The local relaxation flow approach to universality of the local statistics for random matrices

László Erdős[1,*], Benjamin Schlein[2] Horng-Tzer Yau[3,†] and Jun Yin[3]

Institute of Mathematics, University of Munich,
Theresienstr. 39, D-80333 Munich, Germany
lerdos@math.lmu.de [1]

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge
Wilberforce Rd, Cambridge CB3 0WB, UK
b.schlein@dpmms.cam.ac.uk [2]

Department of Mathematics, Harvard University
Cambridge MA 02138, USA
htyau@math.harvard.edu, jyin@math.harvard.edu [3]

Aug 17, 2010

Abstract

We present a generalization of the method of the local relaxation flow to establish the universality of local spectral statistics of a broad class of large random matrices. We show that the local distribution of the eigenvalues coincides with the local statistics of the corresponding Gaussian ensemble provided the distribution of the individual matrix element is smooth and the eigenvalues $\{x_j\}_{j=1}^N$ are close to their classical location $\{\gamma_j\}_{j=1}^N$ determined by the limiting density of eigenvalues. Under the scaling where the typical distance between neighboring eigenvalues is of order $1/N$, the necessary apriori estimate on the location of eigenvalues requires only to know that $\mathbb{E}|x_j - \gamma_j|^2 \leq N^{-1-\varepsilon}$ on average. This information can be obtained by well established methods for various matrix ensembles. We demonstrate the method by proving local spectral universality for sample covariance matrices.

[*]Partially supported by SFB-TR 12 Grant of the German Research Council
[†]Partially supported by NSF grants DMS-0757425, 0804279

Nous présentons une généralisation de la méthode du *flot de relaxation locale* servant à établir l'universalité des statistiques spectrales locales d'une vaste classe de grandes matrices aléatoires. Nous démontrons que la distribution locale des valeurs propres coïncide avec celle de l'ensemble gaussien pourvu que la loi des coefficients individuels de la matrice soit lisse et que les valeurs propres $\{x_j\}_{j=1}^N$ soient près de leurs quantiles classiques $\{\gamma_j\}_{j=1}^N$ determinées par la densité limite des valeurs propres. Dans la normalisation où la distance typique entre les valeurs propres voisines est d'ordre $1/N$, la borne a priori nécessaire sur la position des valeurs propres nécessite uniquement l'établissement de $\mathbb{E}|x_j - \gamma_j|^2 \le N^{-1-\varepsilon}$ en moyenne. Cette information peut être obtenue par des méthodes bien établies pour divers ensembles de matrices. Nous illustrons la méthode en démontrant l'universalité spectrale locale pour des matrices de covariance.

AMS Subject Classification (2010): 15B52, 82B44

Running title: Local relaxation flow

Keywords: Random matrix, sample covariance matrix, Wishart matrix, Wigner-Dyson statistics

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. In particular they coincide with that of the corresponding Gaussian ensemble. The most commonly studied ensembles are

(i) hermitian, symmetric and quaternion self-dual matrices with identically distributed and centered entries that are independent (subject to the natural restriction of the symmetry);

(ii) sample covariance matrices of the form $H = A^*A$, where A is an $M \times N$ matrix with centered real or complex i.i.d. entries.

There are two types of universalities: the edge universality and the bulk universality concerning energy levels near the spectral edges and in the interior of the spectrum, respectively. Since the works of Sinai and Soshnikov [35, 37], the edge universality is commonly approached via the fairly robust moment method [33, 22, 36, 38, 34]; very recently an alternative approach was given in [40].

The bulk universality is a subtler problem. In the simplest case of the hermitian Wigner ensemble, it states that, independent of the distribution of the entries, the local k-point correlation functions of the eigenvalues (see (2.3) for the precise definition later), after appropriate rescaling and in the $N \to \infty$ limit, are given by the determinant of the *sine kernel*

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

Similar statement is expected to hold for all other ensembles mentioned above but the explicit formulas are somewhat more complicated. Detailed formulas for the different Wigner ensembles can be found e.g., in [30]. The various sample covariance ensembles have the same local statistics for their *singular values* as the local eigenvalue statistics of the corresponding Wigner ensembles.

For ensembles of hermitian, symmetric or quaternion self-dual matrices that remain invariant under the transformations $H \to U^* H U$ for any unitary, orthogonal or symplectic matrix U, respectively, the joint probability density function of all the N eigenvalues can be explicitly computed. These ensembles are typically given by the probability density

$$P(H)\mathrm{d}H \sim \exp(-N \mathrm{Tr}\, V(H))\mathrm{d}H, \tag{1.2}$$

where V is a real function with sufficient growth at infinity and $\mathrm{d}H$ is the flat Lebesgue measure on the corresponding symmetry class of matrices. 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 with distribution (1.2) can be computed explicitly:

$$f(x_1, x_2, \ldots, x_N) = (\text{const.}) \prod_{i<j} |x_j - x_i|^\beta \prod_{j=1}^N e^{-N \sum_{j=1}^N V(x_j)}, \tag{1.3}$$

where $\beta = 1, 2, 4$ for hermitian, symmetric and symplectic ensembles, respectively, and const. is a normalization factor. The formula (1.3) defines a joint probability density of N real random variables for any $\beta \geq 1$ even when there is no underlying matrix ensemble. This ensemble is called the *invariant β-ensemble*. Quadratic V corresponds to the Gaussian ensembles; we note that these are the only ensembles that are simultaneously invariant and have i.i.d. matrix entries. These are called the Gaussian Orthogonal, Unitary and Symplectic Ensembles (GOE, GUE, GSE for short) in case of $\beta = 1, 2, 4$, respectively. Somewhat different choices of V lead to two other classical ensembles, the Laguerre and the Jacobi ensembles, that also have matrix interpretation for $\beta = 1, 2, 4$ (e.g., the Laguerre ensemble corresponds to the Gaussian sample covariance matrices which are also called *Wishart matrices*), see [11, 23] for more details. The local statistics can be obtained via a detailed analysis of orthogonal polynomials on the real line with respect to the weight function $\exp(-V(x))$. This approach was originally applied to classical ensembles by Dyson [13], Mehta and Gaudin [31] and Mehta [30] that lead to classical orthogonal polynomials. Later general methods using orthogonal polynomials were developed to tackle a very general class of invariant ensembles by Deift *et.al.*, see [7, 8, 9, 10] and references therein, and also by Bleher and Its [5] and Pastur and Schcherbina [32].

Many natural matrix ensembles are typically not unitarily invariant; the most prominent examples are the Wigner matrices or the sample covariance matrices mentioned in (i) and (ii). For these ensembles, apart from the identically distributed Gaussian case, 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 recently in [14], by Tao and Vu in [39] and in [15]. These works rely on the Wigner matrices with Gaussian divisible distribution, i.e., ensembles of the form

$$\widehat{H} + \sqrt{s}V, \tag{1.4}$$

where $\widehat{H}$ is a Wigner matrix, V is an independent standard GUE matrix and s is a positive constant. Johansson [26] (see also Ben Arous and Péché [3] and the recent paper [27]) proved the bulk universality for the eigenvalues of such matrices by an asymptotic analysis on an *explicit* formula for the correlation functions adapted from Brézin-Hikami [6]. Unfortunately, the similar formula for symmetric or quaternion self-dual Wigner matrices, as well as for real sample covariance matrices, is not very explicit and the technique of [3, 14, 26] cannot be extended to prove universality. Complex sample covariance matrices can however be handled with an analogous formula [3] and universality without any Gaussian component is a work in progress [4].

A key observation of Dyson is that if the matrix $\widehat{H} + \sqrt{s}V$ is embedded into a stochastic matrix flow, i.e. one considers $\widehat{H} + V(s)$ where the matrix elements of $V(s)$ are independent standard Brownian motions with variance s/N, then the evolution of the eigenvalues is given by a system of coupled stochastic differential equations (SDE), 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. Similar stochastic processes can be constructed for symmetric, quaternion self-dual and sample covariance type matrices, and, in fact, on the level of eigenvalue SDE they can be extended to other values of β (see (5.5) and (5.8) for the precise formulas).

The result of [26, 3] 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 $s \gg N^{-3/4}$ [16]. The key observation of [16] is that the local statistics of eigenvalues depend exclusively on the approach to local equilibrium which in general is faster than reaching the global equilibrium. Unfortunately, the identification of local equilibria in [16] still uses explicit representations of correlation functions by orthogonal polynomials (following e.g. [32]), and the extension to other ensembles is not a simple task.

In [20] we introduced an approach based on a new stochastic flow, the *local relaxation flow*, which locally behaves like DBM, but has a faster decay to equilibrium. This method completely circumvented explicit formulas and it resulted in proving universality for *symmetric* Wigner matrices (the method applies to hermitian and quaternion self-dual Wigner matrices as well). As an input of this method, we needed a fairly detailed control on the local density of eigenvalues that could be obtained from our previous works on Wigner matrices [17, 18, 19].

In this paper we will prove a general theorem which states that as long as the eigenvalues are at most $N^{-1/2-\varepsilon}$ distance near their classical location on average, the local statistics is universal and in particular it coincides with the Gaussian case for which explicit formulas have been computed. To introduce this flow, denote by γ_j the location of the j-th eigenvalue that will be defined in

(2.12). We first define the *pseudo equilibrium measure* by

$$\omega_N = C_N \exp\big(-NW\big)\mu_N, \quad W(\mathbf{x}) = \sum_{j=1}^N W_j(x_j), \qquad W_j(x) = \frac{1}{2R^2}(x_j-\gamma_j)^2, \tag{1.5}$$

where μ_N is the probability measure for the eigenvalue distribution of the corresponding Gaussian ensemble. In case of Wigner matrices, μ_N is the measure for the general β ensemble ($\beta \geq 1$ and $\beta = 2$ for GUE):

$$\mu = \mu_N(\mathrm{d}\mathbf{x}) = \frac{e^{-\mathcal{H}(\mathbf{x})}}{Z_\beta}\mathrm{d}\mathbf{x}, \qquad \mathcal{H}(\mathbf{x}) = N\left[\beta\sum_{i=1}^N \frac{x_i^2}{4} - \frac{\beta}{N}\sum_{i<j}\log|x_j - x_i|\right]. \tag{1.6}$$

In this setting, it is natural to view eigenvalues as random points and their equilibrium measure as Gibbs measure with a Hamiltonian $\mathcal{H}$. We will freely use the terminology of statistical mechanics. Note that the additional term W_j in ω_N confines the j-th point x_j near its classical location, but the probability w.r.t. the equilibrium measure μ_N of the event that x_j near its classical location will be shown to be very close to 1. Furthermore, we will prove that the local statistics of the measures ω_N and μ_N are identical in the limit $N \to \infty$ and this justifies the term pseudo equilibrium measure.

The local relaxation flow is defined to be the reversible flow (or the gradient flow) generated by the pseudo-equilibrium measure. The main advantage of the local relaxation flow is that it has a faster decay to global equilibrium (Theorem 4.2) compared with the DBM. The idea behind this construction can be related to the treatment of metastability in statistical physics. Imagine that we have a double well potential and we wish to treat the dynamics of a particle in one of the two wells. Up to a certain time, say t_0, the particle will be confined in the well where the particle initially located. However, the potential of this particle, given by the double well, is not convex. A naive idea is to regain the convexity before the time t_0 is to modify the potential to be a single well! Now as long as we can prove that the particle was confined in the initial well up to t_0, there is no difference between these two dynamics. But the modified dynamics, being w.r.t. a convex potential, can be estimated much more precisely and this estimate can be carried over to the original dynamics up to the time t_0.

In our case, the convexity of the equilibrium measure μ_N is rather weak and in fact, it comes from the quadratic confining potential $\beta x_i^2/4$ of (1.6). So the potential is convex, just not "convex enough". There is no sharp transition like jumping from one metastable state to another as in the double well case. Instead, there are two time scales: in short time the local equilibrium is formed, on longer time, it approaches the global equilibrium. The approach to the local equilibrium is governed by a strong intrinsic convexity in certain directions due to the interactions (see (2.10) later for a precise formula). To reveal this additional convexity, in our previous paper [20] we introduced a pseudo equilibrium measure where we replaced the long range part of the interaction by a mean-field potential term using the classical locations of far away particles. This potential term inherited the intrinsic convexity of the interaction and it could be directly used to enhance the decay to the local statistics. One technical difficulty with this approach was that we needed to handle the singular behavior of the logarithmic interaction potential. In this paper we show that the pseudo equilibrium measure can be defined by adding a Gaussian term. This simple modification turns out to be sufficient and is also model-independent. Since the Gaussian modification is regular, we no longer need to deal with singularities. The price to pay is that we need a slightly stronger local semicircle law which will be treated in Section 8.

The method of local relaxation flow itself proves universality for Wigner matrices with a small Gaussian component $\sqrt{s}V$ (typically of variance $s \geq N^{-\gamma}$ with some $0 < \gamma < 1$). In other words, we can prove universality for a Wigner ensemble whose single entry distribution (the distribution of its matrix elements) is given by $e^{tB}u_0$, where B is the generator of the Ornstein-Uhlenbeck process and u_0 is any initial distribution (We remark that in our approach of decay to equilibrium, the Brownian motion in the construction of DBM is always replaced by the Ornstein-Uhlenbeck process). To obtain universality for Wigner matrices without any Gaussian component, it remains to prove that for a given Wigner matrix ensemble with a single entry distribution ν we can find u_0 and t such that the eigenvalue distributions of the ensembles given by ν and $e^{tB}u_0$ are very close to each other. By the *method of reverse heat flow* introduced in [14], we choose u_0 to be an approximation of $e^{-tB}\nu$. Although the Ornstein-Uhlenbeck evolution cannot be reversed, we can approximately reverse it provided that ν is sufficient smooth and the time is short. This enables us to compare local statistics of Wigner ensembles with and without small Gaussian components assuming that the single entry distribution is sufficiently smooth (see Section 6).

As an application, we will use this method to prove the bulk universality of sample covariance ensembles. The necessary apriori control on the location of eigenvalues will be obtained by a local semicircle law. In addition to sample covariance ensembles, we will outline the modifications needed for proving the bulk universality of symplectic ensembles.

2 Universality for the local relaxation flow

In this section, we consider the following general setup. Suppose $\mu = e^{-N\mathcal{H}}/Z$ is a probability measure on the configuration space $\mathbb{R}^N$ characterized by some Hamiltonian $\mathcal{H} : \mathbb{R}^N \to \mathbb{R}$, where $Z = \int e^{-N\mathcal{H}(\mathbf{x})}\mathrm{d}\mathbf{x} < \infty$ is the normalization. We will always assume that $\mathcal{H}$ is symmetric under permutation of the variables $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$.

We consider time dependent permutational symmetric probability measures with density $f_t(\mathbf{x})$, $t \geq 0$, with respect to the measure $\mu(\mathrm{d}\mathbf{x}) = \mu(\mathbf{x})\mathrm{d}\mathbf{x}$. The dynamics is characterized by the forward equation

$$\partial_t f_t = L f_t, \qquad t \geq 0, \tag{2.1}$$

with a given permutation symmetric initial data f_0. Here the generator L is defined via the Dirichlet form as

$$D(f) = D_\mu(f) = -\int f L f \mathrm{d}\mu = \sum_{j=1}^N \frac{1}{2N} \int (\partial_j f)^2 \mathrm{d}\mu, \quad \partial_j = \partial_{x_j}. \tag{2.2}$$

Formally, we have $L = \frac{1}{2N}\Delta - \frac{1}{2}(\nabla\mathcal{H})\nabla$. In Appendix A we will show that under general conditions on $\mathcal{H}$ the generator can be defined as a self-adjoint operator on an appropriate domain and the dynamics is well defined for any $f_0 \in L^1(\mathrm{d}\mu)$ initial data. Strictly speaking, we will consider a sequence of Hamiltonians $\mathcal{H}_N$ and corresponding dynamics L_N and $f_{t,N}$ parametrized by N, but the N-dependence will be omitted. All results will concern the $N \to \infty$ limit.

The expectation with respect to the density f_t will be denoted by $\mathbb{E}_t$ with $\mathbb{E} := \mathbb{E}_0$. The expectation with respect to the equilibrium measure μ is denoted by $\mathbb{E}^\mu$. For any $n \geq 1$ we define the n-point correlation functions (marginals) of the probability measure $f_t \mathrm{d}\mu$ by

$$p_{t,N}^{(n)}(x_1, x_2, \dots, x_n) = \int_{\mathbb{R}^{N-n}} f_t(\mathbf{x})\mu(\mathbf{x})\mathrm{d}x_{n+1}\dots\mathrm{d}x_N. \tag{2.3}$$

With a slight abuse of notations, we will sometimes also use μ to denote the density of the measure μ with respect to the Lebesgue measure. The correlation functions of the equilibrium measure are denoted by

$$p_{\mu,N}^{(n)}(x_1, x_2, \dots, x_n) = \int_{\mathbb{R}^{N-n}} \mu(\mathbf{x})\mathrm{d}x_{n+1}\dots\mathrm{d}x_N.$$

We now list our main assumptions on the initial distribution f_0 and on its evolution f_t. We first define the subdomain

$$\Sigma_N := \big\{\mathbf{x} \in \mathbb{R}^N,\ x_1 < x_2 < \dots < x_N\big\} \tag{2.4}$$

of ordered sets of points $\mathbf{x}$. In the application to the sample covariance matrices, we will use the subdomain

$$\Sigma_N^+ := \big\{\mathbf{x} \in \mathbb{R}^N,\ 0 < x_1 < x_2 < \dots < x_N\big\} \tag{2.5}$$

of ordered sets of positive points.

Assumption I. The Hamiltonian $\mathcal{H}$ of the equilibrium measure has the form

$$\mathcal{H} = \mathcal{H}_N(\mathbf{x}) = \beta\Big[\sum_{j=1}^N U(x_j) - \frac{1}{N}\sum_{i<j}\log|x_i - x_j|\Big], \tag{2.6}$$

where $\beta \geq 1$. The function $U : \mathbb{R} \to \mathbb{R}$ is smooth with $U'' \geq 0$ and

$$U(x) \geq C|x|^\delta \qquad \text{for some } \delta > 0 \text{ and } |x| \text{ large.} \tag{2.7}$$

The condition $U'' \geq 0$ can be relaxed to $\inf U'' > -\infty$, see remark after (4.11).

Alternatively, in order to discuss the case of the sample covariance matrices, we will also consider the following modification of Assumption I.

Assumption I'. The Hamiltonian $\mathcal{H}$ of the equilibrium measure has the form

$$\mathcal{H} = \mathcal{H}_N(\mathbf{x}) = \beta\Big[\sum_{j=1}^N U(x_j) - \frac{1}{N}\sum_{i<j}\log|x_i - x_j| - \frac{1}{N}\sum_{i<j}\log|x_i + x_j| - \frac{c_N}{N}\sum_j \log|x_j|\Big], \qquad (2.8)$$

where $\beta \geq 1$ and $c_N \geq 1$. The function U satisfies the same conditions as in Assumption I.

It is easy to check that the condition (2.7) guarantees that the following bound holds for the normalization constant

$$|\log Z| \leq CN^m \qquad (2.9)$$

with some exponent m depending on δ.

In Appendix A we will show that for $\beta \geq 1$ the dynamics (2.1) can be restricted to the subdomains Σ_N or Σ_N^+, respectively, i.e, the ordering will be preserved under the dynamics. In the sequel we will thus assume that f_t is a probability measure on Σ_N or Σ_N^+. We continue to use the notation f and μ for the restricted measure. Note that the correlation functions $p^{(k)}$ from (2.3) are still defined on $\mathbb{R}^k$, i.e., their arguments remain unordered.

It follows from Assumption I (or I') that the Hessian matrix of $\mathcal{H}$ satisfies the following bound:

$$\big\langle \mathbf{v}, \nabla^2\mathcal{H}(\mathbf{x})\mathbf{v}\big\rangle \geq \frac{\beta}{N}\sum_{i<j}\frac{(v_i - v_j)^2}{(x_i - x_j)^2}, \qquad \mathbf{v} = (v_1, \ldots, v_N) \in \mathbb{R}^N, \quad \mathbf{x} \in \Sigma_N \quad (\text{or } \mathbf{x} \in \Sigma_N^+). \qquad (2.10)$$

This convexity bound is the key assumption; our method works for a broad class of general Hamiltonians as long as (2.10) holds. In particular, an arbitrary many-body potential function $V(\mathbf{x})$ can be added to the Hamiltonians (2.6), (2.8), as long as V is convex on the open sets Σ_N and Σ_N^+, respectively. The argument in the proof of the main Theorem 2.1 remains unchanged, but the technical details of the regularization of the singular dynamics (Appendix B) becomes more involved. We do not pursue this direction here since we do not need it for the application for Wigner and sample covariance matrices.

Assumption II. There exists a continuous, compactly supported density function $\varrho(x) \geq 0$, $\int_{\mathbb{R}} \varrho = 1$, on the real line, independent of N, such that for any fixed $a, b \in \mathbb{R}$

$$\lim_{N\to\infty}\sup_{t\geq 0}\left|\int \frac{1}{N}\sum_{j=1}^N \mathbf{1}(x_j \in [a,b]) f_t(\mathbf{x})\mathrm{d}\mu(\mathbf{x}) - \int_a^b \varrho(x)\mathrm{d}x\right| = 0. \qquad (2.11)$$

Let $\gamma_j = \gamma_{j,N}$ denote the location of the j-th point under the limiting density, i.e., γ_j is defined by

$$N\int_{-\infty}^{\gamma_j} \varrho(x)\mathrm{d}x = j, \qquad 1 \leq j \leq N, \quad \gamma_j \in \text{supp}\varrho. \qquad (2.12)$$

We will call γ_j the *classical location* of the j-th point. Note that γ_j may not be uniquely defined if the support of ϱ is not connected but in this case the next Assumption III will not be satisfied anyway.

Assumption III. There exists an $\varepsilon > 0$ such that

$$\sup_{t \geq N^{-2\varepsilon}} \int \frac{1}{N} \sum_{j=1}^{N} (x_j - \gamma_j)^2 f_t(\mathrm{d}\mathbf{x}) \mu(\mathrm{d}\mathbf{x}) \leq C N^{-1-2\varepsilon} \tag{2.13}$$

with a constant C uniformly in N.

Under Assumption II, the typical spacing between neighboring points is of order $1/N$ away from the spectral edges, i.e., in the vicinity of any energy E with $\varrho(E) > 0$. Assumption III guarantees that typically the random points x_j remain in the $N^{-1/2-\varepsilon}$ vicinity of their classical location.

The final assumption is an upper bound on the local density. For any $I \in \mathbb{R}$, let

$$\mathcal{N}_I := \sum_{i=1}^{N} \mathbf{1}(x_i \in I)$$

denote the number of points in I.

Assumption IV. For any compact subinterval $I_0 \subset \{E \;:\; \varrho(E) > 0\}$, and for any $\delta > 0$, $\sigma > 0$ there are constants C_n, $n \in \mathbb{N}$, depending on I_0, and σ such that for any interval $I \subset I_0$ with $|I| \geq N^{-1+\sigma}$ and for any $K \geq 1$, we have

$$\sup_{\tau \geq N^{-2\varepsilon}} \int \mathbf{1}\big\{\mathcal{N}_I \geq KN|I|\big\} f_\tau \mathrm{d}\mu \leq C_n K^{-n}, \qquad n = 1, 2, \ldots, \tag{2.14}$$

where ε is the exponent from Assumption III.

The main general theorem is the following:

Theorem 2.1 *Suppose that the Hamiltonian given in* (2.6) *or* (2.8) *satisfy Assumption I or I', respectively. Suppose that Assumptions II, III and IV hold for the solution f_t of the forward equation* (2.1). *Assume that at time $t_0 = N^{-2\varepsilon}$ we have $S_\mu(f_{t_0}) := \int f_{t_0} \log f_{t_0} \mathrm{d}\mu \leq CN^m$ with some fixed exponent m that may depend on ε. Let $E \in \mathbb{R}$ and $b > 0$ such that $\min\{\varrho(x) \;:\; x \in [E-b, E+b]\} > 0$. Then for any $\delta > 0$, $\varepsilon' > 0$, for any integer $n \geq 1$ and for any compactly supported continuous test function $O : \mathbb{R}^n \to \mathbb{R}$, we have,*

$$\begin{aligned} \sup_{t \geq \tau} \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int_{\mathbb{R}^n} \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_n \, O(\alpha_1, \ldots, \alpha_n) \frac{1}{\varrho(E)^n} \\ \times \Big(p_{t,N}^{(n)} - p_{\mu,N}^{(n)} \Big) \Big(E' + \frac{\alpha_1}{N\varrho(E)}, \ldots, E' + \frac{\alpha_n}{N\varrho(E)} \Big) \leq CN^{2\varepsilon'} \Big[b^{-1} N^{-\frac{1+2\varepsilon}{3}} + b^{-1/2} N^{-\delta/2} \Big] \end{aligned} \tag{2.15}$$

for $\tau = N^{-2\varepsilon+\delta}$ where $\varepsilon > 0$ is the exponent from Assumption III.

Suppose in addition to the Assumption I-IV, that there exists an $A > 0$ such that, for any $c' > 0$

$$\mathbb{P}\Big(\sup_{c'N \leq j \leq (1-c')N} |x_j - \gamma_j| \geq N^{-1+A} \Big) \leq C N^{-c \log\log N} \tag{2.16}$$

for some constants c and C only depending on c'. Then for $\tau = N^{-2\varepsilon+\delta}$ we have

$$\begin{aligned} &\sup_{t \geq \tau} \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int_{\mathbb{R}^n} \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_n \, O(\alpha_1, \ldots, \alpha_n) \frac{1}{\varrho(E)^n} \\ &\times \Big(p_{t,N}^{(n)} - p_{\mu,N}^{(n)} \Big) \Big(E' + \frac{\alpha_1}{N\varrho(E)}, \ldots, E' + \frac{\alpha_k}{N\varrho(E)} \Big) \leq C_n N^{2\varepsilon'} \Big[b^{-1} N^{-1+A} + b^{-1/2} N^{-\delta/2} \Big]. \end{aligned} \tag{2.17}$$

This theorem shows that the local statistics of the points x_j in the bulk with respect to the time evolved distribution f_t coincides with the local statistics with respect to the equilibrium distribution μ as long as $t \gg N^{-2\varepsilon}$. In many applications, the local equilibrium statistics can be explicitly computed and in the $b \to 0$ limit it becomes independent of E, in particular this is the case for the classical matrix ensembles (see next section). The restriction on the time $t \gg N^{-2\varepsilon}$ will be removed by the reverse heat flow argument (see Section 6) for matrix ensembles.

Since the eigenvalues fluctuate at least on a scale $1/N$, the best possible exponent in Assumption III is $2\varepsilon \sim 1$, but we will only be able to prove it for some $\varepsilon > 0$ for the ensembles considered in this paper. Similarly, the optimal exponent in (2.16) is $A \sim 0$. If we use these optimal estimates, $2\varepsilon \sim 1$, $A \sim 0$, and we choose $\delta = 2\varepsilon \sim 1$, thus $\tau \sim 1$, then we can choose $b \sim N^{-1}$, i.e., we obtain the universality with essentially no averaging in E. On the other hand, the error estimate is the strongest, of order $\sim N^{-1/2}$, for an averaging on an energy window of size $b \sim 1$. These errors become weaker if time τ is reduced. These considerations are not important in this paper, but will be useful when good estimates on ε and A can be obtained.

Convention: Throughout the paper the letters C, c denote positive constants whose values may change from line to line and they are independent of the relevant parameters. Since we will always take the $N \to \infty$ limit at the end, all estimates are understood for sufficiently large N.

3 Universality for Matrix Ensembles

Now we specialize Theorem 2.1 to Wigner and sample covariance matrices with i.i.d. entries. In the next sections we give the precise definitions of these ensembles; formulas for the equilibrium measure and the dynamics will be deferred until Section 5.

In order to apply Theorem 2.1 to Wigner and sample covariance matrices, we need to check that Assumptions I-IV are satisfied for these ensembles. Assumptions I or I' are satisfied by the definition of the Hamiltonian, the precise formulas are given in Section 5. Assumption II is satisfied since the density of eigenvalues is given by the Wigner semicircle law (3.7) for Wigner matrices [43].

In case of the sample covariance matrices, the singular values of A will play the role of x_j's and their density is given by the Marchenko-Pastur law (3.14) after an obvious transformation (3.15) [29]. In fact, in Section 8 we prove a local version of the Marchenko-Pastur law in analogy with our previous work on the local semicircle law for Wigner matrices [17, 18]. In Section 9 (Theorem 9.1) we will show that Assumption III is satisfied for these ensembles (more precisely, we will prove that Assumption III is satisfied for sample covariance matrices; the proof for Wigner matrices is analogous, and will not be given in details). Assumption IV will be proved in Lemma 8.1 for the sample covariance matrices, for Wigner matrices the proof was given, e.g., in Theorem 4.6 of [19]. We remark that the assumption that the matrix entries are identically distributed, will only be used in checking Assumptions III and IV. Assumption II holds under much more general conditions on the matrix entries. Finally, the apriori estimate on the entropy $S_\mu(f_{t_0})$ follows from the smoothing property of the OU-flow (see Section 5).

3.1 Definition of the Wigner matrix

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

$$h_{\ell k} = h_{k\ell} := N^{-1/2} x_{\ell k}, \quad k < \ell, \tag{3.1}$$

where $x_{\ell k}$ for $\ell < k$ are independent, identically distributed real random variables with distribution ν that has zero expectation and variance 1. The diagonal elements are $h_{kk} = N^{-1/2} x_{kk}$, where x_{kk} are also i.i.d. with distribution $\widetilde{\nu}$ that has zero expectation and variance 2. The eigenvalues of H will be denoted by $x_1 < x_2 < \ldots < x_N$. We will always assume that the distribution ν is continuous hence the eigenvalues are simple with probability one.

In the *hermitian* case we assume that

$$h_{\ell k} = \bar{h}_{k\ell} := N^{-1/2}(x_{\ell k} + i y_{\ell k}), \quad k < \ell, \tag{3.2}$$

where $x_{\ell k}$ and $y_{\ell k}$ are real i.i.d. random variables distributed with the law ν with zero expectation and variance $\frac{1}{2}$. The diagonal elements h_{kk} are real, centered and they have variance one with law $\widetilde{\nu}$. The eigenvalues of H are again denoted by $x_1 < x_2 < \ldots < x_N$.

Finally, for the *quaternion self-dual* case we assume that H is a $2N$ by $2N$ complex matrix that can be viewed as an $N \times N$ matrix with elements consisting of 2×2 blocks of the form

$$\begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}, \tag{3.3}$$

where $z = a + bi, w = c + di$ are arbitrary complex numbers, $a, b, c, d \in \mathbb{R}$. Such a 2 by 2 matrix can be identified with the quaternion $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \in \mathbb{H}$ if the quaternion basis elements $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are identified with the standard Pauli matrices

$$\mathbf{i} = i\sigma_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad \mathbf{j} = i\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \mathbf{k} = i\sigma_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

The complex numbers $z \in \mathbb{C}$ can be naturally identified with diagonal quaternions via the identification

$$z \cong \begin{pmatrix} z & 0 \\ 0 & \bar{z} \end{pmatrix}. \tag{3.4}$$

The dual of the quaternion q is defined to be $q^+ := a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}$ which corresponds to the hermitian conjugate of the matrix (3.3).

Using this identification, H can be viewed as an $N \times N$ matrix with quaternion entries. The matrix H is quaternion self-dual if its entries satisfy $h_{\ell k} = h_{k\ell}^+$, in particular, the diagonal elements h_{kk} are real. We assume that the offdiagonal elements of H are given (in the quaternion notation) by

$$h_{\ell k} = h_{k\ell}^+ := N^{-1/2}(x_{\ell k} + \mathbf{i}y_{\ell k} + \mathbf{j}z_{\ell k} + \mathbf{k}u_{\ell k}), \quad 1 \leq k < \ell \leq N \tag{3.5}$$

where $x_{\ell k}$, $y_{\ell k}$, $z_{\ell k}$ and $u_{\ell k}$ are real i.i.d. random variables with law ν that has zero expectation and variance $\frac{1}{4}$. The diagonal entries are real,

$$h_{kk} = N^{-1/2} x_{kk}, \qquad 1 \leq k \leq N,$$

where x_{kk} has a law $\widetilde{\nu}$ with zero expectation and variance $\frac{1}{2}$. The spectrum of H is doubly degenerate and we will neglect this degeneracy, i.e., we consider only N real (typically distinct) eigenvalues, $x_1 < x_2 < \ldots < x_N$.

The Gaussian ensembles (GOE, GUE and GSE) are special Wigner ensembles with ν and $\widetilde{\nu}$ being Gaussian distribution. These ensembles are invariant under their corresponding symmetry group, i.e., the distribution remains unchanged under the conjugation $H \to UHU^*$. Here U is an arbitrary orthogonal matrix in case of GOE, it is a unitary matrix for GUE and it is a unitary matrix over the quaternions in case of GSE. In the latter case, if one uses the $(2N) \times (2N)$ complex matrix representation, then the symmetry group is $\mathrm{Sp}(N) = \mathrm{Sp}(N, \mathbb{C}) \cap SU(2N)$.

With the given normalization, the eigenvalues are supported asymptotically in $[-2, 2]$ in all three cases, Moreover their empirical density converges weakly to the Wigner semicircle law in probability [43], i.e., for any $J \in C_0(\mathbb{R})$ and for any $\varepsilon > 0$, we have

$$\lim_{N\to\infty} \mathbb{P}\left\{ \Big| \frac{1}{N} \sum_{j=1}^{N} J(x_j) - \int J(x) \varrho_{sc}(x) \mathrm{d}x \Big| \geq \varepsilon \right\} = 0, \tag{3.6}$$

where

$$\varrho_{sc}(x) := \frac{1}{2\pi} \sqrt{(4 - x^2)_+}\,. \tag{3.7}$$

In particular, the typical spacing between neighboring eigenvalues is of order $1/N$ in the bulk of the spectrum.

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

$$\int_{\mathbb{R}} \exp\left[\delta_0 x^2\right] \mathrm{d}\nu(x) < \infty, \qquad \int_{\mathbb{R}} \exp\left[\delta_0 x^2\right] \mathrm{d}\widetilde{\nu}(x) < \infty. \tag{3.8}$$

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

$$\int e^{\delta_0 |x|^\gamma} \mathrm{d}\nu(x) < \infty, \qquad \int e^{\delta_0 |x|^\gamma} \mathrm{d}\widetilde{\nu}(x) < \infty. \tag{3.9}$$

For some statements we will need to assume that the measures $\nu, \widetilde{\nu}$ satisfy the logarithmic Sobolev inequality, i.e., for any density $h \geq 0$ with $\int h \mathrm{d}\nu = 1$ it holds that

$$\int h \log h \, \mathrm{d}\nu \leq C \int |\nabla \sqrt{h}|^2 \mathrm{d}\nu \tag{3.10}$$

and a similar bound holds for $\widetilde{\nu}$. We remark that (3.10) implies (3.8), see, e.g. [28].

3.2 Sample Covariance Matrix

The *real sample covariance* matrix ensemble consists of symmetric $N \times N$ matrices of the form $H = A^* A$. Here A is an $M \times N$ real matrix with $d = N/M$ fixed and we assume that $0 < d < 1$. The elements of A are given by

$$A_{\ell k} = M^{-1/2} x_{\ell k}, \;\; 1 \leq \ell \leq M, \;\; 1 \leq k \leq N, \tag{3.11}$$

where $x_{\ell k}$ are real i.i.d random variables with the distribution ν that is symmetric and has variance 1. In the case of *complex sample covariance* ensemble we assume that

$$A_{\ell k} = M^{-1/2} \left(x_{\ell k} + i y_{\ell k} \right), \;\; 1 \leq \ell \leq M, \;\; 1 \leq k \leq N, \tag{3.12}$$

where $x_{\ell k}$ and $y_{\ell k}$ are symmetric, real i.i.d. random variables with distribution ν that has variance $\frac{1}{2}$. We will assume that ν has Gaussian (3.8) or sometimes only subexponential (3.9) decay. The spectrum of H asymptotically lies in the interval $[\lambda_-, \lambda_+]$, where

$$\lambda_\pm \equiv \left(1 \pm d^{1/2} \right)^2. \tag{3.13}$$

Moreover, analogously to (3.6), the empirical density of eigenvalues converges weakly in probability to the Marchenko-Pastur law

$$\rho_W(x) = \frac{1}{2\pi d} \sqrt{\frac{\big[(\lambda_+ - x)(x - \lambda_-) \big]_+}{x^2}} \, . \tag{3.14}$$

Most of the analysis will be done for the singular values of A that are denoted by $\mathbf{x} = (x_1, \ldots, x_N)$. They are supported asymptotically in $[\sqrt{\lambda_-}, \sqrt{\lambda_+}]$ and therefore the typical spacing between neighboring singular values is of order $1/N$. Their empirical density converges to

$$\widetilde{\varrho}_W(x) := 2x \varrho_W(x^2) = \frac{1}{\pi d} \sqrt{\frac{\big[(\lambda_+ - x^2)(x^2 - \lambda_-) \big]_+}{x^2}} \, . \tag{3.15}$$

We remark that the assumption that ν is symmetric is used only at one technical step, namely when we refer to the large deviation result for the extreme eigenvalues of the sample covariance matrices in [22] (see Lemma 9.2 below). The similar result for Wigner matrices has been proven without the symmetry condition, see Theorem 1.4 in [42].

3.3 Main Theorems

With the remarks at the beginning of Section 3, Theorem 2.1 applies directly to prove universality for Wigner and sample covariance ensembles with a small Gaussian component; we will not state these theorems separately. To remove the small time restriction from Theorem 2.1, we will apply the reverse heat flow argument. This will give our main result:

Theorem 3.1 *Consider an $N \times N$ symmetric, hermitian or quaternion self-dual Wigner matrix H, or an $N \times N$ real or complex sample covariance matrix A^*A. Assume that the single site entries of H or A are i.i.d. with probability distribution $\nu(\mathrm{d}x) = u_0(x)\mathrm{d}x$ and with the standard normalization specified in Sections 3.1 and 3.2. We assume that ν satisfies the logarithmic Sobolev inequality* (3.10) *and in case of the sample covariance matrix we also assume that ν is symmetric. The same conditions are assumed for the distribution $\widetilde{\nu}$ of the diagonal elements in case of the Wigner matrix. Let $f_0 = f_{0,N}$ denote the joint density function of the eigenvalues and let $p_{0,N}^{(k)}$ be the k-point correlation function of f_0. Let ϱ denote the corresponding density of states, i.e., ϱ is given by the Wigner semicircle law* (3.7) *or the Marchenko-Pastur law* (3.14)*, respectively. Let $E \in \mathbb{R}$, $b > 0$ such that $\min\{\varrho(x) \,:\, x \in [E-b, E+b]\} > 0$. If for any $k \geq 1$ there is a constants M_k such that the density function u_0 satisfies*

$$\sum_{j=0}^{M_k} |\partial_x^j \log u_0(x)| \leq C_k(1+|x|)^{C_k} \tag{3.16}$$

for some constants $C_k < \infty$, then for any compactly supported continuous test function $O : \mathbb{R}^k \to \mathbb{R}$ we have

$$\begin{aligned} \lim_{N\to\infty} \int_{E-b}^{E+b} \mathrm{d}E' \int_{\mathbb{R}^k} & \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_k \, O(\alpha_1, \ldots, \alpha_k) \\ & \times \frac{1}{\varrho(E)^k} \Big(p_{0,N}^{(k)} - p_{\mu,N}^{(k)} \Big) \Big(E' + \frac{\alpha_1}{N\varrho(E)}, \ldots, E' + \frac{\alpha_k}{N\varrho(E)} \Big) = 0. \end{aligned} \tag{3.17}$$

Here μ denotes the probability measure of the eigenvalues of the appropriate Gaussian ensemble, i.e. GUE, GOE, GSE for the case of hermitian, symmetric, and, respectively, quaternion self-dual Wigner matrices; and the ensembles of real or complex sample covariance matrices with Gaussian entries (Wishart ensemble) in case of the covariance matrices A^*A. These measures are given in (1.6), with $\beta = 1, 2, 4$, for Wigner matrices, and, expressed in terms of singular values, in (5.6), with $\beta = 1, 2$, for sample covariance matrices.

Remark 1.: In the case of symmetric and hermitian Wigner matrices, the condition (3.16) can be removed by applying the Four-moment theorem of Tao and Vu (Theorem 15 of [39]) as in the proof of Corollary 2.4 of [20]. Similar remark applies to the sample covariance ensembles and to the quaternion self-dual Wigner ensemble provided the corresponding Four-moment theorem is established.

We also remark that a manuscript by Ben-Arous and Péché [4] with a similar statement is in preparation for complex sample covariance matrices that holds for a fixed E', i.e., without averaging over the energy parameter in (3.17).

Remark 2.: After the first version of this manuscript was posted on the arxiv, the question that whether the four moment theorem for sample covariance matrices holds was settled in [41]. In particular, [41] gives an alternative proof of the universality of local statistics for the complex sample covariance ensemble when combined with the result of [3]. For the real sample covariance ensemble the universality was established for distributions whose first four moments match the standard Gaussian variable. An important common ingredient to both our approach and that of [41] is the local Marchenko-Pastur law, established in Proposition 8.1; a slightly different version suitable for the application to prove the four moment theorem is proved in [41].

The four moment theorem in [41] compares the distributions of individual eigenvalues for two different ensembles. For our application to the correlation functions and gap distributions, an alternative approach is to use the recent Green function comparison theorem [21]. This will also remove the smoothness and logarithmic Sobolev inequality restrictions in Theorem 3.1.

We now state our result concerning the eigenvalue gap distribution both for Wigner and sample covariance ensembles. For any $s > 0$ and E with $\rho(E) > 0$ we define the density of eigenvalue pairs with distance less than $s/N\varrho(E)$ in the vicinity of E by

$$\Lambda(E;s) = \frac{1}{2N\ell_N\varrho(E)}\#\Big\{1\le j\le N-1\,:\,x_{j+1}-x_j\le\frac{s}{N\varrho(E)},\;|x_j-E|\le\ell_N\Big\},\tag{3.18}$$

where $\ell_N = N^{-\delta}$ for some $0<\delta\ll 1$.

Theorem 3.2 *Consider an $N\times N$ Wigner or sample covariance matrix as in Theorem 3.1 such that the probability measure $\mathrm{d}\nu = u_0\mathrm{d}x$ of the matrix elements satisfies the logarithmic Sobolev inequality* (3.10) *and, additionally, ν is symmetric in the sample covariance matrix case. Suppose that the initial density u_0 satisfies*

$$\sum_{j=0}^{M}|\partial_x^j\log u_0(x)|\le C(1+|x|)^C\tag{3.19}$$

with some sufficiently large constants C, M that depend on the ε in Assumption III. Then for any E with $\rho(E)>0$ and for any continuous, compactly supported test function $O:\mathbb{R}\to\mathbb{R}$ we have

$$\lim_{N\to\infty}\int_{\mathbb{R}}\mathrm{d}s\,O(s)[\mathbb{E}\,\Lambda(E;s)-\mathbb{E}^{\mu}\,\Lambda(E;s)]=0,\tag{3.20}$$

where μ is the probability measure of the eigenvalues of the appropriate Gaussian ensemble, as in Theorem 3.1.

Theorem 3.2 shows that, in particular, the probability to find no eigenvalue in the interval $[E, E+\alpha/(\varrho(E)N)]$ is asymptotically the same as in the corresponding classical Gaussian ensemble. Theorems 3.1 and 3.2 will follow from Theorem 2.1 and the reverse heat flow argument that we present in Section 6. We remark that the additional condition on the symmetry of ν in the case of sample covariance matrices stems from using a result from [22] on the lowest eigenvalue of these matrices, see Lemma 9.2.

Theorem 3.2 can be proven directly from Theorem 4.1 since the test functions of the form

$$\frac{1}{N}\sum_{i\in J} G(N(x_i - x_{i+1}))$$

determine the distribution of the random variable $\Lambda(E; s)$ uniquely. Here we take J to be the set

$$J := \big\{ i \, : \, \gamma_i \in [E - \ell_N, E + \ell_N] \big\},$$

where γ_i was defined in (2.12). Notice that δ in the definition of ℓ_N has to be small enough so that the edge term near the boundary of the interval is negligible.

4 Local Relaxation Flow

Theorem 4.1 (Universality of Dyson Brownian Motion for Short Time) *Suppose that the Hamiltonian $\mathcal{H}$ given in* (2.6) *satisfies the convexity bound* (2.10) *with $\beta \geq 1$. Let f_t be the solution of the forward equation* (2.1) *with an initial density f_0. Fix a positive $\varepsilon > 0$, set $t_0 = N^{-2\varepsilon}$ and define*

$$Q := \sup_{t\geq t_0} \sum_j \int (x_j - \gamma_j)^2 f_t \mathrm{d}\mu. \tag{4.1}$$

Assume that at time t_0 we have $S_\mu(f_{t_0}) := \int f_{t_0} \log f_{t_0} \mathrm{d}\mu \leq CN^m$ with some fixed exponent m that may depend on ε. Fix $n \geq 1$ and an array of positive integers, $\mathbf{m} = (m_1, m_2, \ldots, m_n) \in \mathbb{N}_+^n$. Let $G : \mathbb{R}^n \to \mathbb{R}$ be a bounded smooth function with compact support and define

$$\mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) := G\Big(N(x_i - x_{i+m_1}), N(x_{i+m_1} - x_{i+m_2}), \ldots, N(x_{i+m_{n-1}} - x_{i+m_n})\Big). \tag{4.2}$$

Then for any sufficiently small $\varepsilon' > 0$, there exist constants C, c, depending only on ε' and G such that for any $J \subset \{1, 2, \ldots, N - m_n\}$ and for any $\tau \geq 3t_0 = 3N^{-2\varepsilon}$, we have

$$\Big| \int \frac{1}{N}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) f_\tau \mathrm{d}\mu - \int \frac{1}{N}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) \mathrm{d}\mu \Big| \leq CN^{\varepsilon'} \sqrt{|J| Q (\tau N)^{-1}} + Ce^{-cN^{\varepsilon'}}, \tag{4.3}$$

where $|J|$ is the number of the elements in J.

The proof of this theorem is similar but much simpler than that of Theorem 2.1 of [20]. The estimate (4.3) improves slightly over the similar estimate in [20] by a factor $|J|/N$ due to the improvement in (4.19). Theorem 2.1 will follow from the fact that in case $\tau \geq N^{-2\varepsilon+\delta}$, the assumption (2.13) guarantees that

$$N^{\varepsilon'}\sqrt{|J|Q(\tau N)^{-1}} \leq N^{\varepsilon'-\delta/2} = N^{-\delta/6} \to 0$$

with the choice $\varepsilon' = \delta/3$ and using $|J| \leq N$. More precise error bound will be obtained by relating b to $|J|$. Therefore the local statistics of observables involving eigenvalue differences coincide in the $N \to \infty$ limit. To complete the proof of Theorem 2.1, we will have to show that the convergence of the observables $\mathcal{G}_{i,\mathbf{m}}$ is sufficient to identify the correlation functions of the x_i's in the sense prescribed in Theorem 2.1. The details will be given in Section 7.

Proof of Theorem 4.1. Without loss of generality we can assume in the sequel that $f_0 \in L^\infty(\mathrm{d}\mu)$. To see this, note that any $f_0 \in L^1(\mathrm{d}\mu)$ can be approximated by a sequence of bounded functions $f_0^{(k)}$ in L^1-norm with arbitrary precision and the dynamics is a contraction in L^1 (see Appendix A), thus f_τ and $f_\tau^{(k)}$ are arbitrarily close in L^1. Since G is bounded on the left hand side of (4.3), this is sufficient to pass to the limit $k \to \infty$.

Every constant in this proof depends on ε' and G, and we will not follow the precise dependence. We can assume that $\varepsilon' < \varepsilon$. Given $\tau > 0$, we define

$$R := \tau^{1/2} N^{-\varepsilon'/2}. \tag{4.4}$$

Notice that the choice of R depending on τ which is the main reason that τ appears in the denominator on the right hand side of (4.3). We now introduce the *pseudo equilibrium measure*, $\omega_N = \omega = \psi\mu_N$, defined by

$$\psi = \frac{Z}{\widetilde{Z}} \exp\big(-NW\big), \quad W(\mathbf{x}) = \sum_{j=1}^{N} W_j(x_j), \qquad W_j(x) = \frac{1}{2R^2}(x_j - \gamma_j)^2,$$

where $\widetilde{Z}$ is chosen such that ω is a probability measure, in particular $\omega = e^{-N\widetilde{\mathcal{H}}}/\widetilde{Z}$ with

$$\widetilde{\mathcal{H}} = \mathcal{H} + W. \tag{4.5}$$

Similarly to (2.9), one can check that

$$|\log \widetilde{Z}| \leq CN^m \tag{4.6}$$

with some exponent m.

Note that the additional term W_j confines the j-th point x_j near its classical location. We will prove that the probability w.r.t. the equilibrium measure μ_N of the event that x_j near its classical location is very close to 1. Thus there is little difference between the two measures ω_N and μ_N and in fact, we will prove that their local statistics are identical in the limit $N \to \infty$. The main advantage of the pseudo equilibrium measure comes from the fact that it has a faster decay to global equilibrium as shown in Theorem 4.2.

The local relaxation flow is defined to be the reversible dynamics w.r.t. ω. The dynamics is described by the generator $\tilde{L}$ defined by

$$\int f \tilde{L} g \mathrm{d}\omega = -\frac{1}{2N} \sum_j \int \partial_j f \partial_j g \mathrm{d}\omega. \tag{4.7}$$

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

$$L = \widetilde{L} + \sum_j b_j \partial_j, \quad b_j = W_j'(x_j) = \frac{x_j - \gamma_j}{R^2}. \tag{4.8}$$

Since the additional potential W_j is uniformly convex with

$$\inf_j \inf_{x \in \mathbb{R}} W_j''(x) \geq R^{-2}, \tag{4.9}$$

by (2.10) and $\beta \geq 1$ we have

$$\Big\langle \mathbf{v}, \nabla^2 \widetilde{\mathcal{H}}(\mathbf{x}) \mathbf{v} \Big\rangle \geq \frac{1}{R^2} \|\mathbf{v}\|^2 + \frac{1}{N} \sum_{i<j} \frac{(v_i - v_j)^2}{(x_i - x_j)^2}, \qquad \mathbf{v} \in \mathbb{R}^N. \tag{4.10}$$

Here we have used $U'' \geq 0$ in the last estimate. If this assumption is replaced by

$$U'' \geq -M \tag{4.11}$$

for some constant M independent of N, then there will be an extra term $-M\|\mathbf{v}\|^2$ in (4.10). Assuming $\tau \leq N^{\varepsilon'}$, we have $R \leq N^{-\varepsilon'/2}$, then this extra term can be controlled by the R^{-2} term and the same proof will go through. Since for the applications in this paper, the condition $U'' \geq 0$ is satisfied, we will not use this remark here.

The R^{-2} in the first term comes from the additional convexity of the local interaction and it enhances the "local Dirichlet form dissipation". In particular we have the uniform lower bound

$$\nabla^2 \widetilde{\mathcal{H}} = \mathrm{Hess}(-\log \omega) \geq R^{-2}. \tag{4.12}$$

This guarantees that the relaxation time to equilibrium ω for the $\widetilde{L}$ dynamics is bounded above by CR^2. We recall the definition of the relative entropy of f with respect to any probability measure $\mathrm{d}\lambda$

$$S_\lambda(f) = \int f \log f \mathrm{d}\lambda, \qquad S_\lambda(f|\psi) = \int f \log(f/\psi) \mathrm{d}\lambda.$$

The first ingredient to prove Theorem 4.1 is the analysis of the local relaxation flow which satisfies the logarithmic Sobolev inequality and the following dissipation estimate. Its proof follows the standard argument in [2] (used in this context in Section 5.1 of [16]). In Appendix B we will explain how to extend this argument onto the subdomain Σ_N. Here we only remark that the key inputs are the convexity bounds (4.10, 4.12) on the Hessian of $\widetilde{\mathcal{H}}$ (4.10).

Theorem 4.2 *Suppose* (4.10) *holds. Consider the forward equation*

$$\partial_t q_t = \widetilde{L} q_t, \qquad t \geq 0, \tag{4.13}$$

with an initial condition q_0 and with the reversible measure ω. Assume that $q_0 \in L^\infty(\mathrm{d}\omega)$. Then we have the following estimates

$$\partial_t D_\omega(\sqrt{q_t}) \leq -\frac{1}{R^2} D_\omega(\sqrt{q_t}) - \frac{1}{2N^2} \int \sum_{i,j=1}^{N} \frac{(\partial_i \sqrt{q_t} - \partial_j \sqrt{q_t})^2}{(x_i - x_j)^2} \mathrm{d}\omega, \tag{4.14}$$

$$\frac{1}{2N^2} \int_0^\infty \mathrm{d}s \int \sum_{i,j=1}^{N} \frac{(\partial_i \sqrt{q_s} - \partial_j \sqrt{q_s})^2}{(x_i - x_j)^2} \mathrm{d}\omega \leq D_\omega(\sqrt{q_0}) \tag{4.15}$$

and the logarithmic Sobolev inequality

$$S_\omega(q) \leq C R^2 D_\omega(\sqrt{q}) \tag{4.16}$$

with a universal constant C. Thus the time to equilibrium is of order R^2:

$$S_\omega(q_t) \leq e^{-Ct/R^2} S_\omega(q_0). \tag{4.17}$$

□

The estimate (4.15) on the second term in (4.10) plays a key role in the next theorem.

Theorem 4.3 *Suppose that Assumption I holds and we have a density $q \in L^\infty$, $\int q \mathrm{d}\omega = 1$. Recall that $\tau = R^2 N^{\varepsilon'}$. Fix $n \geq 1$, $\mathbf{m} \in \mathcal{N}_+^n$, let $G : \mathbb{R}^n \to \mathbb{R}$ be a bounded smooth function with compact support and recall the definition of $\mathcal{G}_{i,\mathbf{m}}$ from* (4.2)*. Then for any $J \subset \{1, 2, \ldots, N-n\}$ we have*

$$\Big| \int \frac{1}{N} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) \mathrm{d}\omega - \int \frac{1}{N} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) q \mathrm{d}\omega \Big| \leq C \Big(\frac{|J| D_\omega(\sqrt{q}) \tau}{N^2} \Big)^{1/2} + C e^{-cN^{\varepsilon'}} \sqrt{S_\omega(q)}. \tag{4.18}$$

Proof. For simplicity, we will consider the case when $\mathbf{m} = (1, 2, \ldots n)$, the general case easily follows by appropriately redefining the function G. Let q_t satisfy

$$\partial_t q_t = \widetilde{L} q_t, \qquad t \geq 0,$$

with an initial condition q. Thanks to the exponential decay of the entropy on time scale $\tau \gg R^2$, see (4.17), and the entropy bound on the initial state q, the difference between the local statistics w.r.t. $q_\tau \omega$ and $q_\infty \omega = \omega$ is subexponentially small in N,

$$\Big| \int \frac{1}{N} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) q_\tau \mathrm{d}\omega - \int \frac{1}{N} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) q_\infty \mathrm{d}\omega \Big| \le \|G\|_\infty \int |q_\tau - 1| \mathrm{d}\omega$$
$$\le C \sqrt{S_\omega(q_\tau)} \le C e^{-cN^{\varepsilon'}} \sqrt{S_\omega(q)},$$

giving the second term on the r.h.s. of (4.18). To compare q with q_τ, by differentiation, we have

$$\int \frac{1}{N} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) q_\tau \mathrm{d}\omega - \int \frac{1}{N} \sum_{i \in J} \mathcal{G}_{i,\mathbf{m}}(\mathbf{x}) q \mathrm{d}\omega$$
$$= \int_0^\tau \mathrm{d}s \int \frac{1}{N} \sum_{i \in J} \sum_{k=1}^n \partial_k G\Big(N(x_i - x_{i+1}), \ldots, N(x_{i+n-1} - x_{i+n}) \Big) [\partial_{i+k-1} q_s - \partial_{i+k} q_s] \mathrm{d}\omega.$$

Here we used the definition of $\widetilde{L}$ from (4.7) and note that the $1/N$ factor present in (4.7) cancels the factor N from the argument of G (4.2). From the Schwarz inequality and $\partial q = 2\sqrt{q}\partial\sqrt{q}$, the last term is bounded by

$$2 \sum_{k=1}^n \left[\int_0^\tau \mathrm{d}s \int \sum_{i \in J} \Big[\partial_k G\big(N(x_i - x_{i+1}), \ldots, N(x_{i+n-1} - x_{i+n}) \big) \Big]^2 (x_{i+k-1} - x_{i+k})^2 \, q_s \mathrm{d}\omega \right]^{1/2}$$
$$\times \left[\int_0^\tau \mathrm{d}s \int \frac{1}{N^2} \sum_{i \in J} \frac{1}{(x_{i+k-1} - x_{i+k})^2} [\partial_{i+k-1} \sqrt{q_s} - \partial_{i+k} \sqrt{q_s}]^2 \mathrm{d}\omega \right]^{1/2}$$
$$\le C \sqrt{\frac{|J| D_\omega(\sqrt{q}) \tau}{N^2}}, \tag{4.19}$$

where we have used (4.15) and that

$$\Big[\partial_k G\Big(N(x_i - x_{i+1}), \ldots, N(x_{i+k-1} - x_{i+k}), \ldots N(x_{i+n-1} - x_{i+n}) \Big) \Big]^2 (x_{i+k-1} - x_{i+k})^2 \le C N^{-2},$$

since G is smooth and compactly supported. This proves Theorem 4.3. □

As a comparison to Theorem 4.3, we state the following result which can be proved in a similar way.

Lemma 4.4 *Let $G : \mathbb{R} \to \mathbb{R}$ be a bounded smooth function with compact support and let a sequence E_i be fixed. Then we have*

$$\Big| \frac{1}{N} \sum_i \int G\big(N(x_i - E_i) \big) \mathrm{d}\omega - \frac{1}{N} \sum_i \int G\big(N(x_i - E_i) \big) q \mathrm{d}\omega \Big| \le C \Big(S_\omega(q) \tau \Big)^{1/2} + C e^{-cN^{\varepsilon'}} \sqrt{S_\omega(q)}. \tag{4.20}$$

Notice that by exploiting the local Dirichlet form dissipation coming from the second term on the r.h.s. of (4.14), we have gained the crucial factor $N^{-1/2}$ in the estimate (4.18) compared with (4.20).

The final ingredient to prove Theorem 4.1 is the following entropy and Dirichlet form estimates.

Theorem 4.5 *Suppose that* (2.10) *holds and recall* $\tau = R^2 N^{\varepsilon'} \geq 3t_0$ *with* $t_0 = N^{-2\varepsilon}$. *Let* $g_t = f_t/\psi$ *so that* $S_\mu(f_t|\psi) = S_\omega(g_t)$. *Assume that* $S_\mu(f_{t_0}) \leq CN^m$ *with some fixed* m. *Then the entropy and the Dirichlet form satisfy the estimates:*

$$S_\omega(g_{\tau/2}) \leq CNR^{-2}Q, \qquad D_\omega(\sqrt{g_\tau}) \leq CNR^{-4}Q. \tag{4.21}$$

Proof. Recall that $\partial_t f_t = Lf_t$. The standard estimate on the entropy of f_t with respect to the invariant measure is obtained by differentiating the entropy twice and using the logarithmic Sobolev inequality. The entropy and the Dirichlet form in (4.21) are, however, computed with respect to the measure ω. This yields the additional second term in the following identity [44] that holds for any probability density ψ_t:

$$\partial_t S_\mu(f_t|\psi_t) = -\frac{2}{N}\sum_j \int (\partial_j \sqrt{g_t})^2\, \psi_t \,\mathrm{d}\mu + \int g_t (L - \partial_t)\psi_t \,\mathrm{d}\mu\ ,$$

where $g_t = f_t/\psi_t$. In our application we set ψ_t to be time independent, $\psi_t = \psi = \omega/\mu$, hence we have

$$\partial_t S_\omega(g_t) = -\frac{2}{N}\sum_j \int (\partial_j \sqrt{g_t})^2 \,\mathrm{d}\omega + \int \widetilde{L} g_t \,\mathrm{d}\omega + \sum_j \int b_j \partial_j g_t \,\mathrm{d}\omega.$$

Since ω is invariant, the middle term on the right hand side vanishes, and from the Schwarz inequality

$$\partial_t S_\omega(g_t) \leq -D_\omega(\sqrt{g_t}) + CN\sum_j \int b_j^2 g_t \,\mathrm{d}\omega \leq -D_\omega(\sqrt{g_t}) + CN\Lambda, \qquad t \geq N^{-2\varepsilon}, \tag{4.22}$$

where we defined

$$\Lambda := QR^{-4} = \sup_{t\geq N^{-2\varepsilon}} R^{-4} \sum_j \int (x_j - \gamma_j)^2 f_t \mathrm{d}\mu. \tag{4.23}$$

Together with (4.16), we have

$$\partial_t S_\omega(g_t) \leq -CR^{-2} S_\omega(g_t) + CN\Lambda, \qquad t \geq N^{-2\varepsilon}. \tag{4.24}$$

To obtain the first inequality in (4.21), we integrate (4.24) from $t_0 = N^{-2\varepsilon}$ to $\tau/2$, using that $\tau = R^2 N^{\varepsilon'}$ and $S_\omega(g_{t_0}) \leq CN^m + N^2 Q$ with some finite m, depending on ε. This apriori bound follows from

$$S_\omega(g_{t_0}) = S_\mu(f_{t_0}|\psi) = S_\mu(f_{t_0}) - \log Z + \log \widetilde{Z} + N\int f_{t_0} W \mathrm{d}\mu \leq CN^m + N^2 Q, \tag{4.25}$$

where we used (2.9) and (4.6). The second inequality in (4.21) can be obtained from the first one by integrating (4.22) from $t = \tau/2$ to $t = \tau$ and using the monotonicity of the Dirichlet form in time. □

Finally, we complete the proof of Theorem 4.1. Recall that $\tau = R^2 N^{\varepsilon'}$ and $t_0 = N^{-2\varepsilon}$. Choose $q_\tau := g_\tau = f_\tau/\psi$ as density q in Theorem 4.3. The condition $q_\tau \in L^\infty$ can be guaranteed by the approximation argument from the beginning of the proof of Theorem 4.1. Then Theorem 4.5, Theorem 4.3 together with (4.25) and the fact that $\Lambda\tau = Q\tau^{-1}N^{2\varepsilon'}$ directly imply that

$$\Big| \int \frac{1}{N}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}} f_\tau \mathrm{d}\mu - \int \frac{1}{N}\sum_{i\in J} \mathcal{G}_{i,\mathbf{m}} \mathrm{d}\omega \Big| \le C N^{\varepsilon'} \sqrt{|J| Q (\tau N)^{-1}} + C e^{-cN^{\varepsilon'}}, \tag{4.26}$$

i.e., the local statistics of $f_\tau \mu$ and ω can be compared. Clearly, equation (4.26) also holds for the special choice $f_0 = 1$ (for which $f_\tau = 1$), i.e., local statistics of μ and ω can also be compared. This completes the proof of Theorem 4.1. □

5 Equilibrium measure and Dyson Brownian motion

We will treat the Wigner and sample covariance ensembles in parallel. Suppose $(x_1, x_2, \dots, x_N)$ denote the eigenvalues of the Gaussian Wigner ensembles. The joint distribution of $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ of the Gaussian Wigner ensembles is given by the following measure on $\mathbb{R}^N$

$$\mu = \mu_{\beta,N}(\mathrm{d}\mathbf{x}) = \frac{e^{-N\mathcal{H}_\beta(\mathbf{x})}}{Z_\beta}\mathrm{d}\mathbf{x}, \qquad \mathcal{H}_\beta(\mathbf{x}) = \beta\left[\sum_{i=1}^N \frac{1}{4}x_i^2 - \frac{1}{N}\sum_{i<j}\log|x_j - x_i|\right] \tag{5.1}$$

where $\beta \ge 1$ is an arbitrary parameter, i.e., this corresponds to choosing $U(x) = x^2/4$ in (2.6). With a slight abuse of notations we will use μ for both the measure $\mathrm{d}\mu$ and its density $e^{-N\mathcal{H}_\beta}/Z_\beta$ with respect to the Lebesgue measure. The specific value $\beta = 1, 2, 4$ correspond to the GUE, GOE and GSE ensembles, respectively.

We define the following generator

$$L = L_{\beta,N} = \sum_{i=1}^N \frac{1}{2N}\partial_i^2 + \beta\sum_{i=1}^N\Bigg(-\frac{1}{4}x_i + \frac{1}{2N}\sum_{j\ne i}\frac{1}{x_i - x_j}\Bigg)\partial_i \tag{5.2}$$

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

$$D(f) = D_\mu(f) = -\int fLf\mathrm{d}\mu = \sum_{j=1}^N \frac{1}{2N}\int(\partial_j f)^2 \mathrm{d}\mu, \quad \text{and} \quad S(f) = S_\mu(f) := \int f\log f\mathrm{d}\mu \tag{5.3}$$

Let $f_t \mathrm{d}\mu$ denote the probability measure on the set Σ_N at the time t with the given generator L. Then f_t satisfies the forward equation

$$\partial_t f_t = L f_t \tag{5.4}$$

with initial condition f_0. This dynamics is the Dyson Brownian motion.

The Dyson Brownian motion is the corresponding system of stochastic differential equations for the vector $\mathbf{x}(t)$ that is given by

$$\mathrm{d}x_i = \frac{\mathrm{d}B_i}{\sqrt{N}} + \beta \left[-\frac{1}{4} x_i + \frac{1}{2N} \sum_{j \neq i} \frac{1}{x_i - x_j} \right] \mathrm{d}t, \qquad 1 \leq i \leq N, \tag{5.5}$$

where $\{B_i \ : \ 1 \leq i \leq N\}$ is a collection of independent standard Brownian motions on $\mathbb{R}$. This SDE is well posed for $\beta \geq 1$, and in particular the points do not cross each other with probability one, i.e., the process is well defined on Σ_N (see, e.g. Section 12.1 of [25])

The treatment of the sample covariance ensembles is fully analogous, but the formulas change slightly. We use the convention in the sample covariance case that x_i denotes the singular values of A and $\lambda_i = x_i^2$ are the eigenvalues of A^*A. Most of the formulas will be in terms of x_i's; in particular we consider the joint distribution function $f_0(\mathbf{x})$ of the singular values. The invariant measure for the singular values is given by (c.f. (5.1)):

$$\mu^W = \mu^W_{\beta,N}(\mathrm{d}\mathbf{x}) = \frac{e^{-N\mathcal{H}^W_\beta(\mathbf{x})}}{Z_\beta} \mathrm{d}\mathbf{x}. \tag{5.6}$$

where

$$\mathcal{H}^W_\beta(\mathbf{x}) = \beta \left[\sum_{i=1}^N \frac{x_i^2}{2d} - \frac{1}{N} \sum_{i<j} \log |x_j^2 - x_i^2| - \Big(\frac{1}{d} - 1 + \frac{1-\beta^{-1}}{N} \Big) \sum_{i=1}^N \log |x_i| \right],$$

where $d = M/N$ and $\beta = 1$ when X is a real matrix, $\beta = 2$ when X is a complex matrix. This formula can be obtained by direct calculation (see also Proposition 2.16 of [23] or Fig. 1 of [11] after appropriate rescaling). Define the generator (c.f. (5.2))

$$L^W = L^W_{\beta,N} = \sum_{i=1}^N \frac{1}{2N} \partial_i^2 + \sum_{i=1}^N \left(-\frac{\beta x_i}{2d} + \frac{\beta}{N} \sum_{j \neq i} \frac{x_i}{x_i^2 - x_j^2} + \frac{1}{2} \Big(\beta \big(\frac{1}{d} - 1 \big) + \frac{\beta - 1}{N} \Big) \frac{1}{x_j} \right) \partial_i. \tag{5.7}$$

Finally, the stochastic differential equation is given by (c.f. (5.5))

$$\mathrm{d}x_i = \frac{\mathrm{d}B_i}{\sqrt{N}} + \left[-\frac{\beta x_i}{2d} + \frac{\beta}{2N} \sum_{j \neq i} \left[\frac{1}{x_i - x_j} + \frac{1}{x_i + x_j} \right] + \frac{1}{2} \Big(\beta \big(\frac{1}{d} - 1 \big) + \frac{\beta - 1}{N} \Big) \frac{1}{x_j} \right] \mathrm{d}t, \qquad 1 \leq i \leq N. \tag{5.8}$$

In applications to Wigner matrices ($\beta = 1, 2, 4$), $f_0 \mathrm{d}\mu$ will be the joint probability density of the eigenvalues of the initial hermitian, symmetric or quaternion self-dual Wigner matrix $\widehat{H}$. The limiting density is the Wigner semicircle law given in (3.7). The Dyson Brownian motion describes the eigenvalues of the matrix valued process

$$\mathrm{d}H_t = \frac{\mathrm{d}\mathbf{B}_t}{\sqrt{N}} - \frac{1}{2} H_t \mathrm{d}t \tag{5.9}$$

with $H_0 = \widehat{H}$. Here $\mathbf{B}_t$ is a symmetric, hermitian or quaternion self-dual matrix-valued process whose offdiagonal elements are standard real, complex or quaternion Brownian motions with variance one and the diagonal elements of are real Brownian motions with variance $2, 1$ and $\frac{1}{2}$, in case $\beta = 1, 2, 4$, respectively. More precisely, let u_t denote the density function of the distribution of one real component of the (ij)-th entry of H_t, $i < j$ (there are two real components for the hermitian matrices and four for the quaternion matrices), then

$$\partial_t u_t = B u_t, \quad B = \frac{1}{2} \frac{\partial^2}{\partial x^2} - \frac{\beta x}{2} \frac{\partial}{\partial x}. \tag{5.10}$$

Let $\gamma(\mathrm{d}x) = \gamma(x)\mathrm{d}x := (\beta/2\pi)^{1/2} e^{-\beta x^2/2} \mathrm{d}x$ denote the reversible measure for this process. The diagonal elements evolve according to an OU process with twice variance. For any $t \geq 0$, the solution to (5.9), H_t, has the same distribution as

$$e^{-t/2} \widehat{H} + (1 - e^{-t})^{1/2} V, \tag{5.11}$$

where V is a GUE, GOE or GSE matrix.

The generator of the induced stochastic process on the eigenvalues is given by (5.2). The equilibrium measure μ is the GUE, GOE or GSE eigenvalue distribution. Theorem 2.1 thus says in this case that the local eigenvalue statistics of a Wigner random matrix with a small Gaussian component coincides with the local statistics of the corresponding Gaussian ensemble. The entropy condition on $S_\mu(f_{t_0})$ in Theorem 2.1 can be easily obtained by

$$S_\mu(f_{t_0}) \leq N^2 S_\gamma(u_{t_0}) \leq C N^m. \tag{5.12}$$

In the real or complex sample covariance case ($\beta = 1, 2$), the matrix elements of A evolve according to the OU process (5.10), i.e. A_t has the same distribution as

$$e^{-t/2} \widehat{A} + (1 - e^{-t})^{1/2} W, \tag{5.13}$$

where W is an $M \times N$ matrix whose elements are i.i.d real or complex Gaussian variables with mean 0 and variance $1/\beta$.

6 Reverse heat flow

To remove the short time restriction from Theorem 2.1 in case of Wigner and sample covariance ensembles and to prove Theorems 3.1 and 3.2, we apply the reverse heat flow argument, presented first in [14] and used also in Corollary 2.4 of [20].

For fixed $\beta = 1, 2$ or 4, recall the Ornstein-Uhlenbeck process from (5.10) with the reversible Gaussian measure $\gamma(\mathrm{d}x)$. Let u be a positive density with respect to γ, i.e., $\int u \mathrm{d}\gamma = 1$ and we write $u(x) = \exp(-V(x))$. Suppose that for any K fixed there are constants C_1, C_2 depending on K such that

$$\sum_{j=1}^{2K} |V^{(j)}(x)| \leq C_1(1 + x^2)^{C_2} \tag{6.1}$$

and the measure $\mathrm{d}\nu = u \mathrm{d}\gamma$ satisfies the subexponential decay condition. We will apply this for the initial distribution $\mathrm{d}\nu = u_0(x)\mathrm{d}x$, so u and u_0 differ by a Gaussian factor.

Proposition 6.1 *Suppose that $\nu = u\gamma$ satisfies the subexponential decay condition and* (6.1) *for some K. Then there is a small constant α_K depending on K such that for $t \leq \alpha_K$ there exists a probability density g_t with mean zero and variance $\frac{1}{2}$ such that*

$$\int \left| e^{tB} g_t - u \right| \mathrm{d}\gamma \leq C\, t^K \tag{6.2}$$

for some $C > 0$ depending on K. Furthermore, g_t can be chosen such that if the logarithmic Sobolev inequality (3.10) *holds for the measure $\nu = u\gamma$, then it holds for $g_t\gamma$ as well, with the logarithmic Sobolev constant changing by a factor of at most* 2.

Furthermore, let $\mathcal{B} = B^{\otimes n}$, $F = u^{\otimes n}$ with some $n \leq CN^2$. Denote by $G_t = g_t^{\otimes n}$. Then we also have

$$\int \left| e^{t\mathcal{B}} G_t - F \right| \mathrm{d}\gamma^{\otimes n} \leq C\, N^2 t^K \tag{6.3}$$

for some $C > 0$ depending on K.

We now explain how to prove Theorems 3.1 and 3.2 from Theorem 2.1 and Proposition 6.1. We choose n to be the number of independent OU processes needed to generate the flow of the matrix elements. By choosing K large enough, we can compare the two measures $e^{t\mathcal{B}} G_t$ and F in the total variational norm; for any observable $J : \mathbb{R}^n \to \mathbb{R}$ of the matrix elements, we have

$$\left| \int J (e^{t\mathcal{B}} G_t - F) \mathrm{d}\gamma^{\otimes n} \right| \leq \|J\|_\infty C\, N^2 t^K.$$

In order to prove Theorems 3.1 and 3.2, appropriate observables J need to be chosen that depend on the matrix elements via the eigenvalues to express the quantities in (3.17) and (3.20). It is easy to see that $\|J\|_\infty$ may grow at most polynomially in N. But we can always choose K large enough to compensate for it with the choice $t = N^{-2\varepsilon+\delta}$ allowed in Theorem 2.1. Here the verifications of the Assumptions I-IV of Theorem 2.1 were explained at the beginning of Section 3. This completes the proof of our main theorems. □

Proof of Proposition 6.1. Define $\theta(x) = \theta_0(t^\alpha x)$ with some small positive $\alpha > 0$ depending on K, where θ_0 is a smooth cutoff function satisfying $\theta_0(x) = 1$ for $|x| \le 1$ and $\theta_0(x) = 0$ for $|x| \ge 2$. Set

$$h_s = u + \theta\xi_s, \quad \text{with} \quad \xi_s := \left[-sB + \frac{1}{2}s^2B^2 + \ldots + (-1)^{K-1}\frac{s^{K-1}}{(K-1)!}B^{K-1}\right]u.$$

By assumption (6.1), h_s is positive and

$$\frac{2}{3}u \le h_s \le \frac{3}{2}u. \tag{6.4}$$

for any $s \le t$ if t is small enough. To see this, take, e.g., $K = 2$ and we have

$$|\theta(x)\xi_s(x)| \le Cs\theta_0(t^\alpha x)\Big[\,\big|V''(x)\big| + \big|xV'(x)\big|\,\Big]\,|u(x)| \le \frac{1}{2}|u(x)|,$$

where we have used $\alpha \ll 1$, $s \le t$ and the assumption (6.1).

Define $v_s = e^{sB}h_s$ and by definition, $v_0 = u$. Then

$$\partial_s v_s = (-1)^{K-1}\frac{s^{K-1}}{(K-1)!}e^{sB}B^Ku + e^{sB}B(\theta-1)\xi_s + e^{sB}(\theta-1)\partial_s\xi_s.$$

Since the Ornstein-Uhlenbeck is a contraction in $L^1(\mathrm{d}\gamma)$, together with (6.1), we have

$$\int |v_t - u|\mathrm{d}\gamma \le C_K\int_0^t\int\Big[t^{K-1}|B^Ku| + |B(\theta-1)\xi_s| + |(\theta-1)\partial_s\xi_s|\Big]\mathrm{d}\gamma\,\mathrm{d}s \le C_Kt^K \tag{6.5}$$

for sufficiently small t. To estimate the last two terms, we also used that on the support of $\theta - 1$ the measure $\mathrm{d}\gamma$ decays subexponentially in t.

Notice that h_t may not be normalized as a probability density w.r.t. γ but this can be easily adjusted. To compute this normalization, take for example, $K = 1$ and we have, by using $s \le t^\alpha$,

$$\Big|\int\theta(x)\xi_s(x)\mathrm{d}\gamma\Big| = \Big|s\int\theta_0(t^\alpha x)Bu(x)\mathrm{d}\gamma\Big| \le \Big|\int\theta_0'(t^\alpha x)u'(x)\mathrm{d}\gamma\Big| \le \int_{|x|\ge t^{-\alpha/2}}\Big|u'(x)\Big|\mathrm{d}\gamma.$$

The last term is bounded by $O(t^M)$ for any $M > 0$ due to that $u(x)\gamma$ has a subexponential decay and using the assumption (6.1) on V.

We have proved that there is a constant $c_t = 1 + O(t^M)$, for any $M > 0$ positive, such that c_th_t is a probability density. Clearly,

$$\alpha_t := \int xc_th_t\mathrm{d}\gamma = O(t^M), \qquad \sigma_t^2 := \int(x-\alpha_t)^2c_th_t\mathrm{d}\gamma = \beta^{-1} + O(t^M),$$

and the same formulas hold if h_t is replaced by v_t since the OU flow preserves expectation and variance. Let g_t be defined by

$$g_t(x)e^{-\beta x^2/2} = c_t\sigma_t^{-1}h_t((x+\alpha_t)\sigma^{-1})e^{-\beta(x+\alpha_t)^2/2\sigma^2}.$$

Then g_t is a probability density w.r.t. γ with zero mean and variance β^{-1}. It is easy to check that the total variation norm of $h_t - g_t$ is smaller than any power of t. Using again the contraction property of e^{tB} and (6.5), we get

$$\int |e^{tB}g_t - u|\mathrm{d}\gamma \le Ct^K \tag{6.6}$$

for sufficiently small t.

Now we check the LSI constant for g_t. Recall that g_t was obtained from h_t by translation and dilation. By definition of the LSI constant, the translation does not change it. The dilation changes the constant, but since our dilation constant is nearly one, the change of LSI constant is also nearly one. So we only have to compare the LSI constants between $\mathrm{d}\nu = u\mathrm{d}\gamma$ and $c_t h_t \mathrm{d}\gamma$. From (6.4) and that c_t is nearly one, the LSI constant changes by a factor less than 2. This proves the claim on the LSI constant.

Finally, the (6.3) directly follows from

$$\int \left|e^{t\mathcal{B}}G_t - F\right| \mathrm{d}\gamma^{\otimes n} \le n\int \left|e^{tB}g_t - u\right| \mathrm{d}\gamma$$

and this completes the proof of Proposition 6.1. $\square$

7 Proof of Theorem 2.1

We start with the identity

$$\begin{aligned}\int_{E-b}^{E+b}\mathrm{d}E'\int_{\mathbb{R}^n}\mathrm{d}\alpha_1\ldots\mathrm{d}\alpha_n\, O(\alpha_1,\ldots,\alpha_n)p^{(n)}_{\tau,N}\Big(E'+\frac{\alpha_1}{N\varrho(E)},\ldots,E'+\frac{\alpha_n}{N\varrho(E)}\Big) \\ = C_{N,n}\int_{E-b}^{E+b}\mathrm{d}E'\int\sum_{i_1\neq i_2\neq\ldots\neq i_n}\widetilde{O}\big(N(x_{i_1}-E'),N(x_{i_1}-x_{i_2}),\ldots N(x_{i_{n-1}}-x_{i_n})\big)f_\tau\mathrm{d}\mu,\end{aligned} \tag{7.1}$$

where $\widetilde{O}(u_1,u_2,\ldots u_n) := O\big(\varrho(E)u_1,\varrho(E)(u_2-u_1),\ldots\big)$ and $C_{N,n} = N^n(N-n)!/N! = 1+O_n(N^{-1})$. By permutational symmetry of $p^{(n)}_{\tau,N}$ we can assume that O is symmetric and we can restrict the last summation to $i_1 < i_2 < \ldots < i_n$ upon an overall factor $n!$. Let S_n denote the set of increasing positive integers, $\mathbf{m} = (m_2,m_3,\ldots,m_n) \in \mathbb{N}_+^{n-1}$, $m_2 < m_3 < \ldots < m_n$. For a given $\mathbf{m} \in S_n$, we change indices to $i = i_1$, $i_2 = i+m_2$, $i_3 = i+m_3,\ldots$, and rewrite the sum on the r.h.s. of (7.1) as

$$\sum_{\mathbf{m}\in S_n}\sum_{i=1}^N\widetilde{O}\big(N(x_i-E'),N(x_i-x_{i+m_2}),N(x_{i+m_2}-x_{i+m_3}),\ldots\big) = \sum_{\mathbf{m}\in S_n}\sum_{i=1}^N Y_{i,\mathbf{m}}(E',\mathbf{x}),$$

where we introduced

$$Y_{i,\mathbf{m}}(E',\mathbf{x}) = \widetilde{O}\big(N(x_i - E'), N(x_i - x_{i+m_2}), \dots, N(x_i - x_{i+m_n})\big).$$

We will set $Y_{i,\mathbf{m}} = 0$ if $i + m_n > N$. Our goal is to estimate the difference

$$\Theta := \left| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \sum_{\mathbf{m}\in S_n} \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x})(f_\tau - 1)\mathrm{d}\mu \right|. \tag{7.2}$$

Let M be an N-dependent parameter chosen at the end of the proof, in fact, M will be chosen as $M = N^c$ with some small positive exponent $c > 0$, depending on n. Let

$$S_n(M) := \{\mathbf{m} \in S_n\ ,\ m_n \le M\}, \quad S_n^c(M) := S_n \setminus S_n(M),$$

and note that $|S_n(M)| \le M^{n-1}$. We have the simple bound $\Theta \le \Theta_M^{(1)}(\tau) + \Theta_M^{(2)}(\tau) + \Theta_M^{(2)}(\infty)$ where

$$\Theta_M^{(1)}(\tau) := \left| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \sum_{\mathbf{m}\in S_n(M)} \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x})(f_\tau - 1)\mathrm{d}\mu \right| \tag{7.3}$$

and

$$\Theta_M^{(2)}(\tau) := \sum_{\mathbf{m}\in S_n^c(M)} \left| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x}) f_\tau \mathrm{d}\mu \right| \tag{7.4}$$

Note that $\Theta_M^{(2)}(\infty)$ is the same as $\Theta_M^{(2)}(\tau)$ but with f_τ replaced by the constant 1, i.e., $f_\infty \mathrm{d}\mu$ is the equilibrium.

Step 1: Small $\mathbf{m}$ *case*

After performing the $\mathrm{d}E'$ integration, we will eventually apply Theorem 4.1 to the function

$$G(u_1, u_2, \dots) := \int_{\mathbb{R}} \widetilde{O}(y, u_1, u_2, \dots,)\mathrm{d}y,$$

i.e., to the quantity

$$\int_{\mathbb{R}} \mathrm{d}E'\ Y_{i,\mathbf{m}}(E',\mathbf{x}) = \frac{1}{N} G\Big(N(x_i - x_{i+m_2}), \dots\Big) \tag{7.5}$$

for each fixed i and $\mathbf{m}$.

For any E and $0 < \xi < b$ define sets of integers $J = J_{E,b,\xi}$ and $J^\pm = J^\pm_{E,b,\xi}$ by

$$J := \big\{ i\ :\ \gamma_i \in [E - b, E + b] \big\}, \quad J^\pm := \big\{ i\ :\ \gamma_i \in [E - (b \pm \xi), E + b \pm \xi] \big\},$$

where γ_i was defined in (2.12). Clearly $J^- \subset J \subset J^+$. With these notations, we have

$$\int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x}) = \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \sum_{i\in J^+} Y_{i,\mathbf{m}}(E',\mathbf{x}) + \Omega^+_{J,\mathbf{m}}(\mathbf{x}). \tag{7.6}$$

The error term $\Omega^+_{J,\mathbf{m}}$, defined by (7.6) indirectly, comes from those $i \notin J^+$ indices, for which $x_i \in [E-b+O(N^{-1}), E+b+O(N^{-1})]$ since $Y_{i,\mathbf{m}}(E',\mathbf{x}) = 0$ unless $|x_i - E'| \le C/N$, the constant depending on the support of O. Thus

$$|\Omega^+_{J,\mathbf{m}}(\mathbf{x})| \le Cb^{-1}N^{-1}\#\{\, i \;:\; |x_i - \gamma_i| \ge \xi/2\} \tag{7.7}$$

for any sufficiently large N, assuming $\xi \gg 1/N$ and using that O is a bounded function. The additional N^{-1} factor comes from the $\mathrm{d}E'$ integration. Taking the expectation with respect to the measure $f_\tau \mathrm{d}\mu$, we get

$$\int |\Omega^+_{J,\mathbf{m}}(\mathbf{x})| f_\tau \mathrm{d}\mu \le Cb^{-1}\xi^{-2}N^{-1} \int \sum_i (x_i-\gamma_i)^2 f_\tau \mathrm{d}\mu = Cb^{-1}\xi^{-2}N^{-1-2\varepsilon} \tag{7.8}$$

using Assumption III (2.13). We can also estimate

$$\begin{aligned}
&\int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \sum_{i\in J^+} Y_{i,\mathbf{m}}(E',\mathbf{x}) \\
&\le \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \sum_{i\in J^-} Y_{i,\mathbf{m}}(E',\mathbf{x}) + Cb^{-1}N^{-1}|J^+ \setminus J^-| \\
&= \int_{\mathbb{R}} \frac{\mathrm{d}E'}{2b} \sum_{i\in J^-} Y_{i,\mathbf{m}}(E',\mathbf{x}) + Cb^{-1}N^{-1}|J^+ \setminus J^-| + \Xi^+_{J,\mathbf{m}}(\mathbf{x}) \\
&\le \int_{\mathbb{R}} \frac{\mathrm{d}E'}{2b} \sum_{i\in J} Y_{i,\mathbf{m}}(E',\mathbf{x}) + Cb^{-1}N^{-1}|J^+ \setminus J^-| + Cb^{-1}N^{-1}|J \setminus J^-| + \Xi^+_{J,\mathbf{m}}(\mathbf{x}),
\end{aligned} \tag{7.9}$$

where the error term $\Xi^+_{J,\mathbf{m}}$, defined by (7.9), comes from indices $i \in J^-$ such that $x_i \notin [E-b, E+b] + O(1/N)$. It satisfies the same bound (7.8) as $\Omega^+_{J,\mathbf{m}}$.

By the continuity of ϱ, the density of γ_i's is bounded by CN, thus $|J^+ \setminus J^-| \le CN\xi$ and $|J \setminus J^-| \le CN\xi$. Therefore, summing up the formula (7.5) for $i \in J$, we obtain from (7.6) and (7.9)

$$\int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x}) f_\tau \mathrm{d}\mu \tag{7.10}$$

$$\le (2b)^{-1} \int \frac{1}{N} \sum_{i\in J} G\Big(N(x_i - x_{i+m_2}),\ldots\Big) f_\tau \mathrm{d}\mu + Cb^{-1}\xi + Cb^{-1}\xi^{-2}N^{-1-2\varepsilon} \tag{7.11}$$

for each $\mathbf{m} \in S_n$. A similar lower bound can be proved analogously and we obtain

$$\left| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x}) f_\tau \mathrm{d}\mu - (2b)^{-1} \int \frac{1}{N} \sum_{i\in J} G\Big(N(x_i - x_{i+m_2}),\dots\Big) f_\tau \mathrm{d}\mu \right| \\ \le Cb^{-1}\xi + Cb^{-1}\xi^{-2}N^{-1-2\varepsilon} \tag{7.12}$$

for each $\mathbf{m} \in S_n$.

Adding up (7.12) for all $\mathbf{m} \in S_n(M)$, we get

$$\left| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \sum_{\mathbf{m}\in S_n(M)} \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x}) f_\tau \mathrm{d}\mu \right. \\ \left. - (2b)^{-1} \int \sum_{\mathbf{m}\in S_n(M)} \frac{1}{N} \sum_{i\in J} G\Big(N(x_i - x_{i+m_2}),\dots\Big) f_\tau \mathrm{d}\mu \right| \le Cb^{-1}\xi + Cb^{-1}\xi^{-2}N^{-1-2\varepsilon}, \tag{7.13}$$

and the same estimate holds for the equilibrium, i.e., if we set $\tau = \infty$ in (7.13). We now subtract the these two formulas and apply (4.3) from Theorem 4.1 to each summand on the second term in (7.13). Choosing $\xi = N^{-(1+2\varepsilon)/3}$ to minimize the two error terms involving ξ, we conclude that

$$\Theta_M^{(1)} = \left| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int \sum_{\mathbf{m}\in S_n(M)} \sum_{i=1}^{N} Y_{i,\mathbf{m}}(E',\mathbf{x}) (f_\tau \mathrm{d}\mu - \mathrm{d}\mu) \right| \\ \le CM^{n-1}\Big(b^{-1}N^{-\frac{1+2\varepsilon}{3}} + b^{-1/2}N^{\varepsilon'-\delta/2}\Big). \tag{7.14}$$

where we have used $\tau = N^{-2\varepsilon+\delta}$ and that $|J| \le CNb$.

Step 2. Large $\mathbf{m}$ *case.*

For a fixed $y \in \mathbb{R}$, $\ell > 0$, let

$$\chi(y,\ell) := \sum_{i=1}^{N} \mathbf{1}\Big\{x_i \in \big[y - \frac{\ell}{N}, y + \frac{\ell}{N}\big]\Big\}$$

denote the number of points in the interval $[y - \ell/N, y + \ell/N]$. Note that for a fixed $\mathbf{m} = (m_2,\dots,m_n)$, we have

$$\sum_{i=1}^{N} |Y_{i,\mathbf{m}}(E',\mathbf{x})| \le C \cdot \chi(E',\ell) \cdot \mathbf{1}\Big(\chi(E',\ell) \ge m_n\Big) \le C \sum_{m=m_n}^{\infty} m \cdot \mathbf{1}\Big(\chi(E',\ell) \ge m\Big), \tag{7.15}$$

where ℓ denotes the maximum of $|u_1| + \ldots + |u_n|$ in the support of $\widetilde{O}(u_1,\dots,u_n)$.

Since the summation over all increasing sequences $\mathbf{m} = (m_2, \ldots, m_n) \in \mathbb{N}_+^{n-1}$ with a fixed m_n contains at most m_n^{n-2} terms, by definition (7.4) we have

$$\Theta_M^{(2)}(\tau) \le C \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \sum_{m=M}^{\infty} m^{n-1} \int \mathbf{1}\Big(\chi(E', \ell) \ge m\Big) f_\tau \mathrm{d}\mu. \tag{7.16}$$

Now we use Assumption IV for the interval $I = [E' - N^{-1+\sigma}, E' + N^{-1+\sigma}]$ with σ chosen in such a way that $N^\sigma \le M^2$. Clearly $\mathcal{N}_I \ge \chi(E', \ell)$ for sufficiently large N, thus we get from (2.14) that

$$\sum_{m=M}^{\infty} m^{n-1} \int \mathbf{1}\Big(\chi(E', \ell) \ge m\Big) f_\tau \mathrm{d}\mu \le C_a \sum_{m=M}^{\infty} m^{n-1} \Big(\frac{m}{N^\sigma}\Big)^{-a}$$

holds for any $a \in \mathbb{N}$. By the choice of σ, we get that $\sqrt{m} \ge N^\sigma$ for any $m \ge M$, and thus choosing $a = k(n+1)$, we get

$$\Theta_M^{(2)}(\tau) \le \frac{C_a}{M^{k-1}}.$$

Together with (7.14), we have thus proved that

$$\Theta \le C M^{n-1} \Big(b^{-1} N^{-\frac{1+2\varepsilon}{3}} + b^{-1/2} N^{\varepsilon' - \delta/2}\Big) + \frac{C_a}{M^{k-1}}. \tag{7.17}$$

Choosing M such that $M^n = N^{\varepsilon'}$ and then choose k large enough so that the last term $\frac{C_a}{M^{k-1}}$ is smaller than, say, N^{-2}. We have thus proved that

$$\Theta \le C N^{2\varepsilon'} [b^{-1} N^{-\frac{1+2\varepsilon}{3}} + b^{-1/2} N^{-\delta/2}] \tag{7.18}$$

for $\tau = N^{-2\varepsilon+\delta}$ and this concludes (2.15).

For the proof of (2.17), we choose $\xi \ge 2N^{-1+A}$, and then by using (2.16) we can estimate $\Omega_{J,\mathbf{m}}^+$ directly as

$$\int |\Omega_{J,\mathbf{m}}^+(\mathbf{x})| f_\tau \mathrm{d}\mu \ll N^{-K} \tag{7.19}$$

for any $K > 0$, instead of (7.8). Therefore, the estimate on the right hand side of (7.12) and the subsequent estimates can be replaced by

$$C b^{-1} \xi + C b^{-1} N^{-K} \tag{7.20}$$

provided $\xi \ge 2N^{-1+A}$. Choosing $\xi = 2N^{-1+A}$ and following the same proof, we can improve the estimate (7.18) to

$$\Theta \le C N^{2\varepsilon'} [b^{-1} N^{-1+A} + b^{-1/2} N^{-\delta/2}] \tag{7.21}$$

for $\tau = N^{-2\varepsilon+\delta}$. This proves (2.17) and we have completed the proof of Theorem 2.1.

□

8 Local Marchenko-Pastur law

In this section we establish that the empirical density of eigenvalues for sample covariance matrices is close to the Marchenko-Pastur law even on short scale. We do this by controlling the difference of the Stieltjes transform, establishing results analogous to Theorem 4.1. and Proposition 4.2 of [16]. In this section, we focus on $0 < d = N/M < 1$, in particular the lower spectral edge $\lambda_- > 0$. The constants appearing in this subsection may depend on d.

Before the detailed proof, we explain the main steps of the argument which is similar to the method we have successively developed in [17, 18, 19]. The proof given here is somewhat complicated by fact that the matrix elements themselves are not independent but are generated as a quadratic expression of independent random variables. The first step, Lemma 8.1, is an apriori bound on the local density on short scales, $\eta \gg 1/N$, using resolvent expansion and a large deviation principle for quadratic forms. Expressing the resolvent of H in terms the resolvents of its minors, we obtain a self-consistent equation (8.20) for the Stieltjes transform m_N of the eigenvalues. This equation is very close to the defining quadratic equation of the Stieltjes transform m_W of the Marchenko-Pastur law, see (8.7), with a perturbation term $Y(z)$. This term can be estimated by large deviation arguments and using the a-priori bound on the local density. Then in Lemma 8.3 we investigate the stability of the self-consistent equation for m_W. Although the perturbed equation has two solutions, only one of them can be close to m_N. To select the correct solution, we use a continuity argument in the spectral parameter z. For $z = z_0$ with a large imaginary part, say $z_0 = 10 + 5i$, the explicit formula (8.27) for the solution can be directly analyzed. For z approaching to the real axis, we prove that the two unperturbed solutions remain far away from each other (8.35). Since the perturbed solutions are also continuous in the spectral parameter, for a sufficiently small perturbation they must remain in the vicinity of the correct solution of the unperturbed equation.

This analysis yields a bound on the difference of Stieltjes transforms, $m_N - m_W$. In Lemma 8.5 we give a better bound on $\mathbb{E}m_N - m_W$. The improvement is due to the fact that the perturbation term Y in the self-consistent equation is random and its expectation is much smaller than its typical size (compare (8.17) and (8.48)). Finally, in Lemma 8.6 we give an independent estimate on $\mathbb{E}m_N - m_W$ that is weaker in terms of $\eta = \text{Im} z$ but it is weaker in κ. When we will verify Assumption III in the following Section 9, we will use both bounds simultaneously.

Lemma 8.1 *Let $0 < E < 10$ and $0 < d < 1$. Consider the interval $I_\eta = [E - \eta, E + \eta]$. Let $\mathcal{N}_I$ denote the number of eigenvalues of $H = A^*A$ in the interval I_η. Suppose that $N^{-1+\varepsilon} \leq \eta \leq E/2$, for some $\varepsilon > 0$. Then there exist constants $C, c > 0$ such that*

$$\mathbb{P}\left(\mathcal{N}_{I_\eta} \geq \frac{KN\eta}{\sqrt{E}}\right) \leq Ce^{-c\sqrt{KN\eta/\sqrt{E}}}, \tag{8.1}$$

for all N, K large enough (independent of E).

We remark that the assumption on η can be relaxed to $CN^{-1} \le \eta \le E/2$. But we do not need this result here. For details, one can refer to Theorem 5.1 of [19].

Proof of Lemma 8.1. We observe, first of all, that

$$\frac{\mathcal{N}_I}{N\eta} \le \frac{C}{N} \mathrm{Im}\,\mathrm{Tr}\ \frac{1}{H - E - i\eta} = \frac{C}{N} \mathrm{Im} \sum_{j=1}^{N} \frac{1}{H - z}(j, j),$$

where we defined $z = E + i\eta$. It follows that

$$\mathbb{P}\left(\mathcal{N}_I \ge KN\eta/\sqrt{E}\right) \le N\mathbb{P}\left(\left|\mathrm{Im}\,\frac{1}{H - z}(1,1)\right| \ge K/\sqrt{E}\right). \tag{8.2}$$

Denoting by a_1 the first column of A and by B the $M \times (N-1)$ matrix consisting of the last $N-1$ column of A, we have

$$H = \begin{pmatrix} a_1 \cdot a_1 & (B^* a_1)^* \\ B^* a_1 & B^* B \end{pmatrix}.$$

Hence

$$\frac{1}{H - z}(1,1) = \frac{1}{a_1 \cdot a_1 - z - a_1 \cdot B(B^*B - z)^{-1} B^* a_1}.$$

Using the identity

$$B(B^*B - z)^{-1}B^* = BB^*(BB^* - z)^{-1},$$

we find

$$\frac{1}{H - z}(1,1) = \frac{1}{a_1 \cdot a_1 - z - a_1 \cdot BB^*(BB^* - z)^{-1} a_1}. \tag{8.3}$$

Denote μ_α's ($\alpha = 1, \ldots, N-1$) the eigenvalues of the $(N-1) \times (N-1)$ matrix B^*B. The μ_α's are also the eigenvalues of $M \times M$ matrix BB^* and the other eigenvalues of BB^* are zeros. Then define v_α ($\alpha = 1, \ldots, N-1$) as the normalized eigenvectors of BB^* associated with non-zero eigenvalues μ_α, i.e., the matrix elements of BB^* are given by

$$(BB^*)_{ij} = \sum_{\alpha=1}^{N-1} \mu_\alpha \bar{v}_\alpha(i) v_\alpha(j). \tag{8.4}$$

Inserting (8.4) into (8.3), we find

$$\frac{1}{H - z}(1,1) = \frac{1}{a_1 \cdot a_1 - z - \frac{1}{M}\sum_{\alpha=1}^{N-1} \frac{\mu_\alpha \xi_\alpha}{\mu_\alpha - z}},$$

where we defined the quantity $\xi_\alpha = M|a_1 \cdot v_\alpha|^2$ (note that $\mathbb{E}\xi_\alpha = 1$). Taking the imaginary part, we find

$$\left|\mathrm{Im}\,\frac{1}{H - z}(1,1)\right| \le \frac{1}{\eta + \frac{\eta}{M}\sum_{\alpha=1}^{N-1} \frac{\mu_\alpha \xi_\alpha}{(\mu_\alpha - E)^2 + \eta^2}} \le \frac{CN\eta}{E\sum_{\alpha:|\mu_\alpha - E| \le \eta} \xi_\alpha},$$

where we used the assumption that $\eta < E/2$. Because the eigenvalues μ_α's are the eigenvalues of the $(N-1)\times(N-1)$ matrix B^*B, they are interlaced with the eigenvalues of H and $|\{\alpha : |\mu_\alpha - E| \le \eta/2\}| \ge \mathcal{N}_I - 1$. It follows from (8.2) that

$$\mathbb{P}\left(\mathcal{N}_I \ge \frac{KN\eta}{\sqrt{E}}\right) \le N\mathbb{P}\left(\sum_{\alpha:|\mu_\alpha - E|\le \eta/2} \xi_\alpha \le \frac{CN\eta}{K\sqrt{E}} \text{ and } \mathcal{N}_I \ge \frac{KN\eta}{\sqrt{E}}\right) \le CNe^{-c\sqrt{KN\eta/\sqrt{E}}} ,$$

where in the last step we used Lemma 4.7 from [19]. The claim follows by the assumption that $N\eta \ge N^\varepsilon$ and that N and K are large enough. □

Proposition 8.1 *Consider sample covariance matrices $H = A^*A$ with A an $M \times N$ matrix with independent and identically distributed complex entries. Let $0 < d < 1$. Recall λ_- and λ_+ in* (3.13) *and define κ as*

$$\kappa = \kappa(E) := |(E - \lambda_-)(E - \lambda_+)| . \tag{8.5}$$

We will often drop the argument E from the notation of κ for brevity. Then for any E, η satisfying $N^{-1+\varepsilon} \le \eta \le \frac{1}{2}E$, $\frac{1}{2}\lambda_- \le E \le 10$, the Stieltjes transform,

$$m_N(z) := \frac{1}{N} Tr\, \frac{1}{H - z} , \quad z = E + i\eta,$$

of the empirical eigenvalue distribution of $H = A^*A$ satisfies

$$\mathbb{P}\left(|m_N(E+i\eta) - m_W(E+i\eta)| \ge \frac{\delta}{\sqrt{\kappa + \delta}}\right) \le Ce^{-c\delta\sqrt{N\eta}} , \tag{8.6}$$

for any δ small enough (independent of E and η) and $N \ge 2$. Here $m_W(z)$ is the unique solution of

$$m_W(z) + \frac{1}{z - (1-d) + z\, d\, m_W(z)} = 0, \tag{8.7}$$

with positive imaginary part for all z with $Im\ z > 0$.

Recall $\lambda_\pm = (1 \pm d^{1/2})^2$ from (3.13). The function m_W defined in (8.7) depends on d and can be written as

$$m_W(z) = \frac{1 - d - z + i\sqrt{(z-\lambda_-)(\lambda_+ - z)}}{2\, d\, z}, \tag{8.8}$$

where $\sqrt{\ }$ denotes the square root on complex plane whose branch cut is the negative real line. Explicit calculation shows $m_W(z)$ is the Stieltjes transform of the Marchenko-Pastur density given in (3.14).

Using (8.6) and (3.14), we have the local Marchenko-Pastur law for the number of eigenvalues in a small interval:

Corollary 8.2 *Consider an interval $I = [E - \eta, E + \eta] \subset [\lambda_-, \lambda_+]$ within the bulk spectrum. Let δ be a suffiiciently small parameter. Suppose that E and η are chosen such that $\delta^{-2} N^{-1+\varepsilon} \le \eta \le C^{-1} \min\{\kappa, \delta^{1/2}\kappa^{3/4}\}$ with a large constant C and with $\kappa = \kappa(E)$ given in (8.5). Then we have the convergence of the counting function, i.e.,*

$$\mathbb{P}\left\{ \left| \frac{\mathcal{N}_\eta(E)}{2\eta N} - \rho_W(E) \right| \ge \delta \right\} \le C e^{-c\delta^2 \sqrt{N\eta\kappa}}, \tag{8.9}$$

where $\mathcal{N}_\eta(E) = |\{\lambda_\alpha : |\lambda_\alpha - E| \le \eta\}|$ denotes the number of eigenvalues of $H = A^*A$ in the interval $I = [E - \eta, E + \eta]$.

Proof of Corollary 8.2. The proof of (8.9) follows from the inequality (8.6) with a similar argument as the proof of Proposition 4.1 of [16]. □

We remark that, similarly to Theorem 3.1 in [19], the assumption on the lower bound $\eta \ge N^{-1+\varepsilon}$ can be relaxed to $\eta \ge KN^{-1}$ and obtain the local Marchenko-Pastur law on the shortest possible scale, at least away from the spectral edges.

Proof of Proposition 8.1. Let a_j be the j-th column of A and let $B^{(j)}$ be the remaining $M \times (N-1)$ matrix obtained from A after removing the j-th column a_j. Let $\mu_\alpha^{(j)}$, $v_\alpha^{(j)}$ be the non-zero eigenvalues and the eigenvectors of the matrix $B^{(j)}[B^{(j)}]^*$ and we define $\xi_\alpha^{(j)} = M|a_j \cdot v_\alpha^{(j)}|^2$. Then we have the formula

$$m_N(z) = \frac{1}{N} \mathrm{Tr}\, \frac{1}{H - z} = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{H - z}(j,j) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{a_j \cdot a_j - z - \frac{1}{M} \sum_{\alpha=1}^{N-1} \frac{\mu_\alpha^{(j)}}{\mu_\alpha^{(j)} - z} \xi_\alpha^{(j)}}$$

that we rewrite as

$$m_N(z) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{a_j \cdot a_j - z - \frac{N-1}{M} - \frac{z}{M} \sum_{\alpha=1}^{N-1} \frac{1}{\mu_\alpha^{(j)} - z} - X^{(j)}},$$

with

$$X^{(j)} = X^{(j)}(z) = \frac{1}{M} \sum_{\alpha=1}^{N-1} \frac{\mu_\alpha^{(j)}}{\mu_\alpha^{(j)} - z} \left(\xi_\alpha^{(j)} - 1 \right). \tag{8.10}$$

Note that the vector a_j is independent of $\mu_\alpha^{(j)}$ and $v_\alpha^{(j)}$. Therefore, we have

$$\mathbb{E} X^{(j)} = 0, \ \ \mathbb{E} \xi_\alpha^{(j)} = 1. \tag{8.11}$$

Define $m_{N-1}^{(j)}(z) \equiv \frac{1}{N-1} \mathrm{Tr}([B^{(j)}]^* B^{(j)} - z)^{-1}$, then

$$\sum_{\alpha=1}^{N-1} \frac{1}{\mu_\alpha^{(j)} - z} = (N-1) m_{N-1}^{(j)}(z) \,.$$

Hence

$$m_N(z) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{1 - z - d - z\, d\, m_N(z) + Y^{(j)}}\ , \tag{8.12}$$

with

$$Y^{(j)} = Y^{(j)}(z) = (a_j \cdot a_j - 1) + \frac{1}{M} - \frac{z}{M}\Big((N-1)m_{N-1}^{(j)}(z) - Nm_N(z)\Big) - X^{(j)}(z). \tag{8.13}$$

For fixed j, denote $b = \sqrt{M}a_j$ with $b = (b_1, \ldots, b_M)$. Drop the superscript j for $\mu_\alpha = \mu_\alpha^{(j)}$, $v_\alpha = v_\alpha^{(j)}$, $B = B^{(j)}$ and $m_{N-1} = m_{N-1}^{(j)}$ for simplicity. We rewrite $X^{(j)}$ as

$$X^{(j)} = \sum_{\ell,k=1}^{M} \sigma_{\ell k}[b_\ell \bar{b}_k - \mathbb{E} b_\ell \bar{b}_k]$$

with

$$\sigma_{\ell k} := \frac{1}{M}\sum_{\alpha=1}^{N-1} \frac{\mu_\alpha \bar{v}_\alpha(\ell) v_\alpha(k)}{\mu_\alpha - z}.$$

So with $\frac{1}{2}\lambda_- \le E = Re(z) \le 10$ and $N^{-1+\varepsilon} \le \eta \le E/2$, we have

$$\sum_{\ell,k} |\sigma_{\ell k}|^2 = \frac{1}{M^2}\sum_\alpha \frac{\mu_\alpha^2}{|\mu_\alpha - z|^2} \le \frac{CK}{M\eta},$$

with some fixed large K (using dyadic decomposition and (8.1), similarly to the argument in Lemma 4.2 of [19]) apart from an event of probability $e^{-c\sqrt{N\eta}}$. Then with Proposition 4.5 of [19], we have

$$\mathbb{P}(\max_j |X^{(j)}| \ge \delta) \le Ce^{-c\delta\sqrt{N\eta}}, \tag{8.14}$$

for sufficiently small $\delta > 0$. Since the eigenvalues of B^*B are interlaced with the eigenvalues of $H = A^*A$, we have

$$|(N-1)m_{N-1}(z) - Nm_N(z)| \le C\eta^{-1}. \tag{8.15}$$

Then using $\mathbb{E}a_j \cdot a_j = \frac{1}{M}\mathbb{E}\sum_{i=1}^{M}|b_i|^2 = 1$ and Proposition 4.5 of [19] for the iid variables b_i's, we obtain

$$\mathbb{P}\left(|a_j \cdot a_j - 1| \ge KM^{-1/2}\right) \le Ce^{-c\min\{K,K^2\}}. \tag{8.16}$$

Combining (8.14), (8.15), and (8.16), we find that

$$\mathbb{P}\left(\left|Y^{(j)}(z)\right| \ge \delta\right) \le Ce^{-c\delta\sqrt{N\eta}} + Ce^{-c\delta^2 N}\ ,$$

for sufficiently small $\delta > 0$. With the assumption $\eta \le Re(z)/2 = E/2$ and $N\eta \ge N^{\varepsilon}$, we obtain

$$\mathbb{P}\left(\max_j \left|Y^{(j)}(z)\right| \ge \delta\right) \le Ce^{-c\delta\sqrt{N\eta}}. \tag{8.17}$$

On the other hand, with the definition of $Y^{(j)}$, for any j, z, z' such that, $|z|, |z'| \le 10$, $\mathrm{Im}(z), \mathrm{Im}(z') \ge \eta$, we have

$$\mathbb{P}\left(\left|Y^{(j)}(z) - Y^{(j)}(z')\right| \ge \eta^{-2}|z - z'|\right) \le e^{-c\sqrt{N\eta}}. \tag{8.18}$$

Together with (8.17), we obtain, for $N^{-1+\varepsilon} \le \eta \le \frac{1}{2}E$, $\frac{1}{2}\lambda_- \le E \le 10$ and sufficiently small δ,

$$\mathbb{P}\left(\max_{z' \in L(z, P_{10})} \max_j \left|Y^{(j)}(z')\right| \ge \delta\right) \le Ce^{-c\delta\sqrt{N\eta}}, \quad P_{10} = 10 + 5i, \tag{8.19}$$

where $L(z, P_{10})$ is the line segment connecting points z and P_{10}. Then the Proposition 8.1 follows from the next lemma. □

Lemma 8.3 *Assume H is a $N \times N$ positive semidefinite matrix with $\|H\| \le 5$. For fixed $0 < d < 1$, we recall the notation $\lambda_\pm = (1 \pm \sqrt{d})^2$. Let $z_0 = E + i\eta$ and $N^{-1+\varepsilon} \le \eta \le \frac{1}{2}E$, $\frac{1}{2}\lambda_- \le E \le 10$. Denote $L(z_0, P_{10})$ the line segment connecting z_0 and $P_{10} = 10 + 5i$. Suppose that for any $z \in L(z_0, P_{10})$, the Stieltjes transform $m_N(z) = \frac{1}{N}\mathrm{Tr}(H - z)^{-1}$ satisfies the following self-consistent relation:*

$$m_N(z) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{1 - z - d - z\, d\, m_N(z) + Y^{(j)}(z)}, \tag{8.20}$$

for some $Y^{(j)}(z)$'s. Then there exists $\delta_0 > 0$ depending only on d, such that, whenever

$$\delta \equiv \max_{z \in L(z_0, P_{10})} \max_j \left|Y^{(j)}(z)\right| \le \delta_0, \tag{8.21}$$

we have

$$|m_N(z_0) - m_W(z_0)| \le C\delta(\kappa + \delta)^{-1/2} \tag{8.22}$$

with $\kappa = \kappa(E) := |(\lambda_+ - E)(E - \lambda_-)|$.

Proof of Lemma 8.3. We begin with a special case: $z_0 = P_{10}$. In this case if $z \in L(z_0, P_{10})$ then $z = z_0 = P_{10}$. With the assumptions on H and $0 < d < 1$, it is easy to see that:

$$\left|-m_N(z)d + \frac{1 - d}{z} - 1\right| \ge \frac{1}{2}, \tag{8.23}$$

which implies

$$|1 - z - d - z\, d\, m_N(z)| \ge \frac{1}{2}|z|. \tag{8.24}$$

Insert it into (8.20), we obtain when $z = P_{10}$,

$$\left| m_N(z) + \frac{1}{z - (1-d) + z\, d\, m_N(z)} \right| \le C\delta. \tag{8.25}$$

Denote the solutions of

$$S + \frac{1}{z - (1-d) + z\, d\, S} = \Delta \tag{8.26}$$

by $S^{\Delta}_{\pm}(z)$. Explicit calculation shows

$$S^{\Delta}_{\pm}(z) = \frac{1 - d - z \pm i(1 + d\Delta)\sqrt{(\lambda^{\Delta}_{+} - z)(z - \lambda^{\Delta}_{-})}}{2dz} + \frac{\Delta}{2}, \tag{8.27}$$

where

$$\lambda^{\Delta}_{\pm} \equiv \left(\frac{\sqrt{1 + \Delta(d - d^2)} \pm \sqrt{d}}{1 + \Delta d} \right)^2. \tag{8.28}$$

With the notations:

$$S_{\pm}(z) \equiv S^{0}_{\pm}(z), \tag{8.29}$$

we note $m_W(z) = S_+(z)$ (see (8.7)). Then the following lemma implies, with $\mathrm{Im}(m_N(z)) > 0$, $\mathrm{Im} S_+(P_{10}) > 0$, $\mathrm{Im} S_-(P_{10}) < 0$, that (8.22) holds for $z_0 = P_{10}$ if δ is small enough.

Lemma 8.4 *Let $S^{\Delta}_{\pm}(z)$ be the solutions of* (8.26)*. Let $z = E + i\eta$ and $\frac{1}{2}\lambda_- \le E \le 10$. For sufficiently small Δ, depending on d,*

$$\max\left\{ |S^{\Delta}_{+}(z) - S_+(z)|,\ |S^{\Delta}_{-}(z) - S_-(z)| \right\} \le C \frac{\Delta}{\sqrt{\kappa(E) + \Delta}}. \tag{8.30}$$

Proof of Lemma 8.4. First, when Δ is small enough, an easy calculation shows

$$\widetilde{\Delta} \equiv \max_{\pm} \left\{ |\lambda^{\Delta}_{\pm} - \lambda_{\pm}| \right\} \le C|\Delta|. \tag{8.31}$$

Therefore, we have

$$\max\left\{ |S^{\Delta}_{+}(z) - S_+(z)|,\ |S^{\Delta}_{-}(z) - S_-(z)| \right\} \le C|\Delta| + C\left| \sqrt{(\lambda^{\Delta}_{+} - z)(z - \lambda^{\Delta}_{-})} - \sqrt{(\lambda_+ - z)(z - \lambda_-)} \right|. \tag{8.32}$$

Let $a = (z - \lambda_-)(\lambda_+ - z)$ and $b = (z - \lambda^{\Delta}_{-})(\lambda^{\Delta}_{+} - z) - (z - \lambda_-)(\lambda_+ - z)$. Note that $|b| \le C\widetilde{\Delta}$ and therefore, by (8.31), $|b| \le C\Delta$. Hence, (8.30) follows from $|(\lambda_+ - z)(z - \lambda_-)| \ge C\kappa(E)$ and from the inequality

$$\left| \sqrt{a + b} - \sqrt{a} \right| \le C \frac{|b|}{\sqrt{|a| + |b|}}. \tag{8.33}$$

which holds for any complex number a and b. $\square$

Now we prove (8.22) for the case $z_0 \neq P_{10}$. We first note that the two solutions of (8.20) are $S_\pm(z)$ when $Y^{(j)} = 0$. One can check that for $z \in L(z_0, P_{10})$, these two solutions are bounded by some constant C_1:

$$|S_\pm(z)| \le C_1, \quad \text{i.e.,} \;\; |z - (1-d) + zdS_\pm(z)| \ge C_1^{-1}, \tag{8.34}$$

and

$$|S_-(z) - S_+(z)| \ge C\sqrt{\kappa(\operatorname{Re} z) + \operatorname{Im} z}. \tag{8.35}$$

Furthermore, $|S_-(z_0) - S_+(z_0)|$ can be bounded by $|S_-(z) - S_+(z)|$ for any $z \in L(z_0, P_{10})$ as follows,

$$|S_-(z_0) - S_+(z_0)| \le C \min_{z \in L(z_0, P_{10})} |S_-(z) - S_+(z)|. \tag{8.36}$$

On the other hand, for any $z \in L(z_0, P_{10})$, we claim that if $m_N(z)$ is close to $S_-(z)$ or $S_+(z)$, then it should be really close to $S_-(z)$ or $S_+(z)$, i.e., if

$$\min\{|m_N(z) - S_-(z)|, \;\; |m_N(z) - S_+(z)|\} \le C_1^{-1}/20, \tag{8.37}$$

then

$$\min\{|m_N(z) - S_-(z)|, \;\; |m_N(z) - S_+(z)|\} \le C\frac{\delta}{\sqrt{\kappa(\operatorname{Re} z) + \delta}}. \tag{8.38}$$

To see this, note that (8.37) together with (8.34) imply

$$|z - (1-d) + zdm_N(z)| \ge \frac{1}{2}C_1^{-1}. \tag{8.39}$$

Then with (8.20) and (8.21), we obtain that (8.25) holds for any $z \in L(z_0, P_{10})$. Using Lemma 8.4 again, we have (8.38).

We have seen that (8.22) and (8.37) (for small δ) hold when $z = P_{10}$. Because $m_N(z)$, $S_\pm(z)$ are continuous functions of z, with (8.38) and (8.37), we can see that when δ is small enough, (8.38) holds for every $z \in L(z_0, P_{10})$. This result shows that $m_N(z)$ must be close to at least one of $S_+(z)$ and $S_-(z)$ and it is close to $S_+(z)$ when $z = P_{10}$.

Now we claim that if $m_N(z_0)$ were close to $S_-(z_0)$, i.e.,

$$|m_N(z_0) - S_-(z_0)| \le C\frac{\delta}{\sqrt{\kappa(E) + \delta}}, \tag{8.40}$$

then $m_N(z_0)$ is also close to $S_+(z_0)$, which implies that $m_N(z_0)$ is always close to $S_+(z_0)$.

Again, with the continuity of $m_N(z)$ and $S_\pm(z)$ and (8.38), if $m_N(z_0)$ is close to $S_-(z_0)$ in the sense of (8.40), then there exists $z \in L(z_0, P_{10})$ such that $m_N(z)$ is close to both of $S_-(z)$ and $S_+(z)$, i.e.,

$$|S_+(z) - S_-(z)| \le 2C\frac{\delta}{\sqrt{\kappa(\operatorname{Re} z) + \delta}}, \tag{8.41}$$

which implies

$$|S_+(z) - S_-(z)| \le 2C \frac{\delta}{\sqrt{C\kappa(E)+\delta}}. \tag{8.42}$$

Combining (8.42) and (8.36), we obtain

$$|S_+(z_0) - S_-(z_0)| \le C \frac{\delta}{\sqrt{\kappa(E)+\delta}}. \tag{8.43}$$

Together with (8.40), we obtain $m_N(z_0)$ is still close to the $S_+(z_0)$. It means that (8.22) holds for all z_0's in our assumption, using the fact $S_+(z_0) = m_W(z_0)$. This completes the proof of Lemma 8.3. □

The following lemma shows that the expectation value of $m_N(z)$ is close to $m_W(z)$.

Lemma 8.5 *Let $z = E + i\eta$, such that $N^{-1+\varepsilon} \le \eta \le \frac{1}{2}E$, $\frac{1}{2}\lambda_- \le E \le 10$, for some $\varepsilon > 0$. Then we have*

$$|\mathbb{E}m_N(z) - m_W(z)| \le \frac{C}{N\eta\kappa^{3/2}}, \qquad \kappa = |(\lambda_+ - E)(E - \lambda_-)|, \tag{8.44}$$

for large enough N depending on ε.

Proof of Lemma 8.5. Using (8.34) and the estimate (8.6) from Proposition 8.1, we have

$$|\mathbb{E}\, m_N(z)| \le C, \ \ |m_W(z)| \le C \tag{8.45}$$

uniformly in $z = E + i\eta$ within the range $N^{-1+\varepsilon} \le \eta \le \frac{1}{2}E$, $\frac{1}{2}\lambda_- \le E \le 10$.

We can assume that $N\eta\kappa^{3/2}$ is much greater than 1, otherwise (8.44) is trivial. Combining $N\eta\kappa^{3/2} > 1$ and $N\eta > N^{\varepsilon}$, we obtain $N\eta\kappa \ge N^{\varepsilon/3}$. Using (8.12), we write $\mathbb{E}m_N(z)$ as

$$\mathbb{E}m_N(z) = -\frac{1}{N}\mathbb{E}\left(\sum_{j=1}^{N} \frac{1}{B - zd(m_N(z) - \mathbb{E}\, m_N(z)) + Y^{(j)}(z)}\right), \tag{8.46}$$

where $B \equiv z - (1-d) - z\, d\, \mathbb{E}m_N(z)$. Then, with (8.6), we know

$$\mathbb{E}\, |m_N - \mathbb{E}m_N|^2 \le O\left(\frac{1}{N\eta\kappa}\right). \tag{8.47}$$

From (8.10), (8.11), (8.13), (8.15) and (8.17), we obtain

$$\mathbb{E}Y^{(j)}(z) = \mathbb{E}\left(\frac{1}{M} - \frac{z}{M}\left((N-1)m^{(j)}_{N-1}(z) - Nm_N(z)\right)\right) = O\left(\frac{1}{N\eta}\right), \tag{8.48}$$

and

$$\mathbb{E}\left|Y^{(j)}\right|^2 \leq O\left(\frac{1}{N\eta}\right). \tag{8.49}$$

Using (8.6), (8.34) and $N\eta\kappa \geq N^{\varepsilon/3}$, we obtain that $|B|$ is bounded from below by a constant C_0. Furthermore, for some $\delta > 0$,

$$\mathbb{P}\left(|B - zd(m_N(z) - \mathbb{E}(m_N(z))) + Y^{(j)}(z)| \leq C_0/2\right) \leq Ce^{-cN^\delta}. \tag{8.50}$$

Denote $a = -zd(m_N(z) - \mathbb{E}m_N(z)) + Y^{(j)}(z)$, then

$$\mathbb{E}'\left(\frac{1}{B+a}\right) = \frac{1}{B} - \mathbb{E}'(a)B^{-2} + O(\mathbb{E}'a^2)|B|^{-3}, \tag{8.51}$$

where $\mathbb{E}'$ is the conditional expectation under the condition: $|B - zd(m_N(z) - \mathbb{E}m_N(z)) + Y^{(j)}(z)| \geq C_0/2$. Because $m_N(z)$ and $Y^{(j)}(z)$ are bounded from above by a polynomial of M, inserting (8.50) into (8.51), we obtain

$$\left|\mathbb{E}m_N(z) + \frac{1}{z - (1-d) + zd\mathbb{E}m_N(z)}\right| \leq C|\mathbb{E}(a)| + C\mathbb{E}a^2. \tag{8.52}$$

Combining this with (8.47), (8.48) and (8.49), we obtain:

$$\left|\mathbb{E}m_N(z) + \frac{1}{z - (1-d) + zd\mathbb{E}m_N(z)}\right| \leq \frac{C}{N\eta\kappa}. \tag{8.53}$$

Using Lemma 8.4, we have

$$\min\left\{|\mathbb{E}m_N(z) - S_+(z)|,\ |\mathbb{E}m_N(z) - S_-(z)|\right\} \leq \frac{C}{N\eta\kappa^{3/2}}. \tag{8.54}$$

Using this inequality, and $S_+(z) = m_W(z)$, we can easily obtain (8.44) for $z = P_{10}$. Consider now $z = E + i\eta \neq P_{10}$. If $\mathbb{E}m_N(z)$ is closer to $S_-(z)$ than $C(N\eta\kappa^{3/2})^{-1}$, then by the continuity of $\mathbb{E}m_N(z)$, there exists $z' \in L(z, P_{10})$, such that, $\mathbb{E}m_N(z')$ is close to both of $S_+(z')$ and $S_-(z')$, i.e.,

$$|S_+(z') - S_-(z')| \leq \frac{C}{N\mathrm{Im}\, z'[\kappa(\mathrm{Re} z')]^{3/2}} \leq \frac{C}{N\eta\kappa^{3/2}}. \tag{8.55}$$

Together with (8.36), we obtain that $\mathbb{E}m_N(z)$ is also close to $S_+(z) = m_W(z)$ and complete the proof. $\square$

Now we give an alternative bound on $\mathbb{E}m_N(z)$.

Lemma 8.6 *Let $z = E + \eta i$, $N^{-1+\varepsilon} \le \eta \le E/2$, $\lambda_-/2 \le E \le 10$ and $\varepsilon > 0$. Suppose $N\kappa\eta \ge N^{\varepsilon'}$ for some $\varepsilon' > 0$, we have*

$$|\mathbb{E} m_N(z) - m_W(z)| \le \frac{C}{N\eta^{3/2}\kappa^{1/2}}, \tag{8.56}$$

when N is sufficiently large (depending on ε').

Proof of Lemma 8.6. We only prove the case of the real sample covariance matrix. The case of the complex sample covariance matrix can be treated similarly.

First, we show

$$\mathbb{E}|m_N(z) - \mathbb{E} m_N(z)| \le \frac{C}{N\eta^{3/2}}. \tag{8.57}$$

Let λ_α and u_α be the eigenvalues and eigenvectors of $H = A^*A$. The derivative of λ_α with respect to the (i,j)-th matrix element A_{ij} is given by

$$\frac{\partial \lambda_\alpha}{\partial A_{ij}} = 2(Au_\alpha)(i) u_\alpha(j). \tag{8.58}$$

Using $\sum_j u_\alpha(j) u_\beta(j) = \delta_{\alpha,\beta}$ and $\sum_i (Au_\alpha)(i)(Au_\beta)(i) = \lambda_\alpha \delta_{\alpha,\beta}$, one can obtain the following result, as in (3.3) of [17],

$$\mathbb{E}|m_N(z) - \mathbb{E} m_N(z)|^2 \le \frac{C}{N^4} \mathbb{E}\left(\sum_\alpha \frac{\lambda_\alpha}{|\lambda_\alpha - z|^4} \right). \tag{8.59}$$

Then with Lemma 8.1, as in (3.6) of [17], we obtain (8.57).

Let $B \equiv z - (1-d) + z\, d\, \mathbb{E} m_N(z)$, $a_1 = zd(m_N(z) - \mathbb{E} m_N(z))$ and $a_2 = Y_j$ for each j. Using the assumption $N\eta\kappa \ge N^{\varepsilon'}$ for some ε', we obtain that $|B|$ is bounded from below by a constant C_0 and for some $\delta > 0$,

$$\mathbb{P}\left(|B - a_1 - a_2| \le C_0/2\right) \le e^{-N^\delta}. \tag{8.60}$$

We have

$$\mathbb{E}' \frac{1}{B - a_1 - a_2} = \frac{1}{B} + O(B^{-2}(\mathbb{E}'|a_1|)) + \frac{1}{B^2}\mathbb{E}'(a_2) + O(B^{-3}\mathbb{E}'(a_2^2)), \tag{8.61}$$

where $\mathbb{E}'$ is the conditional expectation under the condition: $|B - zd(m_N(z) - \mathbb{E} m_N(z)) + Y^{(j)}(z)| \ge C_0/2$. Combining this with (8.60), (8.57), (8.48) and (8.49), we obtain

$$\left| \mathbb{E} m_N(z) + \frac{1}{z - (1-d) + z\, d\, \mathbb{E} m_N(z)} \right| \le \frac{C}{N\eta^{3/2}}. \tag{8.62}$$

As in the proof of Lemma 8.5, with a continuity argument and Lemma 8.4, we can obtain (8.56) from (8.62) and complete the proof. □

9 Verifying Assumption III

The following theorem gives the estimate (2.13) for the singular values of the sample covariance matrices.

Theorem 9.1 *Assume that the single site distribution* $\mathrm{d}\nu$ *of the entries of* $\sqrt{M}A$ *satisfies the logarithmic Sobolev inequality* (3.10). *Recall that* ρ_W *in* (3.14) *denotes the density in Marchenko-Pastur law. Define* $\gamma_j \in [\lambda_-, \lambda_+]$, $(j = 1, 2, \ldots, N)$ *with the relation*

$$\int_{-\infty}^{\gamma_j} \rho_W(x)dx = \frac{j}{N}. \tag{9.1}$$

Denote $x_j = \lambda_j^{1/2}$ *the singular values of* A. *Then there exists* $\delta > 0$, *such that,*

$$\frac{1}{N}\mathbb{E}\sum_{j=1}^{N}\left|x_j - \gamma_j^{1/2}\right|^2 \leq CN^{-1-\delta}. \tag{9.2}$$

Remark. An analogous result holds for the eigenvalues of the Wigner matrices; the proof is similar and we will not give the details here. We only point out that the key ingredients of the argument below are: (i) apriori bound on the extreme eigenvalues (see Lemma 9.2 and the remark afterwards); (ii) concentration of the local density of states (used in Lemma 9.3). We also critically use the fact that the density of states (semicircle law or Marchenko-Pastur law) has a square root singularity at the edges.

The local semicircle law needed in the analogue of Lemma 9.3 for Wigner matrices has been proven for hermitian matrices (see [16] and references therein) but the proof is valid for symmetric and quaternion self-dual matrices as well. The extension to symmetric matrices is trivial. For the case of quaternion self-dual matrices, the only additional observation is that the non-commutativity of the quaternions is irrelevant in the arguments because the common starting point of our papers [16, 17, 18, 19, 20] is an identity on the diagonal elements of the Green's function that involves only complex numbers. For simplicity, we present it only for the (1,1) diagonal element $G_z(1,1)$ of $(H-z)^{-1}$, where H is an $N \times N$ quaternion self-dual matrix and $z \in \mathbb{C}$:

$$G_z(1,1) = \frac{1}{h - z - \mathbf{a}^+ \cdot (B-z)^{-1}\mathbf{a}} \cong \left[h - z - \frac{1}{N}\sum_{\alpha=1}^{N-1}\frac{\xi_\alpha}{\lambda_\alpha - z}\right]^{-1}, \tag{9.3}$$

in particular, $G_z(1,1)$ is a diagonal quaternion thus it can be identified with a complex number via the identification (3.4). Here $h \in \mathbb{R}$, $\mathbf{a} \in \mathbb{H}^{N-1}$ and B is an $(N-1)\times(N-1)$ quaternion self-dual matrix obtained from the following decomposition of H

$$H = \begin{pmatrix} h & \mathbf{a}^+ \\ \mathbf{a} & B \end{pmatrix}. \tag{9.4}$$

The real numbers $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{N-1}$ denote the eigenvalues of B and the nonnegative real numbers ξ_α are given by

$$\xi_\alpha = N(\mathbf{a}^+ \cdot \mathbf{u}_\alpha)(\mathbf{u}_\alpha^+ \cdot \mathbf{a}) = N|\mathbf{a}^+ \cdot \mathbf{u}_\alpha|^2$$

where $\mathbf{u}_\alpha \in \mathbb{H}^{N-1}$ is the normalized eigenvector of B associated with the eigenvalue λ_α. The dot product of two quaternionic vectors, $\mathbf{a}, \mathbf{b} \in \mathbb{H}^{N-1}$ is defined as

$$\mathbf{a}^+ \cdot \mathbf{b} := \sum_{n=1}^{N-1} a_n^+ b_n.$$

The proof of (9.3) is a straightforward computation. This identity is the key to extend our results on the local semicircle law for quaternion self-dual matrices without any further modifications.

Now we return to the the proof of Theorem 9.1 and we start with some preparatory lemmas. First, we recall the following result from [22].

Lemma 9.2 *(Corollary V.2.1 of [22]) Define*

$$n^\lambda(E) \equiv \frac{1}{N}\mathbb{E}\left[\#\{\lambda_j \leq E\}\right], \tag{9.5}$$

then $n^\lambda(\lambda_- - N^{-1/5}) \leq Ce^{-N^\varepsilon}$ *and* $1 - n^\lambda(\lambda_+ + N^{-1/5}) \leq Ce^{-N^\varepsilon}$ *for some* $\varepsilon > 0$. *Therefore for any* $1 \leq j \leq N$,

$$\lambda_- - CN^{-1/5} \leq \mathbb{E}\lambda_j \leq \lambda_+ + CN^{-1/5}. \tag{9.6}$$

Remark. In fact, the error term in [22] is $N^{-2/3+\varepsilon}$ instead of $N^{-1/5}$ but we will use only the weaker bound (9.6) in order to indicate that our proof goes through for Wigner matrices with not necessarily symmetric distributions as well. In the latter case only $N^{-1/4+\varepsilon}$ has been proven by Vu in [42] for compactly supported distribution ν with an effective dependence of the constant C on the support. This effective dependence is necessary to remove the compact support condition as in Lemma 4.1 of [20]. Strictly speaking, the result in [42] was stated only for symmetric Wigner matrices but it holds for hermitian and quaternion self-dual Wigner matrices as well, since the key estimate (equation (5) in [42]) is independent of the matrix ensemble.

Lemma 9.3 *Recall that* ρ_W *in* (3.14) *denotes the density in Marchenko-Pastur law. Let*

$$n_W^\lambda(E) = \int_{-\infty}^{E} \rho_W(x)dx, \tag{9.7}$$

then

$$\int_0^\infty \left| n^\lambda(E) - n_W^\lambda(E) \right| dE \leq CN^{-6/7}, \tag{9.8}$$

and

$$\sup_E \left| n^\lambda(E) - n_W^\lambda(E) \right| \leq CN^{-3/7}. \tag{9.9}$$

Proof of Lemma 9.3. To prove (9.8), with Lemma 9.2, one only needs to prove

$$\int_{\frac{1}{2}\lambda_-}^{2\lambda_+} \Big| |n^\lambda(E) - n^\lambda(\lambda_-/2)| - n_W^\lambda(E) \Big| \, \mathrm{d}E \le C N^{-6/7}. \tag{9.10}$$

To this end, we first note that $|\mathbb{E}\, m_N(z)|$ is bounded uniformly for $z = E + i\eta$, such that $N^{-1+\varepsilon} \le \eta \le E/2$ and $\frac{1}{2}\lambda_- \le E \le 10$ (see (8.45)). Moreover, by (8.56) and (8.44), the conditions of Lemma B.1 in [16] are satisfied and thus we obtain (9.10). Following the proof of Lemma B.1 in [16], we see that this lemma is still valid if $(\log N)^4$ in (B.2), (B.3) and (B.4) is replaced with N^ε for small enough $\varepsilon > 0$.

To prove (9.9), for fixed E, we can assume $n^\lambda(E) > n_W^\lambda(E)$ and denote $\Delta = n^\lambda(E) - n_W^\lambda(E)$. Because $n^\lambda(E)$ is an increasing function and the derivative of $n_W^\lambda(E)$ is bounded by $\|\rho_W\|_\infty = \frac{1}{\pi}(d-d^2)^{-1/2}$, we have:

$$n^\lambda(E') - n_W^\lambda(E') \ge \Delta - C(E'-E) > 0, \text{ when } E \le E' \le E + C^{-1}\Delta. \tag{9.11}$$

Integrating both sides, we obtain

$$\int_E^{E+C^{-1}\Delta} |n^\lambda(E') - n_W^\lambda(E')| \mathrm{d}E' \ge O(\Delta^2). \tag{9.12}$$

Using (9.8), it follows that $\Delta \le O(N^{-3/7})$. □

Similarly to the calculation in Theorem 3.1 of [17], by using the logarithmic Sobolev inequality, we have

Lemma 9.4 *For $j, K \in \mathbb{N}$, such that, $j + K \le N + 1$, let $\lambda_{j,K} = K^{-1}\sum_{i=0}^{K-1} \lambda_{j+i}$, then for any $\delta > 0$*

$$\mathbb{P}\left(|\lambda_{j,K} - \mathbb{E}(\lambda_{j,K})| \ge N^{-1/2+\delta} K^{-1/2} \right) \le C e^{-N^{\delta/2}}, \tag{9.13}$$

with C depending on δ.

□

We will say that an event A, depending on N, occurs with an *extremely high probability* if $\mathbb{P}(A) \ge 1 - N^{-C}$ for any C and sufficiently large N.

Lemma 9.5 *Recall that α_j is defined as $\mathbb{E}(\lambda_j)$. Suppose that there exist sufficiently small positive numbers ε_1, ε_2 and ε_3, such that,*

$$\lambda_- + N^{-2\varepsilon_2} \le \lambda_{j_-} \le \lambda_- + N^{-\varepsilon_2}, \; \lambda_+ - N^{-2\varepsilon_2} \ge \lambda_{j_+} \ge \lambda_+ - N^{-\varepsilon_2} \tag{9.14}$$

and that

$$|\lambda_j - \alpha_j| \le N^{-\frac{1}{2}-\varepsilon_3} \text{ for } j_- < j < j_+, \tag{9.15}$$

hold with an extremely high probability, where we introduced the notations $j_- \equiv N^{1-\varepsilon_1}$ *and* $j_+ \equiv N - N^{1-\varepsilon_1}$*. Then for some* $\varepsilon > 0$*, we have*

$$\frac{1}{N}\sum_j |\alpha_j - \gamma_j|^2 \le N^{-1-\varepsilon}. \tag{9.16}$$

Proof. By symmetry, we only need to prove that (9.16) holds for the sum on the indices j with $\gamma_j \le \alpha_j$. Introduce the notation

$$n^\alpha(E) := \frac{1}{N}\#[\{\alpha_j \le E\}]. \tag{9.17}$$

The estimate (9.13), with $K = 1$, implies that $\max_j |\lambda_j - \alpha_j| \le N^{-1/2+\delta}$ holds with an extremely high probability, for any positive δ. Therefore, we can bound $n^\alpha(E)$ from above by (for any E)

$$n^\lambda(E + N^{-1/2+\delta}) = \frac{1}{N}\mathbb{E}[\#\{\lambda_j \le E + N^{-1/2+\delta}\}] \ge \frac{1}{N}\#[\{\alpha_j \le E\}] - CN^{-100} = n^\alpha(E) - CN^{-100}. \tag{9.18}$$

Similarly, we can obtain the lower bound. Putting them together, we have that:

$$CN^{-100} + n^\lambda(E + N^{-1/2+\delta}) \ge n^\alpha(E) \ge n^\lambda(E - N^{-1/2+\delta}) - CN^{-100} \tag{9.19}$$

holds for any E. The assumption (9.14) implies that

$$\lambda_j \notin [\lambda_- + N^{-\varepsilon_2}, \lambda_+ - N^{-\varepsilon_2}] \tag{9.20}$$

holds with an extremely high probability, for any $j \le j_-$ or $j \ge j_+$. For the other j's, for which λ_j may appear in $[\lambda_- + N^{-\varepsilon_2}, \lambda_+ - N^{-\varepsilon_2}]$, we use (9.15) and obtain the following improved bound on $n^\alpha(E)$: when $\lambda_- + N^{-\varepsilon_2} \le E \le \lambda_+ - N^{-\varepsilon_2}$,

$$CN^{-100} + n^\lambda(E + N^{-1/2-\varepsilon_3}) \ge n^\alpha(E) \ge n^\lambda(E - N^{-1/2-\varepsilon_3}) - CN^{-100}. \tag{9.21}$$

Let $F(E)$ be a continuous and differentiable function, such that $N^{-1/2-\varepsilon_3} \le F(E) \le N^{-1/2+\delta}$, for $0 \le \delta \le \frac{1}{10}\min\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$, $F(E) = N^{-1/2-\varepsilon_3}$ for $\lambda_- + 2N^{-\varepsilon_2} \le E \le \lambda_+ - 2N^{-\varepsilon_2}$, $F(E) = N^{-1/2+\delta}$ for $E \le \lambda_- + N^{-\varepsilon_2}$ or $E \ge \lambda_+ - N^{-\varepsilon_2}$ and $|F'(E)| \le N^{-\delta}$. Combining (9.21) and (9.19), we obtain

$$CN^{-100} + n^\lambda(E + F(E)) \ge n^\alpha(E) \ge n^\lambda(E - F(E)) - CN^{-100}. \tag{9.22}$$

On the other hand, we have

$$\alpha_j - \gamma_j = \int_{\mathbb{R}} \mathbf{1}\left(n_W(E) \ge \frac{j}{N} > n^\alpha(E)\right) \mathrm{d}E, \tag{9.23}$$

for any j, such that, $\alpha_j > \gamma_j$. Therefore we can write

$$\begin{aligned} &\frac{1}{N}\sum_{j:\gamma_j\le\alpha_j}|\alpha_j-\gamma_j|^2 \qquad (9.24)\\ &\quad=\frac{2}{N}\sum_j\int\int_{E'\le E}\mathbf{1}\left(n_W(E)\ge\frac{j}{N}>n^\alpha(E)\right)\mathbf{1}\left(n_W(E')\ge\frac{j}{N}>n^\alpha(E')\right)\mathrm{d}E\mathrm{d}E'\\ &\quad=\frac{2}{N}\sum_j\int\int_{E'\le E}\mathbf{1}\left(n_W(E)\ge\frac{j}{N}>n^\alpha(E)+CN^{-100}\right)\mathbf{1}\left(n_W(E')\ge\frac{j}{N}>n^\alpha(E')\right)\mathrm{d}E\mathrm{d}E', \end{aligned}$$

where in the second line we used the fact that the difference between j/N and $n^\alpha(E)$ must be a multiple of N^{-1}. Since $\max_j\lambda_j\le 10$ holds with an extremely high probability, using (9.22), we can replace $n^\alpha(E)$ with $n^\lambda(E-F(E))$ in (9.24), i.e.,

$$\begin{aligned} &\frac{1}{N}\sum_{j:\gamma_j\le\alpha_j}|\alpha_j-\gamma_j|^2-N^{-10} \qquad (9.25)\\ &\qquad\le\frac{2}{N}\sum_j\int\int_{E'\le E}\mathbf{1}\left(n_W(E)\ge\frac{j}{N}>n^\lambda(E-F(E))\right)\mathbf{1}\left(n_W(E')\ge\frac{j}{N}>n^\alpha(E')\right)\mathrm{d}E\mathrm{d}E'. \end{aligned}$$

Change the variable from E to $t(E)=E-F(E)$. With $|F'|\le N^{-\delta}$, we obtain $F(t)=(1+O(N^{-\delta}))F(E)$ and $\mathrm{d}t/\mathrm{d}E=(1+O(N^{-\delta}))$. Thus

$$\begin{aligned} &\frac{1}{N}\sum_{j:\gamma_j\le\alpha_j}|\alpha_j-\gamma_j|^2-N^{-10} \qquad (9.26)\\ &\qquad\le\frac{3}{N}\sum_j\int\int_{E'\le E(t)}\mathbf{1}\left(n_W(t+2F(t))\ge\frac{j}{N}>n^\lambda(t)\right)\mathbf{1}\left(n_W(E')\ge\frac{j}{N}>n^\alpha(E')\right)\mathrm{d}t\mathrm{d}E', \end{aligned}$$

where $E(t)$ is the inverse function of $t(E)$. Note, when $\mathbf{1}(\cdots)\mathbf{1}(\cdots)=1$ in (9.26), we have $n_W(E')\ge j/N>n_\lambda(t)$. Define the inverse function of n_W as $n_W^{-1}(1)=\lambda_+$, $n_W^{-1}(0)=\lambda_-$ and $n_W^{-1}(n_W(x))=x$ for $0<n_W(x)<1$. Then

$$t+2F(t)\ge E\ge E'\ge n_W^{-1}(n^\lambda(t)). \qquad (9.27)$$

Inserting this inequality into (9.26) and performing the dE' integration, we can see that

$$\begin{aligned} \frac{1}{N}\sum_{j:\gamma_j\le\alpha_j}|\alpha_j-\gamma_j|^2-\frac{1}{N^5}&\le\frac{C}{N}\sum_j\int\mathbf{1}\left(n_W(t+2F(t))\ge\frac{j}{N}>n^\lambda(t)\right)|t-n_W^{-1}(n^\lambda(t))+2F(t)|\mathrm{d}t\\ &\le C\int\left(|n_W(t+2F(t))-n^\lambda(t)|+N^{-1}\right)\cdot|t-n_W^{-1}(n^\lambda(t))+2F(t)|\mathrm{d}t\\ &\le C\left(A_1+A_2+A_3+A_4\right), \qquad (9.28) \end{aligned}$$

where we expanded (9.28) into four terms:

$$A_1 = \int |n_W(t+2F(t)) - n_W(t)| F(t) \mathrm{d}t$$
$$A_2 = \int (|n_W(t) - n^\lambda(t)| + N^{-1}) F(t) \mathrm{d}t$$
$$A_3 = \int |n_W(t+2F(t)) - n_W(t)| \cdot |t - n_W^{-1}(n^\lambda(t))| \mathrm{d}t$$
$$A_4 = \int (|n_W(t) - n^\lambda(t)| + N^{-1}) \cdot |t - n_W^{-1}(n^\lambda(t))| \mathrm{d}t.$$

Since $n_W'(t) = \rho_W(t) \le C$, $F(t) = N^{-1/2-\varepsilon_3}$ when $\lambda_- + 2N^{-\varepsilon_2} \le t \le \lambda_+ - 2N^{-\varepsilon_2}$ and $F(t) \le N^{-1/2+\delta}$ for any E, we obtain $A_1 \le N^{-1-\varepsilon}$, for some $\varepsilon > 0$. Next, from (9.8) and $F(t) \le N^{-1/2+\delta}$ for any t, we can see $A_2 \le (N^{-6/7} + N^{-1}) N^{-1/2+\delta} \le N^{-1-\varepsilon}$.

To prove $A_3 \le N^{-1-\varepsilon}$, we start with writing A_3 as

$$A_3 = \left[\int_{\lambda_+ < t} + \int_{t < \lambda_-} + \int_{\lambda_- \le t \le \lambda_- + E_1} + \int_{\lambda_+ - E_1 \le t \le \lambda_+} + \int_{\lambda_- + E_1 \le t \le \lambda_+ - E_1} \right] \Xi(t) \mathrm{d}t, \tag{9.29}$$

where we set $E_1 \equiv N^{-1/4}$ and

$$\Xi(t) \equiv |n_W(t+2F(t)) - n_W(t)| \cdot |t - n_W^{-1}(n^\lambda(t))|.$$

The first term on the r.h.s. of (9.29) is equal to zero, since n_W is constant outside $[\lambda_-, \lambda_+]$. The second term can be bounded by $N^{-1-\varepsilon}$, for some $\varepsilon > 0$, using the facts $F(t) \le N^{-1/2+\delta}$ and $n_W(\lambda_- + E) \le CE^{3/2}$, i.e.,

$$\int_{t<\lambda_-} \Xi(t) \mathrm{d}t \le C \int_{\lambda_- - N^{-1/2+\delta}}^{\lambda_-} |F(t)|^{3/2} \mathrm{d}t \le N^{-1-\varepsilon}. \tag{9.30}$$

Now we prove that the third and fourth term of (9.29) are less $N^{-1-\varepsilon}$, for some $\varepsilon > 0$.

From the explicit definition of n_W, an easy calculation shows that, for all $t \in (\lambda_-, \lambda_+)$,

$$|t - n_W^{-1}(s)| \le C |n_W(t) - s|^{2/3} \tag{9.31}$$

which in particular implies that

$$\max_{t \in (\lambda_-, \lambda_+)} \left(\max_{|s - n_W(t)| \le CN^{-3/7}} |t - n_W^{-1}(s)| \right) \le CN^{-2/7}. \tag{9.32}$$

Combining this with the fact $|n_W(t+2F(t)) - n_W(t)| \le C \|n_W'\|_\infty F(t) \le CN^{-1/2+\delta}$, we obtain that the third and fourth terms of (9.29) are less than $CN^{-2/7} N^{-1/2+\delta} N^{-1/4} \le N^{-1-\varepsilon}$, for some $\varepsilon > 0$.

To bound the last term of (9.29), we use, once again the bound $|n_W(t+2F(t)) - n_W(t)| \leq CN^{-1/2+\delta}$. From (9.31), we find therefore that

$$\begin{aligned}\int_{\lambda_- + E_1 \leq t \leq \lambda_+ - E_1} \Xi(t)\mathrm{d}t &\leq CN^{-1/2+\delta} \int |n_W(t) - n^\lambda(t)|^{2/3}\mathrm{d}t \\ &\leq CN^{-1/2+\delta}\left(\int |n_W(t) - n^\lambda(t)|\mathrm{d}t\right)^{2/3} \leq CN^{-1-\varepsilon}\end{aligned}$$

At last, we prove $A_4 \leq N^{-1-\varepsilon}$. We rewrite A_4 as

$$A_4 = \int_{t\notin(\lambda_-,\lambda_+)} \Sigma(t)\mathrm{d}t + \int_{t\in(\lambda_-,\lambda_+)} \Sigma(t)\mathrm{d}t, \tag{9.33}$$

where $\Sigma(t) \equiv (|n_W(t) - n^\lambda(t)| + N^{-1}) \cdot |t - n_W^{-1}(n^\lambda(t))|$. When $t \notin (\lambda_-, \lambda_+)$, from (9.9) and (9.32), one can see that

$$|t - n_W^{-1}(n^\lambda(t))| \leq \max_{|s - n_W(t)| \leq CN^{-3/7}} |t - n_W^{-1}(s)| \leq C|t-\lambda_-||t-\lambda_+| + CN^{-2/7}. \tag{9.34}$$

So we have

$$\int_{t\notin(\lambda_-,\lambda_+)} \Sigma(t)\mathrm{d}t = C\int_{t\notin(\lambda_-,\lambda_+)} \mathrm{d}t \left(\left|n^\lambda(t) - n_W(t)\right| + N^{-1}\right)\left(|t-\lambda_-||t-\lambda_+| + N^{-2/7}\right). \tag{9.35}$$

Using Lemma 9.2, we have

$$\begin{aligned}(9.35) \leq C&\int_{\lambda_- - N^{-1/5}}^{\lambda_-} \mathrm{d}t \left(\left|n^\lambda(t) - n_W(t)\right| + N^{-1}\right)\left(|t-\lambda_-||t-\lambda_+| + N^{-2/7}\right) \\ &+ \int_{\lambda_+}^{\lambda_+ + N^{-1/5}} \mathrm{d}t \left(\left|n^\lambda(t) - n_W(t)\right| + N^{-1}\right)\left(|t-\lambda_-||t-\lambda_+| + N^{-2/7}\right) + N^{-10}.\end{aligned} \tag{9.36}$$

Here we also used the fact that for large t, $|n^\lambda(t) - 1|$ decays exponentially fast to zero (see, for example, Lemma 7.3 of [17], which is stated for matrices with complex entries, but can be trivially extended to the case of real entries). Together with (9.8), we obtain that

$$(9.35) \leq C(N^{-6/7} + N^{-1})\left(N^{-1/5} + N^{-2/7}\right) + N^{-10} \leq N^{-1-\varepsilon}, \tag{9.37}$$

for some $\varepsilon > 0$.

When $t \in (\lambda_-, \lambda_+)$, with (9.32) and (9.8), we can see

$$\int_{t\in(\lambda_-,\lambda_+)} \Sigma(t)\mathrm{d}t \leq CN^{-6/7}N^{-2/7}. \tag{9.38}$$

Combining (9.35) and (9.38), we obtain $A_4 \leq N^{-1-\varepsilon}$ for some $\varepsilon > 0$. Together with (9.28), this compeletes the proof of Lemma 9.5. □

Next, we show that the assumptions (9.14) and (9.15) in Lemma 9.5 always hold. First we prove (9.14) in the next Lemma 9.6 with an analogous proof as Lemma 9.5. Then in Lemma 9.7 we show that (9.15) holds when (9.14) holds.

Lemma 9.6 *There exist small positive numbers ε_1 and ε_2, such that,*

$$\lambda_- + N^{-2\varepsilon_2} \leq \lambda_{j_-} \leq \lambda_- + N^{-\varepsilon_2}, \ \lambda_+ - N^{-2\varepsilon_2} \geq \lambda_{j_+} \geq \lambda_+ - N^{-\varepsilon_2} \tag{9.39}$$

hold with an extremely high probability, where we recall the notations $j_- \equiv N^{1-\varepsilon_1}$ and $j_+ \equiv N - N^{1-\varepsilon_1}$.

Proof. As in (9.19), for any E, $\delta > 0$ and sufficiently large N, we have

$$CN^{-100} + n^\lambda(E + N^{-1/2+\delta}) \geq n^\alpha(E) \geq n^\lambda(E - N^{-1/2+\delta}) - CN^{-100}. \tag{9.40}$$

So without any other assumptions, one can obtain (9.28), if we set $F(E) \equiv N^{-1/2+\delta}$ instead of $F(E)$ defined in the proof of Lemma 9.5. With a similar argument as in the proof of Lemma 9.5 but with this redefined $F(E)$, we have

$$\frac{1}{N}\sum_i |\alpha_i - \gamma_i|^2 \leq CN^{-1+C\delta}. \tag{9.41}$$

Then we claim that (9.41) implies that

$$\sup_j |\alpha_j - \gamma_j| \leq N^{-\frac{1}{10}}. \tag{9.42}$$

We prove this claim by contradiction; assume that for some j_0 we have $|\alpha_{j_0} - \gamma_{j_0}| \geq N^{-\frac{1}{10}}$. By symmetry we can assume that $j_0 \leq N/2$, the case $j_0 \geq N/2$ is analogous. We start with the case $j_0 \leq N^{1/2}$. Then $\gamma_{j_0} \leq \lambda_- + CN^{-1/4}$ and in this case α_{j_0} must be larger than γ_{j_0}, otherwise $\alpha_{j_0} \leq \gamma_{j_0} - N^{-\frac{1}{10}} \leq \lambda_- - \frac{1}{2}N^{-\frac{1}{10}}$ would contradict to $\alpha_{j_0} \in [\lambda_- - CN^{-1/5}, \lambda_+ + CN^{-1/5}]$, see (9.6). Using

$$|\gamma_i - \gamma_j| \leq CN^{-2/3}|i-j| \tag{9.43}$$

for any i, j and that α_j is monotone, we obtain that

$$\alpha_j - \gamma_j \geq \alpha_{j_0} - \gamma_{j_0} - CN^{-1/6} \geq \frac{1}{2}N^{-\frac{1}{10}}$$

for any j such that $j_0 \leq j \leq j_0 + N^{1/2}$. Then

$$\sum_{j=j_0}^{j_0+N^{1/2}} |\alpha_j - \gamma_j|^2 \geq cN^{\frac{1}{2}-\frac{1}{5}} \tag{9.44}$$

with some positive $c > 0$ which would contradict to (9.41). Now we consider the case $j_0 \geq N^{1/2}$. The previous argument remains unchanged if $\alpha_{j_0} > \gamma_{j_0}$. If $\alpha_{j_0} < \gamma_{j_0}$, then we use

$$\alpha_j - \gamma_j \leq \alpha_{j_0} - \gamma_{j_0} + CN^{-1/6} \leq -\frac{1}{2}N^{-\frac{1}{10}}$$

for any j such that $j_0 - N^{1/2} \leq j \leq j_0$ and we obtain

$$\sum_{j=j_0-N^{1/2}}^{j_0} |\alpha_j - \gamma_j|^2 \geq cN^{\frac{1}{2}-\frac{1}{5}},$$

which again contradicts to (9.41). This completes the proof of (9.42).

On the other hand, the estimate (9.13), with $K = 1$, implies $\max_j |\lambda_j - \alpha_j| \leq N^{-1/2+\delta}$ holds with an extremely high probability. Combining (9.42) with this fact, we can see that for any small enough ε_1, there exists ε_2 such that (9.39) holds, which completes the proof of Lemma 9.6.

□

The next Lemma guarantees the assumption (9.15) in Lemma 9.5, given (9.14).

Lemma 9.7 *If there exist sufficiently small positive numbers ε_1 and ε_2, such that*

$$\lambda_- + N^{-2\varepsilon_2} \leq \lambda_{j_-} \leq \lambda_- + N^{-\varepsilon_2}, \ \lambda_+ - N^{-2\varepsilon_2} \geq \lambda_{j_+} \geq \lambda_+ - N^{-\varepsilon_2}, \tag{9.45}$$

holds with an extremely high probability, then there exists $\varepsilon_3 > 0$ such that,

$$|\lambda_j - \alpha_j| \leq N^{-\frac{1}{2}-\varepsilon_3}, \text{ for } j_- < j < j_+, \tag{9.46}$$

holds with an extremely high probability, where we recall the notations $j_- \equiv N^{1-\varepsilon_1}$ and $j_+ \equiv N - N^{1-\varepsilon_1}$.

Proof. For simplicity, we only prove the case of $j \leq N/2$, the case $j > N/2$ is analogous. Using (9.13), for any $N/2 \geq j > j_-$, $\delta > 0$, with $K = N^{1/4}$, we have

$$\mathbb{P}(|\lambda_{j,K} - \mathbb{E}(\lambda_{j,K})| \geq N^{-5/8+\delta}) \leq Ce^{-N^\delta}. \tag{9.47}$$

Now we claim that, for $K = N^{1/4}$, $j_- < j < j_+$,

$$|\lambda_{j,K} - \lambda_j| \leq N^{-5/8} \tag{9.48}$$

holds with an extremely high probability, which implies

$$|\mathbb{E}\lambda_{j,K} - \mathbb{E}\lambda_j| \leq CN^{-5/8}. \tag{9.49}$$

To see (9.48), first notice that

$$\mathbb{P}(|\lambda_{j,K} - \lambda_j| \geq N^{-5/8}) \leq \mathbb{P}(\lambda_{j+K} - \lambda_j \geq N^{-5/8}). \tag{9.50}$$

Suppose now that $\lambda_{j+K} - \lambda_j \geq N^{-5/8}$. With the assumption (9.45), we have that, for $j_- < j < N/2$,

$$\lambda_j \in \left(\lambda_- + N^{-2\varepsilon_2}, \lambda_+ - N^{-2\varepsilon_2}\right). \tag{9.51}$$

with an extremely high probability. Divide this interval into small intervals with the length $\frac{1}{2}N^{-5/8}$. By the local Marchenko-Pastur law, i.e., Corollary 8.2, the event that the number of the eigenvalues in each piece is larger than $CN^{1-3\varepsilon_2-5/8}$ holds with an extremely high probability. On the other hand, if $\lambda_{j+K} - \lambda_j \geq N^{-5/8}$, then the total number of eigenvalues in at least one of these intervals is less than $K = N^{1/4}$, which implies that $\lambda_{j+K} - \lambda_j \leq N^{-5/8}$ holds with an extremely high probability. Together with (9.50), we have (9.48). Then combining (9.48), (9.49) and (9.47), we obtain (9.46) and complete the proof. □

Now we are ready to prove Theorem 9.1.

Proof of Theorem 9.1. Note that the assumptions in Lemma 9.5 are proved in Lemma 9.6 and 9.7. Combining Lemma 9.5, 9.6 and 9.7, we obtain (9.16), i.e.,

$$\frac{1}{N}\sum_j |\alpha_j - \gamma_j|^2 \leq N^{-1-\varepsilon}, \tag{9.52}$$

for some constant $\varepsilon > 0$, where α_j is defined as $\mathbb{E}\lambda_j$. Then we claim that for some constant $\varepsilon > 0$,

$$\frac{1}{N}\sum_j |\lambda_j - \alpha_j|^2 \leq N^{-1-\varepsilon} \tag{9.53}$$

holds with an extremely high probability. To see (9.53), first notice that (9.13), with $K = 1$, implies that, for any δ and j

$$|\lambda_j - \alpha_j| \leq N^{-1/2+\delta} \tag{9.54}$$

holds with an extremely high probability. The estimate (9.46) shows that there exist $\varepsilon_1 > 0$ and $\varepsilon_3 > 0$ such that

$$|\lambda_j - \alpha_j| \leq N^{-\frac{1}{2}-\varepsilon_3}, \text{ for } N^{\varepsilon_1} < j < N - N^{\varepsilon_1} \tag{9.55}$$

holds with an extremely high probability. Combining this with (9.54) for the remaining indices $j \leq N^{\varepsilon_1}$ or $j \geq N - N^{\varepsilon_1}$, we obtain (9.53). Together with (9.52), we have:

$$\frac{1}{N}\mathbb{E}\sum_j |\lambda_j - \gamma_j|^2 \leq N^{-1-\varepsilon}, \tag{9.56}$$

for some $\varepsilon > 0$. Using the definition $x_j = \lambda_j^{1/2}$, one has

$$|x_j - \gamma_j^{1/2}| = |\lambda_j - \gamma_j|(x_j + \gamma_j^{1/2})^{-1} \leq C|\lambda_j - \gamma_j|. \tag{9.57}$$

Inserting (9.57) into (9.56), we obtain (9.2) and complete the proof of Theorem 9.1. □

A Existence and restriction of the dynamics

As in Section 2, we consider the Euclidean space $\mathbb{R}^N$ with the normalized measure $\mu = \exp(-N\mathcal{H})/Z$. The Hamiltonian $\mathcal{H}$ is of the form (2.6) or (2.8), for definiteness we discuss the first case, the second case is fully analogous. $\mathcal{H}$ is symmetric with respect to the permutation of the variables $\mathbf{x} = (x_1, \ldots, x_N)$, thus the measure can be restricted to the subset $\Sigma_N \subset \mathbb{R}^N$ defined in (2.4). In this appendix we outline how to define the dynamics (2.1) with its generator, formally given by $L = \frac{1}{2N}\Delta - \frac{1}{2}(\nabla\mathcal{H})\nabla$, on Σ_N. The condition $\beta \geq 1$ and the specific factors $\prod_{i<j} |x_j - x_i|^\beta$ will play a key role in the argument, in particular, we will see that $\beta = 1$ is the critical threshold for this method to work.

We first recall the standard definition of the dynamics on $\mathbb{R}^N$. The quadratic form

$$\mathcal{E}(u, v) := \int_{\mathbb{R}^N} \nabla u \cdot \nabla v \, \mathrm{d}\mu$$

is a closable Markovian symmetric form on $L^2(\mathbb{R}^N, \mathrm{d}\mu)$ with a domain $C_0^\infty(\mathbb{R}^N)$ (see Example 1.2.1 and Theorem 3.1.3 of [24]). This form can be closed with a form domain $H^1(\mathbb{R}^N, \mathrm{d}\mu)$ defined as the closure of C_0^∞ in the norm $\|\cdot\|_+^2 = \mathcal{E}(\cdot,\cdot) + \|\cdot\|_2^2$. The closure is called the Dirichlet form. It generates a strongly continuous Markovian semigroup T_t, $t > 0$, on L^2 (Theorem 1.4.1 [24]) and it can be extended to a contraction semigroup to $L^1(\mathbb{R}^N, \mathrm{d}\mu)$, $\|T_t f\|_1 \leq \|f\|_1$ (Section 1.5 [24]). The generator L of the semigroup, is defined via the Friedrichs extension (Theorem 1.3.1 [24]) and it is a positive self-adjoint operator on its natural domain $D(L)$ with C_0^∞ being the core. The generator is given by $L = \frac{1}{2N}\Delta - \frac{1}{2}(\nabla\mathcal{H})\nabla$ on its domain (Corollary 1.3.1 [24]). By the spectral theorem, T_t maps L^2 into $D(L)$, thus with the notation $f_t = T_t f$ for some $f \in L^2$, it holds that

$$\partial_t f_t = L f_t, \qquad t > 0, \quad \text{and} \quad \lim_{t\to 0+0} \|f_t - f\|_2 = 0.$$

Moreover, by approximating f by L^2 functions and using that T_t is contraction in L^1 (Section 1.5 in [24]), the differential equation holds even if the initial condition f is only in L^1. In this case the convergence $f_t \to f$, as $t \to 0+0$, holds only in L^1. We remark that T_t is also a contraction on L^∞, by duality.

Now we restrict the dynamics to $\Sigma = \Sigma_N$. Repeating the general construction with $\mathbb{R}^N$ replaced by Σ_N, we obtain the corresponding generator $L^{(\Sigma)}$ and the semigroup $T_t^{(\Sigma)}$.

To establish the relation between L and $L^{(\Sigma)}$, we first define the symmetrized version of Σ

$$\widetilde{\Sigma} := \mathbb{R}^N \setminus \Big\{ \mathbf{x} \,:\, \exists i \neq j \quad \text{with} \quad x_i = x_j \Big\}.$$

Denote $X := C_0^\infty(\widetilde{\Sigma})$. The key information is that X is dense in $H^1(\mathbb{R}^N, \mathrm{d}\mu)$ which is equivalent to the density of X in $C_0^\infty(\mathbb{R}^N, \mathrm{d}\mu)$. We will check this property below. Then the general argument above directly applies if $\mathbb{R}^N$ is replaced by $\widetilde{\Sigma}_N$ and it shows that the generator L is the the same (with the same domain) if we start from X instead of $C_0^\infty(\mathbb{R}^N, \mathrm{d}\mu)$ as a core.

Note that both L with $L^{(\Sigma)}$ are local operators and L is symmetric with respect to the permutation of the variables. For any function f defined on Σ, we define its symmetric extension onto $\widetilde{\Sigma}$ by $\widetilde{f}$. Clearly $L\widetilde{f} = \widetilde{L^{(\Sigma)} f}$ for any $f \in C_0^\infty(\Sigma)$. Since the generator is uniquely determined by its action on its core, and the generator uniquely determines the dynamics, we see that for any $f \in L^1(\Sigma, \mathrm{d}\mu)$, one can determine $T_t^{(\Sigma)} f$ by computing $T_t\widetilde{f}$ and restricting it to Σ. In other words, the dynamics (2.1) is well defined when restricted to $\Sigma = \Sigma_N$.

Finally, we have to prove the density of X in $C_0^\infty(\mathbb{R}^N, \mathrm{d}\mu)$, i.e., to show that if $f \in C_0^\infty(\mathbb{R}^N)$, then there exists a sequence $f_n \in C_0^\infty(\widetilde{\Sigma})$ such that $\mathcal{E}(f - f_n, f - f_n) \to 0$. The structure of $\widetilde{\Sigma}$ is complicated since in addition to the one codimensional coalescence hyperplanes $x_i = x_j$ (and $x_i = 0$ in case of Σ^+), it contains higher order coalescence subspaces with higher codimensions. We will show the approximation argument in a neighborhood of a point $\mathbf{x}$ such that $x_i = x_j$ but $x_i \neq x_k$ for any other $k \neq i, j$. The proof uses the fact that the measure $\mathrm{d}\mu$ vanishes at least to first order, i.e., at least as $|x_i - x_j|$, around $\mathbf{x}$, thanks to $\beta \geq 1$. This is the critical case; the argument near higher order coalescence points is even easier, since they have lower codimension and the measure μ vanishes at even higher order.

In a neighborhood of $\mathbf{x}$ we can change to local coordinates such that $r := x_i - x_j$ remains the only relevant coordinate. Thus the task is equivalent to show that any $g \in C_0^\infty(\mathbb{R})$ can be approximated by a sequence $g_\varepsilon \in C_0^\infty(\mathbb{R} \setminus \{0\})$ in the sense that

$$\int_{\mathbb{R}} |g'(r) - g'_\varepsilon(r)|^2 |r| \mathrm{d}r \to 0 \tag{A.1}$$

as $\varepsilon \to 0$. It is sufficient to consider only the positive semi-axis, i.e., $r > 0$. Extending the functions to two dimensional radial functions, $G(x) := g(|x|)$, $G_\varepsilon(x) = g_\varepsilon(|x|)$, this statement is equivalent to the fact that a point in two dimensions has zero capacity.

B Bakry-Emery argument on a subdomain

The estimate (4.14) in Theorem 4.2 is based on the Bakry-Emery argument [2] for the dissipation of the Dirichlet form. This method uses a lower bound on the Hessian of $\widetilde{\mathcal{H}}$ and an integration by parts. Since the dynamics is restricted to $\Sigma = \Sigma_N$, we need to check that the boundary term in the integration by parts vanishes.

In our application, this argument will be used for the Hamiltonian $\widetilde{\mathcal{H}}$ (see (4.5)) and its generator $\widetilde{L} = \frac{1}{2N}\Delta - \frac{1}{2}(\nabla\widetilde{\mathcal{H}})\nabla$, but for simplicity, we omit the tilde from the notation below. With $h = h_t =$

$\sqrt{q_t}$ a standard calculation (see (5.8) of [16] with somewhat different notations) shows that

$$\begin{aligned}
\partial_t \frac{1}{2N}\int_\Sigma (\nabla h)^2 e^{-N\mathcal{H}} \mathrm{d}\mathbf{x} &= \frac{1}{N}\int_\Sigma \nabla h \nabla\Big(Lh + \frac{1}{2N} h^{-1}(\nabla h)^2\Big) e^{-N\mathcal{H}} \mathrm{d}\mathbf{x} \\
&= \frac{1}{N}\int_\Sigma \Big[\nabla h L \nabla h - \frac{1}{2}\nabla h(\nabla^2 \widetilde{\mathcal{H}})\nabla h + \frac{1}{2N}(\nabla h)\nabla[h^{-1}(\nabla h)^2]\Big] e^{-N\mathcal{H}} \mathrm{d}\mathbf{x} \\
&= \frac{1}{2N}\int_\Sigma \Big[-\nabla h(\nabla^2\mathcal{H})\nabla h - \sum_{i,j}\big(\partial^2_{ij} h - \frac{\partial_i h \partial_j h}{h}\big)^2\Big] e^{-N\mathcal{H}} \mathrm{d}\mathbf{x} \\
&\le -\frac{1}{2N}\int_\Sigma \nabla h(\nabla^2\mathcal{H})\nabla h \, e^{-N\mathcal{H}} \mathrm{d}\mathbf{x}
\end{aligned} \tag{B.1}$$

assuming that the quantities in each step are well defined and that the boundary term

$$\int_{\partial\Sigma} \partial_i h \, \partial^2_{ij} h \, e^{-N\mathcal{H}} = 0 \tag{B.2}$$

in the integration by parts in the third line vanishes. In [16] we argued with a somewhat specific form of q, an information not directly available here.

The rigorous proof in the general case uses a regularization and a cutoff argument. First we regularize the function $q = q_t \in D(L)$, $t > 0$, by defining

$$q^\varepsilon(\mathbf{x}) := \frac{q(\mathbf{x}) + \varepsilon}{1+\varepsilon}, \qquad h^\varepsilon := \sqrt{q^\varepsilon},$$

for some $\varepsilon > 0$. This has the advantage that the derivatives of h^ε can be bounded by those of q^ε. We consider a cutoff function $\theta \in C_0^\infty(\Sigma)$ to be specified later and we insert θ in the calculation (B.1). Since L is an elliptic operator with smooth coefficients away from the boundary $\partial\Sigma$, by standard parabolic regularity we know that q and thus h are smooth functions inside Σ. Thus each step in the cutoff version of (B.1) is justified with an additional term coming from the derivative hitting θ in the integration by parts. After repeating the steps in (B.1), we obtain

$$\begin{aligned}
\partial_t \frac{1}{2N}\int_\Sigma \theta(\nabla h^\varepsilon)^2 e^{-N\mathcal{H}} \mathrm{d}\mathbf{x} &= \frac{1}{N}\int_\Sigma \theta \nabla h^\varepsilon \nabla\Big(Lh^\varepsilon + \frac{1}{2Nh^\varepsilon}(\nabla h^\varepsilon)^2\Big) e^{-N\mathcal{H}} \mathrm{d}\mathbf{x} \\
\le -\frac{1}{2N}\int_\Sigma \theta \nabla h^\varepsilon(\nabla^2\mathcal{H})\nabla h^\varepsilon \, e^{-N\mathcal{H}} \mathrm{d}\mathbf{x} &- \frac{1}{2}N\int_\Sigma \sum_{i,j}(\partial_j\theta)(\partial_i h^\varepsilon)(\partial_i\partial_j h^\varepsilon) \, e^{-N\mathcal{H}} \mathrm{d}\mathbf{x}.
\end{aligned} \tag{B.3}$$

We now show that, by an appropriate choice of a sequence of cutoff functions, the second term in (B.3) vanishes. We first define the set of higher order coalescences where at least three point coincide as

$$Q := \{\mathbf{x} \in \partial\Sigma \ : \ \exists\, i \ \text{ s.t. } \ x_i = x_{i+1} = x_{i+2}\}.$$

We remark that in case of Assumption I' we formally introduce $x_0 = 0$ to this definition, so that Q will include also three point singularities of the type $x_1 = x_2 = 0$. For any $\delta > 0$ we define the set

$$Q_\delta := \{\mathbf{x} \in \Sigma \ : \ \text{dist}(\mathbf{x}, Q) \leq \delta\}$$

is the δ-neighborhood of the three-point singularity set within Σ. Introduce an additional small positive parameter $\eta \ll \delta$. We now choose the cutoff function θ of the form $\theta = \theta_1\theta_2$, depending on the parameters δ and η, such that

(i) $\theta_1(\mathbf{x}) \equiv 1$ if $\text{dist}(\mathbf{x}, \partial\Sigma) \geq 2\eta$, $\theta_1(\mathbf{x}) \equiv 0$ if $\text{dist}(\mathbf{x}, \partial\Sigma) \leq \eta$ and $|\nabla\theta_1| \leq O(\eta^{-1})$;

(ii) $\theta_2(\mathbf{x}) \equiv 1$ if $\text{dist}(\mathbf{x}, Q) \geq 2\delta$, $\theta_2(\mathbf{x}) \equiv 0$ if $\text{dist}(\mathbf{x}, Q) \leq \delta$ and $|\nabla\theta_2| \leq O(\delta^{-1})$.

Here and in the sequel we make the convention that a quantity of order δ^k with some $k \in \mathbb{R}$ (sometimes denoted by $O(\delta^k)$) denotes a number that is comparable with δ^k with implicit constants that may depend on N. However, N is fixed in this argument, so this dependence is irrelevant. Similar convention holds for $O(\eta^k)$.

We state two estimates on the solution q_t of (4.13) that will be proven at the end of the section.

Lemma B.1 *Assume that $q_0 \in L^\infty$. Then the solution q_t of* (4.13) *satisfies a uniform supremum bound on the closure of Σ,*

$$\sup_{t \geq 0} \sup_{\mathbf{x} \in \overline{\Sigma}} q_t(\mathbf{x}) < \infty. \tag{B.4}$$

Furthermore, q_t is regular away from the higher order coalescence singularities with the estimate

$$\sup \Big\{ |\nabla^k q_t(\mathbf{x})| \ : \ \mathbf{x} \in \Sigma \cap K, \ \mathit{dist}(\mathbf{x}, Q) \geq \delta \Big\} \leq C(t, k, N, K)\delta^{-k} \tag{B.5}$$

where K is a compact set and the constant depends only on the indicated parameters. In particular, q_t is regular up to the boundary $\partial\Sigma \setminus Q_\delta$, i.e., at the two-point coalescence points away from higher order coalescences.

Using this lemma, we can treat the second term on the r.h.s. of (B.3). We split the integration into two regimes. First we consider the regime where $\theta_2\nabla\theta_1 \neq 0$, i.e., an (2η)-neighborhood of $\partial\Sigma \setminus Q_\delta$. On this set we note the local density scales at as η^β, thanks to the term $|x_i - x_j|^\beta$ in $e^{-N\mathcal{H}}$. Thus the measure of the support of $\nabla\theta$ near $\partial\Sigma \setminus Q_\delta$ scales as $\eta^{1+\beta}$, while $|\nabla\theta| \leq C\eta^{-1}$ (assuming $\eta \leq \delta$). Since (B.5) guarantees that the derivatives of h^ε remain locally bounded (with a bound depending on ε, δ, t and N), the boundary term near $\partial\Sigma \setminus Q_\delta$ vanishes as $\eta \to 0$.

To estimate the integral on the support of $\nabla\theta_2$, i.e., on a subset of $Q_{2\delta}$, we use that θ can be replaced with θ_2 after taking the $\eta \to 0$ limit. Since we have $|\nabla\theta_2| = O(\delta^{-1})$ and $|\nabla^k h_t^\varepsilon| \leq C_\varepsilon |\nabla^k q_t^\varepsilon| \leq C_{\varepsilon,t,N}\delta^{-k}$ with $k = 1, 2$, the integrand scales at most δ^{-4}. Since the local density scales at least as $\delta^{3\beta}$ due to a factor of the type $|x_i - x_{i+1}|^\beta |x_{i+1} - x_{i+2}|^\beta |x_i - x_{i+2}|^\beta$, the total measure of $Q_{2\delta}$ is of order $\delta^{2+3\beta}$. Hence the integral on $Q_{2\delta}$ scales at most as $\delta^{2+3\beta-4} \leq \delta$ in the δ parameter and therefore the contribution of the neighborhood of higher order singularities to the second term in (B.3) vanishes as $\delta \to 0$.

After having removed θ and the second term from (B.3), we let $\varepsilon \to 0$ and this gives the desired result (4.14).

To complete the argument, finally, we need to prove Lemma B.1.

Proof of Lemma B.1. The bound (B.4) follows immediately, since $q_0 \in L^\infty$ and the semigroup T_t a contraction in L^∞ (see Appendix A).

The second statement of Lemma B.1 follows from a standard regularization argument for a typical two-point singularity at $x_i = x_{i+1}$ that was already outlined in [20]. Fix a point $\mathbf{x}^* \in \partial\Sigma \setminus Q_\delta$ and assume that $x_i^* = x_{i+1}^*$, but for all other pairs $|x_j^* - x_{j+1}^*| \geq \delta$. We remark that the neighborhood of two (or more) independent singularities, e.g., $x_i = x_{i+1}$ and $x_j = x_{j+1}$, $|i-j| \geq 2$, can be treated similarly by applying the same regularization argument separately. We omit these details here.

Let B be a neighborhood of size $O(\delta)$ around $\mathbf{x}^*$. Choose a local coordinate system $\Phi(\mathbf{x}) = (u, \mathbf{y}) \in \mathbb{R}_+ \times \mathbb{R}^{N-1}$ in B such that $u = \frac{1}{2}(x_{i+1} - x_i) > 0$. Within $\Phi(B)$, we can write

$$\widetilde{L} = \frac{1}{4N}\Big[\partial_u^2 + \frac{\beta}{u}\partial_u\Big] + L_{reg},$$

where L_{reg} is an elliptic operator with second derivatives in the $\mathbf{y}$ variables and with coefficients regular on the scale δ (since all other singularities are at least at a distance $O(\delta)$ away from $\Phi(B)$).

For the $\beta = 1$ case, by introducing a function $\widehat{q}_t(a, b, \mathbf{y}) := q_t(\sqrt{a^2 + b^2}, \mathbf{y})$ of $N+1$ variables, we see that $\widehat{q}_t$ satisfies $\partial_t \widehat{q}_t = \widehat{L}\widehat{q}_t$, where

$$\widehat{L} = \frac{1}{N}\big[\partial_a^2 + \partial_b^2\big] + L_{reg},$$

i.e., L becomes an elliptic operator $\widehat{L}$ with bounded and regular coefficients in the new variables. A similar transformation is possible for any integer $\beta \geq 1$, where u is considered as the radial part of a $(\beta+1)$-dimensional variable.

We claim that the singular point $u = 0$ becomes a removable singularity in the variables (a, b) around $(0, 0)$. Note that the singular set is a codimension two subspace in the $(a, b, \mathbf{y}, t)$ space-time coordinate system which becomes a line segment in the (a, b, t) space-time system if we disregard the variable $\mathbf{y}$. Note that $\mathbf{y}$ plays no role in this argument since every coefficient is regular in $\mathbf{y}$. The parabolic equation $\partial_t \widehat{q}_t = \widehat{L}\widehat{q}_t$ holds in a strong sense away from the origin $(a, b) = (0, 0)$ in these two variables, and moreover $\widehat{q}_t$ is bounded by (B.4). We can thus apply Theorem II of [1] with $p = 2$, $r = \infty$ to see that $\widehat{q}_t$ must coincide with the regular solution obtained by using the fundamental solution to the equation in a small space-time neighborhood of the singular set. This proves that $\widehat{q}_t$, and hence q_t, is a smooth function up to the boundary $\partial\Sigma \setminus Q_\delta$.

To obtain the quantitative estimate (B.5), we consider the regularity of the coefficients of L_{reg}. Due to the special structure of $\mathcal{H}$, every term in $L = \frac{1}{2N}\Delta - \frac{1}{2}(\nabla\mathcal{H})\nabla$ is either regular on any small scales, or it scales as (length)$^{-2}$. Since the neighborhood B is at least at distance $O(\delta)$ away from the other singularities, the coefficients of L_{reg} are regular on scale δ. Therefore the solution q_t is regular on scale δ on B and this gives the δ-scaling of the estimate (B.5). This completes the proof of Lemma B.1. □

Acknowledgement. The authors thank Alice Guionnet for pointing out some errors in the preliminary version of the manuscript.

References

[1] Aronson, D.G.: Removable singularities for linear parabolic equations.*Arch. Rat. Mech. Anal.* **17** No. 1, 79-84 (1964).

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

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

[4] Ben Arous, G., Péché, S.: Private communication.

[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] Deift, P.: Orthogonal polynomials and random matrices: a Riemann-Hilbert approach. *Courant Lecture Notes in Mathematics* **3**, American Mathematical Society, Providence, RI, 1999

[8] Deift, P., Gioev, D.: Random Matrix Theory: Invariant Ensembles and Universality. *Courant Lecture Notes in Mathematics* **18**, American Mathematical Society, Providence, RI, 2009

[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] Dumitriu, I., Edelman, A.: Matrix models for beta-ensembles. *J. Math. Phys.* **43**, no. 11, 5830-5847 (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] Erdős, L., Péché, G., Ramírez, J., Schlein, B., and Yau, H.-T., Bulk universality for Wigner matrices. *Commun. Pure Appl. Math.* **63**, No. 7, 895–925 (2010)

[15] Erdős, L., Ramírez, J., Schlein, B., Tao, T., Vu, V. and Yau, H.-T.: Bulk universality for Wigner hermitian matrices with subexponential decay. *Int. Math. Res. Notices.* **2010**, No. 3, 436-479 (2010)

[16] Erdős, L., Ramirez, J., Schlein, B., Yau, H.-T.: Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. *Electr. J. Prob.* **15**, Paper 18, 526–604 (2010)

[17] 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 (2009)

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

[19] Erdős, L., Schlein, B., Yau, H.-T.: Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Notices.* **2010**, No. 3, 436-479 (2010)

[20] Erdős, L., Schlein, B., Yau, H.-T.: Universality of random matrices and local relaxation flow. Preprint. arXiv:0907.5605

[21] Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices. Preprint. arXiv:1001.3453

[22] Feldheim, O. and Sodin, S.: A universality result for the smallest eigenvalues of certain sample covariance matrices. Preprint. arXiv:0812.1961

[23] Forrester, P.: Log-gases and random matrices. Book in preparation, see http://www.ms.unimelb.edu.au/∼matpjf/matpjf.html.

[24] Fukushima, M., Oshima, Y., Takeda, M.: Dirichlet Forms and Symmetric Markov Processes. Walter de Gruyter, 1994

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

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

[27] Johansson, K.: Universality for certain hermitian Wigner matrices under weak moment conditions. Preprint arxiv.org/abs/0910.4467

[28] Ledoux, M.: The concentration of measure phenomenon. Mathematical Surveys and Monographs, **89** American Mathematical Society, Providence, RI, 2001.

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

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

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

[32] Pastur, L., Shcherbina M.: Bulk universality and related properties of Hermitian matrix models. *J. Stat. Phys.* **130** (2008), no.2., 205-250.

[33] Péché, S.: Universality results for largest eigenvalues of some sample covariance matrix ensembles. *Probab. Th. Rel. Fields* **143**, no. 3-4, 481–516 (2009)

[34] Ruzmaikina, A.: Universality of the edge distribution of eigenvalues of Wigner random matrices with polynomially decaying distributions of entries. *Comm. Math. Phys.* **261** (2006), 277–296.

[35] 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.

[36] Sodin, S.: The spectral edge of some random band matrices. Preprint arXiv:0906.4047

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

[38] Péché, S. and Soshnikov, A.: Wigner Random Matrices with Non-symmetrically Distributed Entries, *Journal of Statistical Physics*, **129**(5/6) (2007), 857–884.

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

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

[41] Tao, T. and Vu, V.: Random covariance matrices: Universality of local statistics of eigenvalues. Preprint. arXiv:0912.0966

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

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

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