [6, 24] on the correlation functions. Unfortunately, the similar formula for symmetric matrices is not very explicit and the technique of [20, 24] cannot be extended to prove universality for symmetric Wigner matrices.

A key observation of Dyson is that if the parameter $s$ in the matrix $\hat{H} + sV$ is varied and $s^2$ is interpreted as time, then the evolution of the eigenvalues is given by a coupled system of stochastic differential equations, commonly called the Dyson Brownian motion (DBM) [12]. If we replace the Brownian motions by the Ornstein-Uhlenbeck processes to keep the variance constant, then the resulting dynamics on the eigenvalues, which we still call DBM, has the GUE eigenvalue distribution as the invariant measure. Thus the result of Johansson can be interpreted as stating that the local statistics of GUE is reached via DBM for time of order one. In fact, by analyzing the dynamics of DBM with ideas from the hydrodynamical limit, we have extended Johansson’s result to $s^2 \gg N^{-3/4}$ [19]. The key observation of [19] is that the local statistics of eigenvalues depend exclusively on the approach to local equilibrium which in general is faster than reaching global equilibrium. Unfortunately, the identification of local equilibria still uses explicit representations of
correlation functions by orthogonal polynomials (following e.g. [29]), and the extension to other ensembles is not a simple task.

Therefore, the universality for symmetric random matrices remained open and the only partial result is given by Tao-Vu (Theorem 23 in [35]) for Wigner matrices with the first four moments of the matrix elements matching those of GOE.

In [18], together with B. Schlein and H.T. Yau, we have introduced a general approach based on a new stochastic flow, the local relaxation flow, which locally behaves like DBM, but has a faster decay to equilibrium. This approach completely circumvents explicit formulae. It is thus applicable to prove the universality for a very broad class of matrices that includes hermitian, symmetric and symplectic Wigner matrices and in principle it works also for Wishart matrices and general $\beta$-ensembles. The heart of the proof is a convex analysis and the model specific information involve only estimates on the accuracy of the local density of states. For simplicity of the presentation, we will focus on the hermitian and symmetric cases, the necessary modifications for the other cases are technical.

We present results only about the convergence of the local correlation functions; this implies, among others, that the distribution of the gap (difference between neighboring eigenvalues) is universal as well (Wigner surmise). In particular, short gaps are suppressed, i.e. the eigenvalues tend to repel each other. This feature is characteristic to the strongly correlated point process of eigenvalues of random matrices in contrast to the Poisson process of independent points.

Universality of local eigenvalue statistics is believed to hold for a much broader class of matrix ensembles than we have introduced. Wigner has originally invented random matrices to mimic the eigenvalues of the then unknown Hamiltonian of heavy nuclei; lacking any information, he assumed that the matrix elements are i.i.d. random variables subject to the hermitian condition. Conceivably, the matrix elements need not be fully independent or identically distributed for universality. There is little known about matrices with correlated entries, apart from the unitary invariant ensembles that represent a very specific correlation. In case of a certain class of Wigner matrices with weakly correlated entries, the semicircle law and its Gaussian fluctuation have been proven [31, 32].

Much more studied are various classes of random matrices with independent but not identically distributed entries. The most prominent example is the Anderson model [2], i.e. a Schrödinger operator on a regular square lattice with a random potential. Restricted to a finite box, it can be represented by a matrix whose diagonal elements are i.i.d. random variables; the deterministic off-diagonal elements are given by the Laplacian. In space dimensions three or higher and for weak randomness, the Anderson model is conjectured to exhibit metal-insulator transition. Near the spectral edges, the eigenfunctions are localized [22, 1] and the local eigenvalue statistics is Poissonian [28]; in particular there is no level repulsion. It is conjectured, but not yet proven, that in the middle of the spectrum the eigenfunctions are extended (some results on the quantum diffusion and delocalization of eigenfunctions are available in a certain scaling limit [14, 7]). Furthermore, in the delocalization regime the local eigenvalue statistics are expected to be given by GUE or GOE statistics, depending whether the time reversal symmetry is broken by magnetic field or not. Based upon this conjecture, local eigenvalue statistics is used to compute the phase diagram numerically. It is very remarkable that the random Schrödinger operator, represented by a very sparse random matrix, exhibits the same universality class as the full Wigner matrix, at least in a certain energy range.

An intermediate class of ensembles between these two extremes is the family of random band matrices. These are hermitian or symmetric random matrices $H$ with independent but not identically distributed entries. The variance of $H_{ij}$ depends only on $|i - j|$ and it becomes negligible if $|i - j|$ exceeds a given parameter $W$, the band-width; for example, $\mathbb{E}|H_{ij}|^2 \sim \exp(-|i - j|/W)$. It is conjectured that for narrow bands, $W \ll \sqrt{N}$, the local eigenvalue statistics is Poisson, while for broad bands, $W \gg \sqrt{N}$ it is given by
GUE or GOE, depending on the symmetry class. (Localization properties of $H$ for $W \ll N^{1/8}$ has been recently shown [30] but not local statistics.) To mimic the three dimensional Anderson model, the rows and columns of $H$ may be labelled by a finite domain of the three dimensional lattice, $i, j \in \Lambda \subset \mathbb{Z}^3$. The only rigorous result for this three dimensional band matrix concerns the density of states by establishing that Wigner semicircle law holds as $W \to \infty$ [11].

Finally, we mention that universality of local eigenvalue statistics is often investigated by supersymmetric techniques in the physics literature. These methods are extremely powerful to extract the results by saddle point computations, but the analysis justifying the saddle point approximation usually lacks mathematical rigor. It is a challenge to the mathematical physics community to put the supersymmetric method on a solid mathematical basis; so far only the density of states has been investigated rigorously by using this technique [11].

2 Local semicircle law, delocalization and level repulsion

Each approach that proves bulk universality for general Wigner matrices requires first to analyze the local density of eigenvalues. The Wigner semicircle law [38] (and its analogue for Wishart matrices, the Marchenko-Pastur law [25]) has traditionally been among the first results established on random matrices. Typically, however, the empirical density is shown to converge weakly on macroscopic scales, i.e. on intervals that contain $O(N)$ eigenvalues. Based upon our results [15, 16, 17], here we show that the semicircle law holds on much smaller scales as well.

To fix the notation, we assume that in the symmetric case the matrix elements of $H$ are given by

$$h_{\ell k} = N^{-1/2}x_{\ell k},$$

(2.1)

where $x_{\ell k}$ for $\ell < k$ are independent, identically distributed random variables with the distribution $\nu$ that has zero expectation and variance 1. The diagonal elements $x_{\ell \ell}$ are also i.i.d. with distribution $\tilde{\nu}$ that has zero expectation and variance two. In the hermitian case we assume that

$$h_{\ell k} = N^{-1/2}(x_{\ell k} + iy_{\ell k})$$

(2.2)

where $x_{\ell k}$ and $y_{\ell k}$ are real i.i.d. random variables with zero expectation and variance $\frac{1}{2}$. The diagonal elements are also centered and have variance one. The eigenvalues of $H$ will be denoted by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$. The Gaussian ensembles (GUE and GOE) are special Wigner ensembles with Gaussian single-site distribution.

We will often need to assume that the distributions $\nu$ and $\tilde{\nu}$ have Gaussian decay, i.e. there exists $\delta_0 > 0$ such that

$$\int_{\mathbb{R}} \exp[\delta_0 x^2]d\nu(x) < \infty, \quad \int_{\mathbb{R}} \exp[\delta_0 x^2]d\tilde{\nu}(x) < \infty. \quad (2.3)$$

In several statements we can relax this condition to assuming only subexponential decay, i.e. that there exists $\delta_0 > 0$ and $\gamma > 0$ such that

$$\int_{\mathbb{R}} \exp[\delta_0 |x|^{\gamma}]d\nu(x) < \infty, \quad \int_{\mathbb{R}} \exp[\delta_0 |x|^{\gamma}]d\tilde{\nu}(x) < \infty. \quad (2.4)$$

The matrix elements have thus variance of order $1/N$. This normalization guarantees that the spectrum remains bounded as $N \to \infty$, in fact the spectrum converges to $[-2, 2]$ almost surely. Therefore the typical spacing between neighboring eigenvalues is of order $1/N$. 

4
For any \( I \subset \mathbb{R} \) let \( N_I \) denote the number of eigenvalues in \( I \). Wigner’s theorem [38] states that for any fixed interval \( I \)

\[
\frac{N_I}{N} \rightarrow \int_I \varrho_{sc}(x) \, dx
\]

almost surely as \( N \to \infty \), where

\[
\varrho_{sc}(x) := \frac{1}{2\pi} \sqrt{4 - x^2}
\]

is the density of the semicircle law. This result can be interpreted as a law of large numbers for the empirical eigenvalue density on macroscopic scales, i.e., for intervals that contain \( O(N) \) eigenvalues. The following result shows that the semicircle law holds on intervals \( I \) of length \( |I| = \eta \geq K/N \) for sufficiently large \( K \).

**Theorem 2.1** [17, Theorem 3.1] Suppose that (2.3) holds. Let \( \kappa > 0 \) and fix an energy \( E \in [-2 + \kappa, 2 - \kappa] \). Consider the interval \( I = [E - \frac{2}{\kappa}, E + \frac{2}{\kappa}] \) of length \( \eta \) about \( E \). Then there exist positive constants \( C, c \), depending only on \( \kappa \), and a universal constant \( c_1 \) such that for any \( \eta \leq c_1 \kappa \) there is \( K = K_\delta \) such that

\[
\mathbb{P} \left\{ \left| \frac{N_I}{N} - \varrho_{sc}(E) \right| \geq \delta \right\} \leq Ce^{-c\delta^2 \sqrt{N\eta}} \tag{2.5}
\]

holds for all \( \eta \) satisfying \( K/N \leq \eta \leq 1/K \).

In particular, this result shows that \( N_I/N \eta \) converges in probability as long as \( \eta = \eta(N) \) is such that \( \eta(N) \to 0 \) and \( N\eta(N) \to \infty \). The Gaussian decay condition (2.3) can be relaxed to (2.4) if \( \eta \geq N^{-1+\varepsilon} \) with any \( \varepsilon > 0 \) at the expense of a weaker bound on the right hand side of (2.5), see Section 5 of [20]. The estimate also deteriorates if the energy is close to the edge, see Proposition 4.1 of [19] for a more precise statement. Based upon our proofs, similar estimates were given in [35, Theorem 56] for energies in the bulk and somewhat stronger bounds in [36, Theorem 1.7] for the edge.

**Sketch of the proof.** For any \( z = E + i\eta, \eta > 0 \), let

\[
m(z) = m_N(z) = \frac{1}{N} \text{Tr} \frac{1}{H - z} = \frac{1}{N} \sum_{\alpha=1}^{N} \frac{1}{\lambda_\alpha - z}
\]

be the Stieltjes transform of the empirical density of states and let

\[
m_{sc}(z) = \int \frac{\varrho_{sc}(x)}{x - z} \, dx
\]

be the Stieltjes transform of the semicircle law. Clearly \( \varrho_{E} = \frac{1}{\pi} \text{Im} \, m(z) \) gives the normalized density of states of \( H \) around \( E \) regularized on a scale \( \eta \). Therefore it is sufficient to establish the convergence of \( m(z) \) to \( m_{sc}(z) \) for small \( \eta = \text{Im} \, z \).

The first step of the proof is to provide an upper bound on \( N_I \). Let \( B^{(k)} \) denote the \( (N - 1) \times (N - 1) \) minor of \( H \) after removing the \( k \)-th row and \( k \)-th column. Let \( \lambda^{(k)}_\alpha, \alpha = 1, 2, \ldots, N - 1 \) denote the eigenvalues of \( B^{(k)} \) and \( u^{(k)}_\alpha \) denote its eigenvectors. Computing the \((k, k)\) diagonal element of the resolvent \((H - z)^{-1}\) we easily obtain the following expression for \( m(z) \)

\[
m(z) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{H - z} (k, k) = \frac{1}{N} \sum_{k=1}^{N} \left[ h_{kk} - z - \frac{1}{N} \sum_{\alpha=1}^{N-1} \xi^{(k)}_\alpha \lambda^{(k)}_\alpha - z \right]^{-1}, \tag{2.7}
\]
\[ \xi^{(k)} = N |a^{(k)} \cdot u^{(k)}_a|^2, \]  
(2.8)

and \( a^{(k)} \) is the \( k \)-th column of \( H \) without the diagonal element \( h_{kk} \). Taking the imaginary part, and using \( N I \leq C \text{Im} m(z) \), we have

\[ N I \leq CN\eta^2 \sum_{k=1}^{N} \left| \sum_{\alpha : \lambda^{(k)}_\alpha \in I} \xi^{(k)}_\alpha \right|^{-1}. \]  
(2.9)

It is an elementary fact that the eigenvalues of \( H \) and \( B^{(k)} \), for each fixed \( k \), are interlaced, i.e. the number of \( \lambda^{(k)}_\alpha \) in \( I \) is at least \( N_I - 1 \). For each fixed \( k \) the random variables \( \{ \xi^{(k)}_\alpha : \alpha = 1, 2, \ldots N - 1 \} \) are almost independent and have expectation value one, thus the probability of the event

\[ \Omega_k := \left\{ \sum_{\alpha : \lambda^{(k)}_\alpha \in I} \xi^{(k)}_\alpha \leq \delta (N_I - 1) \right\} \]

is negligible for small \( \delta \) [17, Lemma 4.7]. On the complement of all \( \Omega_k \) we thus have from (2.9) that

\[ N_I \leq \frac{CN^2\eta^2}{\delta (N_I - 1)}, \]

from which it follows that \( N_I \leq CN\eta \) with very high probability.

The second step of the proof is to establish that \( m(z) \) and \( m_{sc}(z) \) are close. Let \( m^{(k)}(z) \) denote the Stieltjes transform of the empirical distribution of the eigenvalues \( \lambda^{(k)}_\alpha \) of \( B^{(k)} \). Then it follows from (2.7) that

\[ m(z) = \frac{1}{N} \sum_{k=1}^{N} h_{kk} - z - \left( 1 - \frac{1}{N} \right)m^{(k)}(z) - X_k \]  
(2.10)

holds, where

\[ X_k = \frac{1}{N} \sum_{\alpha = 1}^{N-1} \frac{\xi^{(k)}_\alpha - 1}{\lambda^{(k)}_\alpha - z}. \]

Fixing the matrix \( B^{(k)} \), we view \( X_k \) as a random variable of the independent \( a^{(k)} \) vector alone. Using again that the nominators \( \xi^{(k)}_\alpha - 1 \) are almost independent and have zero expectation, we obtain that \( X_k \) is bounded by \((N\eta)^{-1}\) with high probability [17, Lemma 6.1]. The interlacing property guarantees that \( m(z) \) and \( m^{(k)}(z) \) are close. Since \( h_{kk} \) is also small, we obtain from (2.10) that

\[ m(z) = -\frac{1}{N} \sum_{k=1}^{N} \frac{1}{m(z) + z + \varepsilon_k}. \]  
(2.11)

where \( \varepsilon_k \) are small with very high probability. Note that the Stieltjes transform of the semicircle law is the solution of the equation

\[ m_{sc}(z) = -\frac{1}{m_{sc}(z) + z} \]  
(2.12)

that is stable away from the spectral edges, \( z = \pm 2 \). Comparing the solution of (2.11) and (2.12) we obtain that \( |m - m_{sc}| \) is small. Strictly speaking, this argument applies only for \( \eta \geq (\log N)^4/N \) since the smallness
of each $\varepsilon_k$ is guaranteed only apart from a set of probability $e^{-c\sqrt{MN}}$ [17, Lemma 4.2] and there are $N$ possible values of $k$. On very short scale, our proof uses an additional expansion of the denominators in (2.11) up to second order and we use that the expectation of $X_k$, the main contribution to $\varepsilon_k$, vanishes [17, Section 6].

The second result concerns the delocalization of eigenvectors. The motivation comes from the Anderson model. In the infinite volume, the extended states regime is usually characterized by the absolute continuity of the spectrum; such characterization is meaningless for finite matrices. However, the lack of concentration of the eigenfunctions for the finite volume approximations of the Anderson Hamiltonian is already a signature of the extended states regime.

If $v$ is an $\ell^2$-normalized eigenvector of $H$, then the size of the $\ell^p$-norm of $v$, for $p > 2$, gives information about delocalization. Complete delocalization occurs when $\|v\|_p \leq N^{-1/2 + 1/p}$ (note that $\|v\|_p \geq CN^{-1/2 + 1/p}\|v\|_2$). The following result shows that eigenvectors are fully delocalized with a very high probability.

**Theorem 2.2** [17, Corollary 3.2] Under the conditions of Theorem 2.1, for any $|E| < 2$, fixed $K$ and $2 < p < \infty$ we have

$$P\left( \exists v : H v = \lambda v, \ |\lambda - E| \leq \frac{K}{N}, \ \|v\|_2 = 1, \ \|v\|_p \geq MN^{-1/2 + 1/p} \right) \leq C e^{-c\sqrt{MN}}$$

for $M$ and $N$ large enough.

The proof is an easy consequence of Theorem 2.1 and will be omitted here.

The local semicircle law asserts that the empirical density on scales $\eta \gg O(1/N)$ is close to the semicircle density. On even smaller scales $\eta \lesssim O(1/N)$, the empirical density fluctuates, but its average, $\bar{\varrho}(E)$, remains bounded uniformly in $\eta$. This is a type of Wegner estimate that plays a central role in the localization theory of random Schrödinger operators. In particular, it says that the probability of finding at least one eigenvalue in an interval $I$ of size $\eta = \varepsilon/N$ is bounded by $C \varepsilon$ uniformly in $N$ and $\varepsilon \leq 1$, i.e. no eigenvalue can stick to any value. Furthermore, if the eigenvalues were independent (Poisson process), then the probability of finding $n = 1, 2, 3, \ldots$ eigenvalues in $I$ were proportional with $\varepsilon^n$. For random matrices in the bulk of the spectrum this probability is much smaller. This phenomenon is known as level repulsion and the precise statement is the following:

**Theorem 2.3** [17, Theorem 3.4 and 3.5] Suppose (2.3) holds and the measure $\nu$ is absolutely continuous with a strictly positive and smooth density. Let $|E| < 2$ and $I = [E - \eta/2, E + \eta/2]$ with $\eta = \varepsilon/N$. Then for any fixed $n$,

$$P(N_I \geq n) \leq \begin{cases} C_1 \varepsilon^n & \text{[hermitian case]} \\ C_2 \varepsilon^{n(n+1)/2} & \text{[symmetric case]} \end{cases} \quad (2.13)$$

uniformly in $\varepsilon \leq 1$ and for all sufficiently large $N$.

The exponents are optimal as one can easily see from the Vandermonde determinant in the joint probability density (1.2) for unitary ensembles. The sine kernel behavior (1.1) implies level repulsion (and even a lower bound on $P(N_I \geq n)$), but usually not on arbitrarily small scales since sine kernel is typically proven only as a weak limit (see (3.2) later).
Sketch of the proof. The starting point is formula (2.7) together with
\[ N_I \leq CN\eta \Im m(E + i\eta). \]
This implies
\[ N_I \leq C\eta \sum_{k=1}^{N} \frac{1}{(a_k^2 + b_k^2)^{1/2}} \quad (2.14) \]
with
\[ a_k := \eta + \frac{1}{N} \sum_{a=1}^{N-1} \frac{\eta \xi_{\alpha}(k)}{(\lambda_{\alpha}(k) - E)^2 + \eta^2}, \quad b_k := h_{kk} - E - \frac{1}{N} \sum_{a=1}^{N-1} \frac{(\lambda_{\alpha}(k) - E)\xi_{\alpha}(k)}{(\lambda_{\alpha}(k) - E)^2 + \eta^2}, \]
where \( a_k \) and \( b_k \) are the imaginary and real part, respectively, of the reciprocal of the summands in (2.7) and \( \xi_{\alpha}(k) \) was defined in (2.8). The proof of Theorem 2.1 relied only on the imaginary part, i.e. \( b_k \) in (2.14) was neglected. In the proof of Theorem 2.3, however, we make an essential use of \( b_k \) as well. Since typically \( 1/N \lesssim |\lambda_{\alpha}(k) - E| \), we note that \( a_k^2 \) is much smaller than \( b_k^2 \) if \( \eta \ll 1/N \) and this is the relevant regime for the Wegner estimate and for the level repulsion.

Assuming a certain smoothness condition on the distribution \( d\nu \), the distribution of the variables \( \xi_{\alpha}(k) \) will also be smooth even if we fix an index \( k \) and we condition on the minor \( B(k) \), i.e. if we fix the eigenvalues \( \lambda_{\alpha}(k) \) and the eigenvectors \( u_{\alpha}(k) \). Although the random variables \( \xi_{\alpha}(k) = |a(k) \cdot u_{\alpha}(k)|^2 \) are not independent for different \( \alpha \)'s, they are sufficiently decorrelated so that the distribution of \( b_k \) inherits some smoothness from \( a(k) \). Sufficient smoothness on the distribution of \( b_k \) makes the expectation value \( (a_k^2 + b_k^2)^{-\eta/2} \) finite for any \( p > 0 \). This will give a bound on the \( p \)-th moment on \( N_I \) which will imply (2.13).

We present this idea for hermitian matrices and for the simplest case \( k = 1 \). From (2.14) we have
\[ \mathbb{P}(N_I \geq 1) \leq \mathbb{E} N_I^2 \leq C(N\eta)^2 \mathbb{E} \frac{1}{a_1^2 + b_1^2}. \]
Dropping the superscript \( k = 1 \) and introducing the notation
\[ d_{\alpha} = \frac{N(\lambda_{\alpha} - E)}{N^2(\lambda_{\alpha} - E)^2 + \varepsilon^2}, \quad c_{\alpha} = \frac{\varepsilon}{N^2(\lambda_{\alpha} - E)^2 + \varepsilon^2}, \]
we have
\[ \mathbb{P}(N_I \geq 1) \leq C\varepsilon^2 \mathbb{E} \left[ \left( \sum_{a=1}^{N-1} c_{\alpha}(\xi_{\alpha})^2 \right)^2 + \left( h - E - \sum_{a=1}^{N-1} d_{\alpha}\xi_{\alpha} \right)^2 \right]^{-1}. \quad (2.15) \]
From the local semicircle law we know that with very high probability, there are several eigenvalues \( \lambda_{\alpha} \) within a distance of \( O(1/N) \) of \( E \). Choosing four such eigenvalues, we can guarantee that for some index \( \gamma \)
\[ c_{\gamma}, c_{\gamma+1} \geq C\varepsilon, \quad d_{\gamma+2}, d_{\gamma+3} \geq C \quad (2.16) \]
for some positive constant \( C \). If \( \xi_{\alpha} \)'s were indeed independent and distributed according to the square of a complex random variable \( z_{\alpha} \) with a smooth and decaying density \( d\mu(z) \) on the complex plane, then the expectation in (2.15) would be bounded by
\[ \sup_E \int \frac{1}{(c_{\gamma}|z_{\gamma}|^2 + c_{\gamma+1}|z_{\gamma+1}|^2)^2 + (E - d_{\gamma+2}|z_{\gamma+2}|^2 - d_{\gamma+3}|z_{\gamma+3}|^2)^2} \prod_{j=0}^{3} d\mu(z_{\gamma+j}). \quad (2.17) \]
Simple calculation shows that this integral is bounded by $C\varepsilon^{-1}$ assuming the lower bounds (2.16). Combining this bound with (2.15), we obtain (2.13) for $n = 1$. The proof for the general $n$ goes by induction. The difference between the hermitian and the symmetric cases manifests itself in the fact that $\xi_\alpha$‘s are squares of complex or real variables, respectively. This gives different estimates for integrals of the type (2.17), resulting in different exponents in (2.13).

\section{Sine kernel universality}

Let $f(\lambda_1, \lambda_2, \ldots, \lambda_N)$ denote the symmetric joint density function of the eigenvalues of the $N \times N$ Wigner matrix $H$. For any $k \geq 1$ we define the $k$-point correlation functions (marginals) by

$$p_N^{(k)}(\lambda_1, \ldots, \lambda_k) = \int_{\mathbb{R}^{N-k}} f(\lambda_1, \lambda_2, \ldots, \lambda_N) d\lambda_{k+1} \ldots d\lambda_N.$$

We will use the notation $p_{N,GUE}^{(k)}$ and $p_{N,GOE}^{(k)}$ for the correlation functions of the GUE and GOE ensembles.

We consider the rescaled correlation functions about a fixed energy $E$ under a scaling that guarantees that the local density is one. The sine-kernel universality for the GUE ensemble states that the rescaled correlation functions converge weakly to the determinant of the sine-kernel, $K(x) = \frac{\sin \pi x}{\pi x}$, i.e.

$$\lim_{\delta \to 0} \lim_{N \to \infty} \frac{1}{[\theta_{sc}(E)]^k} p_{N,GUE}^{(k)}(E + \frac{x_1}{N\theta_{sc}(E)}, \ldots, E + \frac{x_k}{N\theta_{sc}(E)}) \to \det(K(x_\ell - x_j))_{\ell,j=1}^k$$

as $N \to \infty$ for any fixed energy $|E| < 2$ in the bulk of the spectrum [27, 13]. Similar result holds for the GOE case; the sine kernel being replaced with a similar but somewhat more complicated universal function, see [26]. Our main result is that universality (3.1) holds for general hermitian or symmetric Wigner matrices after averaging in the energy $E$:

\begin{theorem} \cite{18}
Let $H$ be an $N \times N$ symmetric or hermitian Wigner matrix with normalization defined at the beginning of Section 2. Suppose that the distribution $\nu$ of the matrix elements has subexponential decay (2.4). Let $k \geq 1$ and $O : \mathbb{R}^k \to \mathbb{R}$ be a continuous, compactly supported function. Then for any $|E| < 2$, we have

$$\lim_{\delta \to 0} \lim_{N \to \infty} \frac{1}{N^{k-1}} \int_{E+\delta}^E \int_{E-\delta}^E d\alpha_1 \ldots d\alpha_k O(\alpha_1, \ldots, \alpha_k)$$

$$\times \frac{1}{[\theta_{sc}(v)]^k} \left( p_{N}^{(k)}(\cdot) - \left(\frac{\alpha_1}{N\theta_{sc}(v)}, \ldots, \frac{\alpha_k}{N\theta_{sc}(v)}\right) \right) = 0,$$

where $\#$ stands for GOE or GUE for the symmetric or hermitian cases, respectively.

For the hermitian case, the first result on universality beyond the GUE was due to Johansson \cite{24} (based upon \cite{6}) under the condition that $\nu$ has a Gaussian component with a positive variance independent of $N$. His method was extended in \cite{4} to Wishart matrices. The variance of the necessary Gaussian component was reduced to $N^{-3/4+\varepsilon}$ in \cite{19} under the additional technical assumptions that the measure $\nu$ is smooth and it satisfies the logarithmic Sobolev inequality. The local statistics was identified via orthogonal polynomials. The Gaussian component assumption was first removed completely in \cite{20} under the condition that the density of the probability measure $\nu$ is positive and it possesses a certain number of derivatives. Shortly after \cite{20} appeared on the arXiv, the same result using a different method has been posted \cite{35} without
any regularity condition on $\nu$ provided that the third moment vanishes and $\nu$ is supported on at least three points. Combining the two methods, all conditions on $\nu$ apart from the subexponential decay (2.4) were removed in a short joint paper [21].

The methods of [20] and [35] both rely on the explicit formula of Brézin and Hikami [6], exploited also in [24], for the correlation functions of the Wigner matrix with Gaussian convolution. This formula reduces the problem to a saddle point analysis. The saddle points are identified by solving an equation involving the Stieltjes transform $m_N(z)$ (2.6) with $\eta = \text{Im} z$ corresponding to the variance of the Gaussian component: precise information on $m_N(z)$ for a smaller $\eta$ implies that a smaller Gaussian component is sufficient.

In our work [20] we used the convergence of $m_N(z)$ to $m_{sc}(z)$ for very small $\eta = N^{-1+\varepsilon}$ established along the proof of Theorem 2.1. To remove this tiny Gaussian component, we have compared the local eigenvalue statistics of a given Wigner matrix $H$ with that of $\hat{H}_s + sV$ for which the saddle point analysis applies. Here $s^2 = \eta = N^{-1+\varepsilon}$ and the new Wigner matrix $\hat{H}_s$ was chosen such that the law of $\hat{H}_s + sV$ be very close to $H$. Since Gaussian convolution corresponds to running a heat flow on the matrix elements, $\hat{H}_s$ could, in principle, be obtained by running the reverse heat flow on the elements of $H$. Although the reverse heat flow is undefined for most initial conditions, one can construct an approximation to the reverse heat flow that is well defined and yields $\hat{H}_s$ with a required precision assuming sufficient smoothness on $\nu$. Technically, we use Ornstein-Uhlenbeck process instead of the heat flow to keep the variance constant. We also mention that the result of [20] is valid for any fixed energy $E$, i.e. $dv$ averaging in (3.2) is not necessary.

Tao and Vu [35] have directly compared local statistics of the Wigner matrix $H$ and that of the matrix with order one Gaussian component for which Johansson has already proved universality. Their main technical result [35, Theorem 15] states that the local eigenvalue statistics of two Wigner matrices coincide as long as the first four moments of their single site distributions match. It is then an elementary lemma from probability theory ([35, Corollary 23] based upon [8]) to match to order four a given random variable with another random variable with a Gaussian component.

The proof of Theorem 3.1 for the symmetric case requires a new idea since the formula of Brézin and Hikami is not available. While the four moment theorem of [35] also applies to this case, there is no reference ensemble available. In the next sections we describe our new approach that proves universality for both hermitian and symmetric matrices without relying on any explicit formulae.

4 Dyson Brownian motion

The joint distribution of the eigenvalues $x = (x_1, x_2, \ldots, x_N)$ of the Gaussian ensembles is given by the following measure

$$
\mu = \mu_N(dx) = \frac{e^{-\mathcal{H}(x)}}{Z_\beta} dx,
\mathcal{H}(x) = N \left[ \beta \sum_{i=1}^{N} x_i^2 \frac{1}{4} - \frac{\beta}{N} \sum_{i<j} \log |x_j - x_i| \right]
$$

where $\beta = 1$ for GOE and $\beta = 2$ for GUE. For definiteness, we consider the $\beta = 1$ GOE case and we assume that the eigenvalues are ordered, i.e. $\mu$ is restricted to $\Sigma_N = \{ x \in \mathbb{R}^N : x_1 < x_2 < \ldots < x_N \}$.

Suppose the matrix elements evolve according to the Ornstein-Uhlenbeck process on $\mathbb{R}$, i.e. the density of their distribution $\nu_t = u_t(x)dx$ satisfies

$$
\partial_t u_t = \mathcal{L} u_t, \quad \mathcal{L} = \frac{1}{2} \frac{\partial^2}{\partial x^2} - \frac{x}{2} \frac{\partial}{\partial x}
$$

(4.2)
The Ornstein-Uhlenbeck process (4.2) induces a stochastic process, the Dyson Brownian motion, on the eigenvalues with a generator given by

$$L = \sum_{i=1}^{N} \frac{1}{2N} \partial_i^2 + \sum_{i=1}^{N} \left( -\frac{\beta}{4} x_i + \frac{\beta}{2N} \sum_{j \neq i} \frac{1}{x_i - x_j} \right) \partial_i$$

(4.3)

acting on $L^2(\mu)$. The measure $\mu$ is invariant and reversible with respect to the dynamics generated by $L$. Let

$$D(f) = -\int fLfd\mu = \sum_{j=1}^{N} \frac{1}{2N} \int (\partial_j f)^2 d\mu$$

(4.4)

be the corresponding Dirichlet form. Denote the distribution of the eigenvalues at time $t$ by $f_t(x)\mu(dx)$. Then $f_t$ satisfies

$$\partial_t f_t = Lf_t$$

(4.5)

with initial condition $f_0$ given by the eigenvalue density of the Wigner ensemble. Dyson Brownian motion is the corresponding system of stochastic differential equations for the eigenvalues $x(t)$ that is given by (see, e.g. Section 12.1 of [23])

$$dx_i = dB_i \sqrt{N} + \left[ -\frac{\beta}{4} x_i + \frac{\beta}{2N} \sum_{j \neq i} \frac{1}{x_i - x_j} \right] dt, \quad 1 \leq i \leq N,$$

(4.6)

where $\{B_i : 1 \leq i \leq N\}$ is a collection of independent Brownian motions. Note that the equations (4.5) and (4.6) are defined for any $\beta \geq 1$, independently of the original matrix models. Our main technical result (Theorem 5.1) holds for general $\beta \geq 1$.

## 5 Local Relaxation Flow

The Hamiltonian of the invariant measure $\mu$ of the Dyson Brownian motion is convex, with Hessian bounded from below

$$\text{Hess } \mathcal{H} \geq \frac{\beta N}{2}$$

on the set $\Sigma_N$. By the Bakry-Emery criterion, this guarantees that $\mu$ satisfies the logarithmic Sobolev inequality and the relaxation time to equilibrium is of order one (note the additional $1/N$ factor in the Dirichlet form (4.4) that rescales time).

We now introduce the local relaxation measure, which has the local statistics of GOE (or GUE) but generates a faster decaying dynamics. Let $\gamma_j$ be the semicircle location of the $j$-th eigenvalue, i.e.

$$\gamma_j = n_{sc}^{-1}(j/N), \quad n_{sc}(E) := \int_{-\infty}^{E} q_{sc}(x)dx.$$

We fix a regularization parameter $\eta \ll 1$ and we replace the interaction potential between $x_j$ and far away particles by a regularized mean field potential

$$W_j(x) = -\frac{\beta}{N} \sum_{k:|k-j| \geq N\eta} \log(|x - \gamma_k| + \eta)$$

(5.1)
Strictly speaking, $W_j(x)$ is defined by this formula only in an interval of size $N\eta$ about $\gamma_j$ and we use a quadratic extension beyond, but we leave this technicality aside.

The local relaxation measure $\omega_N = \omega$ is a Gibbs measure defined by the Hamiltonian

$$\tilde{H} = N \sum_{j=1}^{N} \left\{ \beta \frac{x_j^2}{4} + W_j(x_j) \right\} + \beta \sum_{i<j} \log |x_i - x_j| - \frac{\beta}{2} \sum_{j:|j-i|>N\eta} \log(|x_i - x_j| + \eta).$$

We often write $\omega = \psi \mu$ where $\psi$ is the Radon-Nykodim derivative. The local relaxation flow is defined to be the reversible dynamics w.r.t. $\omega$ characterized by the generator $\tilde{L}$ defined by

$$\int f \tilde{L} g \, d\omega = -\frac{1}{2N} \sum_j \int \partial_j f \partial_j g \, d\omega. \quad (5.2)$$

Explicitly, $\tilde{L}$ is given by

$$L = \tilde{L} + \sum_j b_j \partial_j, \quad b_j = \frac{1}{N} \sum_{k:|k-j|>N\eta} \frac{\text{sgn}(x_j - x_k)}{|x_j - x_k| + \eta} + W'_j(x_j), \quad (5.3)$$

Simple calculation shows that the mean field potential is uniformly convex with

$$\inf_{j} \inf_{x \in \mathbb{R}} W''_j(x) \geq c \eta^{-1/3}. \quad (5.4)$$

This will guarantee that the relaxation time to equilibrium $\omega$ for the $\tilde{L}$ dynamics is of order $\eta^{-1/3}$.

We recall the definition of the relative entropy of with respect to any probability measure $d\lambda$

$$S_\lambda(f) = \int f \log f \, d\lambda, \quad S_\lambda(f|\psi) = \int f \log(f/\psi) \, d\lambda.$$

Our main technical result is the following theorem that states that the relaxation time $\tau$ for specific local observables is much shorter than order one.

**Theorem 5.1 (Universality of Dyson Brownian Motion for Short Time)** Suppose that $S_\mu(f_0|\psi) \leq CN^m$ for some $m$ fixed. Let $\tau = \eta^{1/3}N^\varepsilon$ with some $\varepsilon > 0$ and assume that $\eta \geq N^{-2/55+\varepsilon}$. Assume that there is a positive number $\Lambda$ such that

$$\sup_{0 \leq t \leq \tau} N \sum_j b^2_j f_t \, d\mu \leq C \eta^{-2} \Lambda. \quad (5.5)$$

Let $G$ be a bounded smooth function with compact support. Then for any fixed $n \geq 1$ and $J \subset [1, \ldots, N]$ we have

$$\left| \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) f_{\tau} \, d\mu - \int \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n})) \, d\mu \right| \leq \sqrt{\frac{CA}{N^{1-\varepsilon} \eta^{5/3}}}.$$

We emphasize that Theorem 5.1 applies to all $\beta \geq 1$ ensembles and the only assumption concerning the distribution $f_t$ is in (5.5). In case of the original Wigner ensembles $\beta = 1, 2$, the critical constant $\Lambda$ can be estimated under an additional assumption.
Lemma 5.2 Let \( f_0 \) be the joint density of the eigenvalues of a Wigner matrix. Suppose that the measure \( d\nu \) of its single site distribution satisfies the logarithmic Sobolev inequality. Then the constant \( \Lambda \) in (5.5) can be estimated as
\[
\Lambda \leq C_\sigma \eta^{-2} N^{4/5+\sigma} 
\] for any \( \sigma > 0 \).

For the proof of this lemma, we can estimate \( b_j \) as
\[
|b_j| \leq \frac{1}{N} \sum_{k:|j-k|>N\eta} \frac{\text{sgn}(x_j-x_k)}{|x_j-x_k|+\eta} - \frac{\text{sgn}(x_j-\gamma_k)}{|x_j-\gamma_k|+\eta} \leq C\eta^{-2} \frac{1}{N} \sum_{k=1}^{N} |x_k-\gamma_k| 
\] as long as \( x_k \) is sufficiently near \( \gamma_k \) so that \( \text{sgn}(x_j-\gamma_k) = \text{sgn}(\gamma_j-\gamma_k) \) holds for \( |j-k| > N\eta \). The average difference between \( x_k \) and \( \mathbb{E}x_k \) can be estimated using the logarithmic Sobolev inequality for \( \nu \). The average of \( |\mathbb{E}x_k-\gamma_k| \) is estimated in Proposition 4.2 of [17] that was a consequence of the local semicircle law. Combining these results with information on the lowest and largest eigenvalues [37], we can show that
\[
\frac{1}{N} \sum_k |x_k-\gamma_k| \leq N^{-3/5+\varepsilon} \text{ and this yields (5.6).}
\]

Combining Lemma 5.2 with Theorem 5.1 and choosing \( \eta \) appropriately, we see that the local eigenvalue statistics of \( f_\tau \) with \( \tau \geq N^{-1/55+\varepsilon} \) coincides with that of the global equilibrium measure, i.e. with GOE or GUE. For hermitian matrices, the same statement was already proven in [19] even for \( \tau \geq N^{-1+\varepsilon} \) by using Brézin-Hikami formula, but the current approach is purely analytical and it applies to symmetric matrices as well. Using the reverse heat flow argument, we can show that the local statistics of \( f_0 \) is also given by GOE or GUE assuming that the initial distribution \( \nu \) is sufficiently smooth. The smoothness condition and the additional requirement that \( \nu \) satisfies the logarithmic Sobolev inequality can be removed by applying the four moment theorem of [35].

6 Proof of Theorem 5.1

We first list the key new ideas of behind the proof of Theorem 5.1, then we formulate the corresponding results.

I. The key concept is the introduction of the local relaxation flow (5.2) which has the following two properties: (1) The invariant measure for this flow, the local relaxation measure \( \omega \) has the same local eigenvalue statistics as the GOE or GUE. (2) The relaxation time of the local relaxation flow is much shorter than that of the DBM, which is of order one.

II. Suppose we have a density \( q \) w.r.t. \( \omega \) that evolves with the local relaxation flow. Then, by differentiating the Dirichlet form w.r.t. \( \omega \) we will prove that the difference between the local statistics of \( q\omega \) and \( \omega \) can be estimated in terms of the Dirichlet form of \( q \) w.r.t. \( \omega \). Hence if the Dirichlet form is small, the local statistics of \( q\omega \) is independent of \( q \).

III. It remains to show that the Dirichlet form of \( q = f_\mu \omega \) is small for \( t \) sufficiently large (but still much less than order one). To do that, we study the evolution of the entropy of \( f_\mu \omega \) relative to \( \omega \). This provides estimates on the entropy and Dirichlet form which serve as inputs for the Step II to conclude the universality.
The first ingredient to prove Theorem 5.1 is the analysis of the local relaxation flow which satisfies the logarithmic Sobolev inequality and the following dissipation estimate.

**Theorem 6.1** Suppose (5.4) holds. Consider the equation

\[ \partial_t q_t = \mathcal{L} q_t \]  

with reversible measure \( \omega \). Then we have the following estimates

\[
\begin{align*}
\partial_t D_\omega(\sqrt{q_t}) &\leq -C\eta^{-1/3} D_\omega(\sqrt{q_t}) - \frac{1}{2N^2} \int \sum_{|i-j| \leq N\eta} \frac{1}{(x_i - x_j)^2} (\partial_i \sqrt{q_t} - \partial_j \sqrt{q_t})^2 d\omega, \\
\frac{1}{2N^2} \int_0^\infty ds \int \sum_{|i-j| \leq N\eta} \frac{1}{(x_i - x_j)^2} (\partial_i \sqrt{q_s} - \partial_j \sqrt{q_s})^2 d\omega &\leq D_\omega(\sqrt{q_0})
\end{align*}
\]

and the logarithmic Sobolev inequality

\[
S_\omega(q_t) \leq C\eta^{1/3} D_\omega(\sqrt{q_t})
\]

with a universal constant \( C \). Thus the time to equilibrium is of order \( \eta^{1/3} \):

\[
S_\omega(q_t) \leq e^{-C\eta^{-1/3} S_\omega(q_0)}.
\]

The proof follows the standard argument in [3] (used in this context in [19]). The key input is the following lower bound on the Hessian of \( \mathcal{H} \)

\[
\frac{1}{2N^2} \langle \nabla (\nabla^2 \mathcal{H}) \nabla \rangle \geq C\eta^{-1/3} \frac{1}{N} \| \nabla \|^2 + \frac{1}{2N^2} \sum_{|i-j| \leq N\eta} \frac{1}{(x_i - x_j)^2} (v_i - v_j)^2.
\]

The first term is due to convexity of the mean field potential (5.4). The second term comes from the additional convexity of the local interaction and it corresponds to “local Dirichlet form dissipation”. The estimate (6.3) on this additional term plays a key role in the next theorem.

**Theorem 6.2** Suppose that the density \( q_0 \) satisfies \( S_\omega(q_0) \leq CN^m \) with some \( m > 0 \) fixed. Let \( G \) be a bounded smooth function with compact support and let \( J \subset \{1, 2, \ldots, N\} \). Set \( \tau = \eta^{1/3} N^\epsilon \). Then for any \( n \geq 1 \) fixed we have

\[
\left| \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n}))d\omega - \frac{1}{N} \sum_{i \in J} G(N(x_i - x_{i+n}))q_0d\omega \right| \leq C \sqrt{\frac{D_\omega(\sqrt{q_0})\tau}{N}} + Ce^{-cN^\epsilon}.
\]

**Sketch of the proof.** Let \( q_t \) satisfy

\[ \partial_t q_t = \mathcal{L} q_t \]

with an initial condition \( q_0 \). Thanks to the exponential decay of the entropy on time scale \( \tau \gg \eta^{1/3} \), see (6.5), difference between the local statistics w.r.t \( q_t, \omega \) and \( q_\infty, \omega = \omega \) is subexponentially small in \( N \). To compare \( q_0 \) with \( q_\infty \), by differentiation, we have

\[
\int \frac{1}{N} \sum_i G(N(x_i - x_{i+n}))q_t d\omega - \int \frac{1}{N} \sum_i G(N(x_i - x_{i+n}))q_0 d\omega
\]
Theorem 6.3 Suppose the assumptions of Theorem 5.1 hold. Let
\[ S_\omega \text{ term on the r.h.s. of (6.2), we gain the crucial factor } N \]
where \( g \) of (6.7) will become \( t \)

From the Schwarz inequality and \( \partial q \) in time.

Together with (6.4) and (5.5), we have
\[ \text{be obtained from integrating (6.10) from } t \]

The entropy and the Dirichlet form in (6.9) are, however, computed with respect to the measure \( \mu \) which, after integrating it from \( t \) to \( \tau /2 \), proves the first inequality in (6.9). The second inequality can be obtained from integrating (6.10) from \( t = \tau /2 \) to \( t = \tau \) and using the monotonicity of the Dirichlet form in time.

\[ \square \]

The final ingredient to prove Theorem 5.1 is the following entropy and Dirichlet form estimates.

**Theorem 6.3** Suppose the assumptions of Theorem 5.1 hold. Let \( \tau = \eta^{1/3} N^\varepsilon \) and let \( g_t = f_t / \psi \) so that \( S_\mu (f_t | \psi) = S_\omega (g_t) \). Then the entropy and the Dirichlet form satisfy the estimates:
\[ S_\omega (g_t | 2) \leq C \eta^{-5/3} \Lambda, \quad D_\omega (\sqrt{\eta t}) \leq C \eta^{-2} \Lambda. \]

**Sketch of the proof.** Recall that \( \partial_t f_t = L f_t \). The standard estimate on the entropy of \( f_t \) with respect to the invariant measure is obtained by differentiating it twice and using the logarithmic Sobolev inequality. The entropy and the Dirichlet form in (6.9) are, however, computed with respect to the measure \( \omega \). This yields the additional second term in the following identity [39] that holds for any probability density \( \psi_t \):
\[ \partial_t S_\mu (f_t | \psi_t) = -\frac{2}{N} \sum_j \int (\partial_j \sqrt{\psi_t})^2 \psi_t \, d\mu + \int g_t (L - \partial_t) \psi_t \, d\mu, \]
where \( g_t = f_t / \psi_t \). In our application we set \( \psi_t = \psi = \omega / \mu \), hence we have
\[ \partial_t S_\omega (g_t) = -\frac{2}{N} \sum_j \int (\partial_j \sqrt{\omega t})^2 \, d\omega + \int \tilde{L} g_t \, d\omega + \sum_j \int b_j \partial_j g_t \, d\omega. \]

Since \( \omega \) is invariant, the middle term on the right hand side vanishes, and from the Schwarz inequality
\[ \partial_t S_\omega (g_t) \leq -D_\omega (\sqrt{g_t}) + C N \sum_j \int b_j^2 g_t \, d\omega. \]

Together with (6.4) and (5.5), we have
\[ \partial_t S_\omega (g_t) \leq -C \eta^{-1/3} S_\omega (g_t) + C \eta^{-2} \Lambda. \]

which, after integrating it from \( t = 0 \) to \( \tau /2 \), proves the first inequality in (6.9). The second inequality can be obtained from integrating (6.10) from \( t = \tau /2 \) to \( t = \tau \) and using the monotonicity of the Dirichlet form in time.

\[ \square \]
Finally, we sketch the proof of Theorem 5.1. With the choice of $\tau = \eta^{1/3}N^{\varepsilon}$ and $q_0 = f_{\tau}/\psi$, Theorems 6.1, 6.2 and 6.3 directly imply

$$\left| \int \frac{1}{N} \sum_{i \in A} G(N(x_i - x_{i+n})) f_{\tau} d\mu - \int \frac{1}{N} \sum_{i \in A} G(N(x_i - x_{i+n})) d\omega \right| \leq \sqrt{\frac{CA}{N^{1-\varepsilon}} + Ce^{-cN^{\varepsilon}}},$$

(6.12)

i.e. the local statistics of $f_{\tau}\mu$ and $\omega$ are close. Clearly, equation (6.12) also holds for the special choice $f_0 = 1$ (for which $f_\tau = 1$), i.e. local statistics of $\mu$ and $\omega$ can also be compared. This completes the proof of Theorem 5.1. 

\[ \square \]

References

[1] Aizenman, M., and Molchanov, S.: Localization at large disorder and at extreme energies: an elementary derivation, Commun. Math. Phys. 157, 245–278 (1993)

[2] Anderson, P.: Absences of diffusion in certain random lattices, Phys. Rev. 109, 1492–1505 (1958)

[3] Bakry, D., Émery, M.: Diffusions hypercontractives. in: Séminaire de probabilités, XIX, 1983/84, 1123 Lecture Notes in Mathematics, Springer, Berlin, 1985, 177–206.

[4] Ben Arous, G., Peché, S.: Universality of local eigenvalue statistics for some sample covariance matrices. Comm. Pure Appl. Math. LVIII. (2005), 1–42.

[5] Bleher, P., Its, A.: Semiclassical asymptotics of orthogonal polynomials, Riemann-Hilbert problem, and universality in the matrix model. Ann. of Math. 150 (1999): 185–266.

[6] Brézin, E., Hikami, S.: Correlations of nearby levels induced by a random potential. Nucl. Phys. B 479 (1996), 697–706, and Spectral form factor in a random matrix theory. Phys. Rev. E 55 (1997), 4067–4083.

[7] Chen, T.: Localization lengths and Boltzmann limit for the Anderson model at small disorders in dimension 3. J. Stat. Phys. 120, no.1-2, 279–337 (2005).

[8] Curto, R., Fialkow, L.: Recursiveness, positivity and truncated moment problems. Houston J. Math. 17, no. 4., 603-635 (1991).

[9] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: 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):1335–1425.

[10] Deift, P., Kriecherbauer, T., McLaughlin, K.T-R, Venakides, S., Zhou, X.: Strong asymptotics of orthogonal polynomials with respect to exponential weights. Comm. Pure Appl. Math. 52 (1999): 1491–1552.

[11] Diestorti, M., Pinson, H., Spencer, T.: Density of states for random band matrices. Commun. Math. Phys. 232, 83–124 (2002)
[12] Dyson, F.J.: A Brownian-motion model for the eigenvalues of a random matrix. J. Math. Phys. 3, 1191-1198 (1962).

[13] Dyson, F.J.: Correlations between eigenvalues of a random matrix. Commun. Math. Phys. 19, 235-250 (1970).

[14] L. Erdős, M. Salmhofer, H.-T. Yau, Quantum diffusion for the Anderson model in scaling limit. Ann. Inst. H. Poincare 8 no. 4, 621-685 (2007)

[15] Erdős, L., Schlein, B., Yau, H.-T.: Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. Ann. Probab. 37, No. 3, 815–852 (2008)

[16] Erdős, L., Schlein, B., Yau, H.-T.: Local semicircle law and complete delocalization for Wigner random matrices. Commun. Math. Phys. 287, 641–655 (2009)

[17] Erdős, L., Schlein, B., Yau, H.-T.: Wegner estimate and level repulsion for Wigner random matrices. To appear in Inv. Math. Preprint arxiv.org/abs/0811.2591

[18] Erdős, L., Schlein, B., Yau, H.-T.: Universality of random matrices and local relaxation flow. Submitted to Inv. Math. arxiv.org/abs/0907.5605

[19] Erdős, L., Ramírez, J., Schlein, B., Yau, H.-T.: Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. Submitted to Electr. J. Prob. arxiv.org/abs/0905.2089

[20] L. Erdős, S. Péché, J. Ramírez, B. Schlein and H-T. Yau, Bulk universality for Wigner matrices. Submitted to Comm. Pure Appl. Math. Preprint arXiv.org:0905.4176.

[21] L. Erdős, J. Ramírez, B. Schlein, T. Tao, V. Vu and H-T. Yau, Bulk universality for Wigner hermitian matrices with subexponential decay. To appear in Math. Res. Letters. Preprint arXiv:0906.4400

[22] J. Fröhlich and T. Spencer, Absence of diffusion in the Anderson tight binding model for large disorder or low energy, Commun. Math. Phys. 88, 151–184 (1983)

[23] Guionnet, A.: Large random matrices: Lectures on Macroscopic Asymptotics. École d’Été de Probabilités de Saint-Flour XXXVI-2006. Springer.

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

[25] Marchenko, V.A., Pastur, L.: The distribution of eigenvalues in a certain set of random matrices. Mat. Sb. 72, 507–536 (1967).

[26] Mehta, M.L.: Random Matrices. Academic Press, New York, 1991.

[27] Mehta, M.L., Gaudin, M.: On the density of eigenvalues of a random matrix. Nuclear Phys. 18, 420-427 (1960).

[28] Minami, N.: Local fluctuation of the spectrum of a multidimensional Anderson tight binding model. Commun. Math. Phys. 177, 709–725 (1996).

[29] Pastur, L., Shcherbina M.: Bulk universality and related properties of Hermitian matrix models. J. Stat. Phys. 130 (2008), no.2., 205-250.
[30] Schenker, J.: Eigenvector localization for random band matrices with power law band width. *Commun. Math. Phys.* **290**, 1065-1097 (2009)

[31] Schenker, J. and Schulz-Baldes, H.: Semicircle law and freeness for random matrices with symmetries or correlations. *Math. Res. Letters* **12**, 531-542 (2005)

[32] Schenker, J. and Schulz-Baldes, H.: Gaussian fluctuations for random matrices with correlated entries. *Int. Math. Res. Not. IMRN* 2007, **15**, Art. ID rnm047.

[33] Sinai, Y. and Soshnikov, A.: A refinement of Wigner’s semicircle law in a neighborhood of the spectrum edge. *Functional Anal. and Appl.* **32** (1998), no. 2, 114–131.

[34] Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** (1999), no.3. 697-733.

[35] Tao, T. and Vu, V.: Random matrices: Universality of the local eigenvalue statistics. Preprint arXiv:0906.0510.

[36] Tao, T. and Vu, V.: Random matrices: Universality of local eigenvalue statistics up to the edge. Preprint arXiv:0908.1982

[37] Vu, V.: Spectral norm of random matrices. *Combinatorica*, **27** (6) (2007), 721-736.

[38] Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.* **62** (1955), 548-564.

[39] Yau, H. T.: Relative entropy and the hydrodynamics of Ginzburg-Landau models, *Lett. Math. Phys.* **22** (1991) 63–80.