Optimal Smoothed Analysis and Quantitative Universality for the Smallest Singular Value of Random Matrices

Haoyu Wang *

November 9, 2022

Abstract

The smallest singular value and condition number play important roles in numerical linear algebra and the analysis of algorithms. In numerical analysis with randomness, many previous works make Gaussian assumptions, which are not general enough to reflect the arbitrariness of the input. To overcome this drawback, we prove the first quantitative universality for the smallest singular value and condition number of random matrices.

Moreover, motivated by the study of smoothed analysis that random perturbation makes deterministic matrices well-conditioned, we consider an analog for random matrices. For a random matrix perturbed by independent Gaussian noise, we show that this matrix quickly becomes approximately Gaussian. In particular, we derive an optimal smoothed analysis for random matrices in terms of a sharp Gaussian approximation.
## Contents

1 Introduction 3
   1.1 Overview and our contributions 3
   1.2 Models and main results 4
   1.3 Outline of proofs 6
   1.4 Notations and paper organizations 7

2 Applications in Numerical Analysis and Algorithms 8
   2.1 Accuracy of least-square solution 8
   2.2 Complexity of conjugate gradient method 8

3 Smoothed Analysis and Gaussian Approximation 9
   3.1 Singular value dynamics 9
   3.2 Local relaxation at the hard edge 13
   3.3 Bootstrap arguments 15

4 Quantitative Universality 16
   4.1 Quantitative resolvent comparison 16
   4.2 Proof of Theorem 2 17

5 Condition Number 18
   5.1 Smoothed analysis 18
   5.2 Quantitative universality 19

6 Beyond Strictly Square Matrices 20

A Auxiliary Results 22

B Proofs for Smoothed Analysis 23
   B.1 Proof of Proposition 3.1 23
   B.2 Proof of Lemma 3.10 26

C Proofs for Quantitative Universality 36
1 Introduction

In numerical analysis and analysis of algorithms, the smallest singular value and condition number of matrices play important roles [TB97]. Broadly speaking, the smallest singular value reflects the invertibility of a matrix, and a better non-degeneracy condition of the matrix will result in faster algorithms or ensures better stability in computations. The condition number is a crucial quantity as well. It measures the hardness and accuracy of numerical computations, and its most well-known practical meaning is the loss of precision in solving linear equations [Sma85]. For more concrete state-of-the-art applications, the smallest singular value affect the stability in fast algorithms of solving linear systems [PV21, Nie22]. In [GPV21], the condition number determines the iteration complexity in fast algorithms for $p$-norm regression. There are a lot of other applications of these two quantities, and we do not intend to provide a complete review here.

In the seminal work of Spielman and Teng [ST04], smoothed analysis is introduced to understand why some algorithms with poor worst-case performance can work successfully in practice. Roughly speaking, smoothed analysis is an interpolation between the worst-case analysis and the average-case analysis. In the context of solving linear equations $Ax = b$, even if the matrix $A$ has large condition number (and consequently large loss of precision), a random perturbation $A + H$ will become well-conditioned with high probability. Smoothed analysis for random perturbation of deterministic matrices has been well studied since its invention [Wsc04, VH14, TV10b, FV16, SS20] etc. It has also been applied in many algorithms, for example Gaussian elimination [SST06], matrix inversion [BC10], conjugate descent method [MT16], tensor decomposition [BCMV14], the $k$-means method [AMR11], etc.

Classical results of smoothed analysis focus on random perturbation of deterministic matrices, and show that this perturbation indeed makes the original matrix better for the implementation of algorithms (in the sense such as having smaller condition number). Recently, numerical computations with random data has become more and more common. In the context of random matrices, we show that if a random matrix is perturbed by independent Gaussian noise, it quickly becomes approximately Gaussian. The smallest singular value and condition number for a Gaussian matrix has been well studied [Ede88], and therefore such Gaussian approximations enable us to analyze the more tractable Gaussian ensemble instead with benign approximation error. Specifically, we prove an optimal smoothed analysis of random matrices in terms of a sharp Gaussian approximation.

Moreover, in numerical analysis with randomness, many previous works make Gaussian assumptions. Such assumptions are not general enough to reflect the arbitrariness of the input. To overcome this drawback, we prove the first quantitative universality for the smallest singular value and condition number of random matrices. Both the smoothed analysis for random matrices and the quantitative universality will play useful roles in computations with randomness.

1.1 Overview and our contributions

From the mathematical perspective, the edge universality has been a classical problem in random matrix theory and it has tremendous progress in the past decade. However, most previous results are qualitative statements, and most quantitative rate of convergence to the limiting law were only known for integrable models. Recently, the rate of convergence to the Tracy-Widom distribution for the largest eigenvalue was first obtained in [Bou22] for generalized Wigner matrices and in [Wan19] for sample covariance matrices. These results were further improved by Schnelli and Xu in [SX22a, SX21, SX22b]. Our work is the first result about the quantitative universality for the smallest singular value and the condition number.

For an $M \times N$ matrix $H$ with $\lim_{N \to \infty} N/M \to d \in (0, 1]$, the empirical spectral distribution
of the sample covariance matrix $H^\top H$ converges to the Marchenko-Pastur law
\[
\rho_{\text{MP}}(x) = \frac{1}{2\pi d} \sqrt{(x - (1 - \sqrt{d})^2)((1 + \sqrt{d})^2 - x)} \frac{1}{x^2}, \quad x \in [(1 - \sqrt{d})^2, (1 + \sqrt{d})^2].
\]

In the case $d < 1$, the Marchenko-Pastur law has no singularity and such situations are called the soft edge case in the literature of random matrix theory. Moreover, note that all eigenvalues are of order constant with high probability. In particular, thanks to the square-root decay near the spectral edge of the Marchenko-Pastur law, the extreme eigenvalues of the covariance matrix $H^\top H$ are highly concentrated at the spectral endpoints at the scale $N^{-2/3}$ and the fluctuation is the Tracy-Widom law. In this case, the optimal rate of convergence to Tracy-Widom distribution has been established in [SX21], and the quantitative universality for the smallest singular value will be an easy consequence.

However, in the case $\lim N/M = 1$, things become much more complicated. Note that in this case the Marchenko-Pastur law has a singularity at the origin
\[
\rho_{\text{MP}}(x) = \frac{1}{2\pi} \sqrt{\frac{4 - x}{x}}, \quad x \in [0, 4].
\]

This singularity of the limiting spectral distribution makes the behaviours of the eigenvalue near the left endpoint $x = 0$ very different. In particular, the typical eigenvalue spacing near the left edge is of order $N^{-1}$, which is much smaller than the soft edge case where the typical spacing is of order $N^{-2/3}$. In random matrix theory, this case is called the hard edge case. In this model, the left edge of the spectrum with singularity is called the hard edge and the right edge is called the soft edge. The study of the spectral statistics near the hard edge is a notoriously tricky problem. Due to some technical difficulties, the study of the hard edge case is mostly restricted to square matrices with $M \equiv N$. Because of this difficulty, we will also focus on the study of square $N \times N$ random matrices.

For the study of the smallest singular value of an $N \times N$ random matrix, the first result was obtained by Edelman for the Gaussian matrix in [Ede88]. Later, the distribution in the Gaussian ensemble was shown to be universal by Tao and Vu in [TV10a] and further generalized to sparse matrices by Che and Lopatto in [CL19]. However, both of their universality results are qualitative statements with unknown error terms.

In computer science and numerical computations, the accuracy or complexity of algorithms depend on the smallest singular value or condition number in a delicate way. Therefore, a qualitative universality is not enough to measure the performance of general random input. Thus, quantitative estimates are necessary for practical purposes. In this work, we prove the first quantitative version of the universality for the smallest singular value and the condition number. This solves a long-standing open problem\(^1\). Along the way to prove quantitative universality, we obtain a sharp estimate for matrices with Gaussian perturbations, and hence establish the optimal smoothed analysis for random matrices.

1.2 Models and main results

Thanks to the motivations from theoretical computer science, we mainly focus on real matrices. However, as mentioned in [CL19], our whole proof works for complex matrices as well.

\(^1\)Problem 11 in http://www.aimath.org/WWN/randommatrices/randommatrices.pdf, Open problems: AIM workshop on random matrices.
Let $H = (H_{ij})$ be an $N \times N$ matrix with independent real valued entries with mean 0 and variance $N^{-1}$,

$$H_{ij} = N^{-1/2}h_{ij}, \quad \mathbb{E}h_{ij} = 0, \quad \mathbb{E}h_{ij}^2 = 1. \quad (1)$$

We assume the entries $h_{ij}$ have a sub-exponential decay, that is, there exists a constant $\theta > 0$ such that for $u > 1$,

$$\mathbb{P}(|h_{ij}| > u) \leq \theta^{-1} \exp(-u^\theta). \quad (2)$$

We remark that this assumption is mainly for convenience, and other conditions such as the existence of a sufficiently high moment would also be enough.

For an $N \times N$ matrix $H$, let $\sigma_1(H) \leq \cdots \leq \sigma_N(H)$ denote the singular values in non-decreasing order and use $\kappa(H) := \sigma_N(H)/\sigma_1(H)$ to denote the condition number. Throughout this paper, we let $G$ be an $N \times N$ Gaussian matrix with i.i.d. entries $\mathcal{N}(0, N^{-1})$.

To state the main results, we first introduce two important probabilistic notions that are commonly used throughout the whole paper.

**Definition 1** (Overwhelming probability). Let $\{\mathcal{E}_N\}$ be a sequence of events. We say that $\mathcal{E}_N$ holds with overwhelming probability if for any (large) $D > 0$, there exists $N_0(D)$ such that for all $N \geq N_0(D)$ we have

$$\mathbb{P}(\mathcal{E}_N) \geq 1 - N^{-D}.$$  

**Definition 2** (Stochastic domination). Let $\mathcal{X} = \{X_N\}$ and $\mathcal{Y} = \{Y_N\}$ be two families of nonnegative random variables. We say that $\mathcal{X}$ is stochastically dominated by $\mathcal{Y}$ if for all (small) $\varepsilon > 0$ and (large) $D > 0$, there exists $N_0(\varepsilon, D) > 0$ such that for $N \geq N_0(\varepsilon, D)$ we have

$$\mathbb{P}(X_N > N^\varepsilon Y_N) \leq N^{-D}.$$  

The stochastic domination is always uniform in all parameters. If $\mathcal{X}$ is stochastically dominated by $\mathcal{Y}$, we use the notation $\mathcal{X} \prec \mathcal{Y}$.

Our main result is the following.

**Theorem 1.** Let $H$ be an $N \times N$ random matrix satisfying (1) and (2). For any $\varepsilon > 0$ and any $\lambda > N^{-1/2+\varepsilon}$, we have

$$\left| \sigma_1(H + \lambda G) - \sqrt{1 + \lambda^2}\sigma_1(G) \right| \leq \frac{\sqrt{1 + \lambda^2}}{N^2 \log(1 + \lambda^2)}. \quad (3)$$

**Theorem 2.** Let $H$ be an $N \times N$ matrix satisfying (1) and (2). For any $\delta \in (0, 1)$ and $\varepsilon > 0$, we have

$$\mathbb{P} \left( N\sigma_1(G) > r + N^{-\delta} \right) - N^\varepsilon \left( N^{-1+\delta} \vee N^{-\frac{\delta}{2}} \right) \leq \mathbb{P} \left( N\sigma_1(H) > r \right) \leq \mathbb{P} \left( N\sigma_1(G) > r - N^{-\delta} \right) + N^\varepsilon \left( N^{-1+\delta} \vee N^{-\frac{\delta}{2}} \right), \quad (4)$$

where $a \vee b = \max\{a, b\}$ denotes the maximum between $a$ and $b$.

For the smallest singular value, one aspect of the complex-valued case is particularly interesting in the sense that the complex Gaussian model is explicitly integrable, i.e., the distribution of its smallest singular value is given by an exact formula. Specifically, let $G_C$ be an $N \times N$ matrix whose entries are i.i.d complex Gaussians whose real and imaginary parts are i.i.d. copies of $\mathcal{N}(0, N^{-1})$. For the complex Gaussian ensemble, Edelman proved in [Ede88] that the distribution
of the (renormalized) smallest singular value of a complex Gaussian ensemble is independent of $N$ and can be computed explicitly

$$P(N\sigma_1(G_C) \leq r) = \int_0^r e^{-x} dx = 1 - e^{-r}.$$  

Thanks to this exact formula for the integrable model, the edge universality for the smallest singular value can be quantified in terms of the Kolmogorov-Smirnov distance to the explicit law.

More precisely, let $H_C$ be an $N \times N$ random matrix satisfying

$$(H_C)_{ij} = N^{-1/2}h_{ij}, \quad \mathbb{E} h_{ij} = 0, \quad \mathbb{E}[(\text{Re} h_{ij})^2] = \mathbb{E}[(\text{Im} h_{ij})^2] = \frac{1}{2}, \quad \mathbb{E}[(\text{Re} h_{ij})(\text{Im} h_{ij})] = 0.$$  

and the sub-exponential decay assumption (2). Then we have the following rate of convergence to the explicit law.

**Corollary 1.** Let $H_C$ be a complex $N \times N$ random matrix defined as above, then for any $\varepsilon > 0$ we have

$$P(N\sigma_1(H_C) \leq r) = 1 - e^{-r} + O(N^{-\frac{1}{2}+\varepsilon}). \quad (5)$$

Based on the analysis of the smallest singular value, combined with results on the quantitative results of the largest singular values, we can also derive optimal smoothed analysis and quantitative universality for the condition number.

**Theorem 3.** Let $H$ be an $N \times N$ random matrix satisfying (1) and (2). For any $\varepsilon > 0$ and any $\lambda > N^{-1/2+\varepsilon}$, we have

$$|\kappa(H + \lambda G) - \kappa(G)| < \frac{1}{\log(1 + \lambda^2)}. \quad (6)$$

**Theorem 4.** Let $H$ be an $N \times N$ random matrix satisfying (1) and (2). For any $\varepsilon > 0$, we have

$$P\left(\frac{\kappa(G)}{N} > r + N^{-\frac{1}{2}+\varepsilon}\right) - N^{-\frac{1}{2}-\varepsilon} \leq P\left(\frac{\kappa(H)}{N} > r\right) \leq P\left(\frac{\kappa(G)}{N} > r - N^{-\frac{1}{2}+\varepsilon}\right) + N^{-\frac{1}{2}-\varepsilon} \quad (7)$$

As mentioned previously, the general hard edge case $\lim N/M = 1$ without $M \equiv N$ is notoriously difficult problem in random matrix theory. To further generalize our results, we have the following slight extension towards the general case.

**Theorem 5.** Let $H$ be an $M \times N$ random matrix satisfying (1) and (2) with $M = N + O(N^{\alpha(1)})$. Then all results in Theorem 1, Theorem 2, Theorem 3 and Theorem 4 are still true.

1.3 Outline of proofs

The central idea of this paper is based on the Erdős-Schlein-Yau dynamical approach in random matrix theory. In their seminal work [ESYY12], the so-called three-step strategy is developed to prove universality phenomena for random matrices. Roughly speaking, this framework is the following three ideas.

(i) A priori estimates for spectral statistics. This is based on the analyzing the resolvent of the matrix, and such analysis is called local law in random matrix theory. Local law states that the spectral density converges to the limiting law on microscopic scale. This local law implies the eigenvalue rigidity phenomenon, which states that the eigenvalues are close to their typical locations. Such a priori control of the eigenvalue locations will play a significant role in further analysis.
Local relaxation of eigenvalues. This step is designed to prove universality for matrices with a tiny Gaussian component. We perturb the matrix by some independent Gaussian noise, and then under this perturbation, the dynamics of the eigenvalues is governed by the Dyson Brownian motion (DBM). Moreover, the spectral distribution of the Gaussian ensemble is the equilibrium measure of DBM. The ergodicity of DBM results in a fast convergence to the local equilibrium, and hence implies the universality for matrix with small Gaussian perturbation.

Density arguments. For any probability distribution of the matrix elements, there exists a distribution with a small Gaussian component (in the sense of Step (ii)) such that the two associated random matrices have asymptotically identical spectral statistics. Typically, such an asymptotic identity is guaranteed by some moment matching conditions and the comparison of resolvents.

For a systematic discussion of this method, we refer to the monograph [EY17]. Following this strategy, our main techniques can also be divided into the following three steps.

- The first step is the local semicircle law for the symmetrization of the random matrix $H$. This local law guarantees the optimal rigidity estimates for the singular values. This step is based on classical works in random matrix theory such as [BEK+14, AEK14, AEK17].

- The second step is to interpolate the general matrix $H$ with the Gaussian matrix $G$, and estimate the dynamics of the singular value. More specifically, we consider the interpolation $H_t = e^{-t/2}H + \sqrt{1-e^{-t}}G$, which solves the matrix Ornstein-Uhlenbeck stochastic differential equation

$$\frac{dH_t}{\sqrt{N}} = dB_t - \frac{1}{2}H_t dt.$$ 

Note that this interpolation $H_t$ is equivalent to the matrix perturbation in our smoothed analysis. We consider a weighted Stieltjes transform (defined in (15)). A key innovation of our work is that, combined with a symmetrization trick, the evolution of the weighted Stieltjes along the dynamics of $H_t$ satisfies a stochastic PDE that can be well approximated by a deterministic advection equation. This deterministic PDE yields a rough estimate for $|\sigma_k(H_t) - \sigma_k(G)|$. Finally, using a delicate bootstrap argument, we show that the estimates for $|\sigma_k(H_t) - \sigma_k(G)|$ are self-improving. Iterating the bootstrap argument to optimal scale, we derive the optimal smoothed analysis for the smallest singular value.

- The last step is a quantitative resolvent comparison. In particular, the difference between the resolvent of two different random matrices are explicitly controlled in terms of the difference of their fourth moments. This comparison is proved via the Lindeberg exchange method. Together with the optimal smoothed analysis, this comparison theorem establishes the quantitative universality.

1.4 Notations and paper organizations

Throughout this paper, we denote $C$ a generic constant which does not depend on any parameter but may vary form line to line. We write $A \lesssim B$ if $A \leq CB$ holds for some constant $C$, and similarly write $A \gtrsim B$ if $A \geq C^{-1}B$. We also denote $A \sim B$ if both $C^{-1}B \leq A \leq CB$ hold. When $A$ and $B$ are complex valued, $A \sim B$ means $\text{Re} A \sim \text{Re} B$ and $\text{Im} A \sim \text{Im} B$. We use $[A, B] := [A, B] \cap \mathbb{Z}$ to denote the set of integers between $A$ and $B$. We use $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$ to denote the maximum and minimum between $a$ and $b$, respectively.
The paper is organized as follows. In Section 2, we discuss some applications of our results in numerical analysis and algorithms. In Section 3, we discuss the smoothed analysis for the smallest singular value of random matrices via the study of singular value dynamics. In Section 4, we use the smoothed analysis to establish a full quantitative universality for the smallest singular value. In Section 5, we use the results on the smallest singular value to derive smoothed analysis and quantitative universality for the condition number. In Section 6, we extend the result for square matrices to a slightly more general non-square case. Finally, in the Appendix, we collect some auxiliary results and provide the deferred technical proofs.

2 Applications in Numerical Analysis and Algorithms

In this section, we discuss some applications of our results in numerical analysis and algorithms. There are numerous circumstances where the smallest singular value and condition number play important roles. We do not intend to mention all of them and just focus on two simple scenarios in the framework of solving linear systems to illustrate the usefulness of our results. We expect our results can be applied in more complicated models and more advanced algorithms.

2.1 Accuracy of least-square solution

Consider the linear least-square optimization

$$\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2,$$

where $A$ is an $M \times N$ matrix satisfying $M - N = N^{o(1)}$, and $b \in \mathbb{R}^N$ is a fixed vector. The loss of precision of this problem, denoted by $\text{LoP}(A, b)$, is the number of correct digits in the entries of the data $(A, b)$ minus the same quantity for the computed solution. Let $\text{LoP}(A)$ denote the loss of precision for the worst $b$. Then, as shown in [Hig02], we have

$$\text{LoP}(A) = \log MN^{3/2} + 2 \log \kappa(A) + O(1).$$

Let $H$ be an $M \times N$ random matrix satisfying (1) and (2). Also let $G$ be an $M \times N$ Gaussian matrix. By Theorem 3 and Theorem 5, for any $\varepsilon > 0$, with overwhelming probability we have

$$\text{LoP}(H + \lambda G) \leq \log MN^{3/2} + 2 \log \kappa(G) + N^{-1+\varepsilon} \frac{1}{\log(1 + \lambda^2)} + O(1).$$

Also, using Theorem 4 and 5, for any $\varepsilon > 0$, with probability at least $1 - N^{-1/3 - \varepsilon}$, we have

$$\text{LoP}(H) \leq \log MN^{3/2} + 2 \log \kappa(G) + N^{-\frac{2}{3} + \varepsilon} + O(1)$$

The error terms are smaller than the $O(1)$ term. These results imply that a general random matrix and its Gaussian perturbation can ensure accuracy as good as in the Gaussian case.

2.2 Complexity of conjugate gradient method

Consider the linear equation

$$A^\top Ax = c,$$

where $A$ is an $M \times N$ matrix with $M - N = N^{o(1)}$, and $c \in \mathbb{R}^N$ is a fixed vector. This linear system can be solved via the conjugate gradient algorithm. Let $T_\delta(A)$ denote the needed iterations
to obtain an $\delta$-approximation of the true solution in the worst case. Then it is known (see e.g. [TB97]) that

$$T_\delta(A) = \frac{1}{2} \kappa(A) \delta.$$ 

Let $H$ be an $M \times N$ random matrix satisfying (1) and (2). Also let $G$ be an $M \times N$ Gaussian matrix. By Theorem 3 and Theorem 5, for any $\varepsilon > 0$, with overwhelming probability we have

$$|T_\delta(H + \lambda G) - T_\delta(G)| \leq \frac{\delta N \varepsilon}{\log(1 + \lambda^2)} \lesssim N \varepsilon \delta \lambda^2.$$ 

This shows that as long as $\lambda^2 \gg \delta$, the Gaussian perturbation $H + \lambda G$ has time complexity as good as the Gaussian ensemble.

Similarly, using Theorem 4 and Theorem 5, for any $\varepsilon > 0$, with probability at least $1 - N^{-1/3 - \varepsilon}$, we have

$$T_\delta(H) \leq T_\delta(G) + \delta N^{1/3 + \varepsilon}.$$ 

This shows that as long as the required accuracy satisfies $\delta \ll N^{-1/3}$, the time complexity for a general random matrix is as good as the Gaussian ensemble.

3 Smoothed Analysis and Gaussian Approximation

3.1 Singular value dynamics

In smoothed analysis, we are interested in matrix perturbation of the form $H + \lambda G$. After normalization of the variance, it is equivalent to study matrix of the form $H_t = e^{-t/2}H + \sqrt{1 - e^{-t}} G$. More specifically, we have

$$H + \lambda G = \sqrt{1 + \lambda^2} \left( \frac{1}{\sqrt{1 + \lambda^2}} H + \frac{\lambda}{\sqrt{1 + \lambda^2}} G \right) = \sqrt{1 + \lambda^2} H_{\log(1 + \lambda^2)}. \quad (8)$$ 

Let $B$ be an $N \times N$ matrix Brownian motion, i.e. $B_{ij}$ are independent standard Brownian motions. Then the evolution of $H_t$ is governed by the following matrix-valued Ornstein-Uhlenbeck process:

$$dH_t = \frac{1}{\sqrt{N}} dB_t - \frac{1}{2} H_t dt. \quad (9)$$ 

Let $\{s_k(t)\}_{k=1}^N$ denote the singular values of $H_t$, then $\{s_k(t)\}_{k=1}^N$ satisfy the following system of stochastic differential equations [ESYY12, equation (5.8)],

$$ds_k = \frac{dB_k}{\sqrt{N}} + \left[ -\frac{1}{2} s_k + \frac{1}{2N} \sum_{\ell \neq k} \left( \frac{1}{s_k - s_\ell} + \frac{1}{s_k + s_\ell} \right) \right] dt, \quad 1 \leq k \leq N. \quad (10)$$ 

To handle these SDEs, an important idea is the following symmetrization trick (see [CL19, equation (3.9)]):

$$s_\ell(t) = -s_i(t), \quad B_\ell(t) = -B_i(t), \quad \forall t \geq 0, \quad 1 \leq i \leq N.$$ 

With these notations, we label the indices from $-1$ to $-N$ and $1$ to $N$, so that the zero index is omitted. Unless otherwise stated, this will be the convention and we will not emphasize it explicitly in the following parts of the paper. After symmetrization, for the real case we have

$$ds_k = \frac{dB_k}{\sqrt{N}} + \left[ -\frac{1}{2} s_k + \frac{1}{2N} \sum_{\ell \neq \pm k} \frac{1}{s_k - s_\ell} \right] dt, \quad -N \leq k \leq N, k \neq 0. \quad (11)$$
Now we use the coupling method introduced in [LSY19] to analyze these dynamics. Consider the interpolation between a general matrix $H$ and a Gaussian matrix $G$. Let $\{\sigma_k(H)\}_{k=−N}^N$ and $\{\sigma_k(G)\}_{k=−N}^N$ be the (symmetrized) singular values of $H$ and $G$, respectively. For $\nu \in [0,1]$, define

$$s_k^{(\nu)}(0) = (1-\nu)\sigma_k(H) + \nu\sigma_k(G).$$

With this initial condition, we denote the unique solution of (11) by $\{s_k^{(\nu)}(t)\}$. Also, let $\{\sigma_k(H,t)\}$ and $\{\sigma_k(G,t)\}$ denote the solutions of (11) with initial conditions $\{\sigma_k(H)\}$ and $\{\sigma_k(G)\}$, respectively.

It is well known that the empirical measure of the eigenvalues of $H^*H$ converges to the Marchenko-Pastur distribution

$$\rho_{MP}(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}},$$

For $1 \leq k \leq N$, we define the typical position of the singular value $\sigma_k$ as the quantile $\gamma_k$ satisfying

$$\int_{-\infty}^{\gamma_k} \rho_{MP}(x) dx = \frac{k}{N}.$$ 

We also define $\gamma_{-k} = -\gamma_k$. By a change of variable, we have

$$\int_{-\infty}^{\gamma_k} \rho_{\text{sc}}(x) dx = \frac{N+k}{2N}, \quad \int_{-\infty}^{\gamma_{-k}} \rho_{\text{sc}}(x) dx = \frac{N-k}{2N}, (12)$$

where $\rho_{\text{sc}}(x) = \frac{1}{2\pi\sqrt{(4-x^2)+}}$ is the semicircle law.

An important input of our proof is the following uniform rigidity estimates. For any fixed $\varepsilon > 0$, consider the set of good trajectories

$$\mathcal{A}_\varepsilon = \left\{ \left| s_k^{(\nu)}(t) - \gamma_k \right| < N^{-\frac{2}{3}+\varepsilon} (N + 1 - |k|)^{-\frac{2}{3}} \right\} \text{ for all } 0 \leq t \leq 1, -N \leq k \leq N, 0 \leq \nu \leq 1. \right.$$ 

(13)

Such rigidity estimates for fixed $t$ and $\nu = 0$ or 1 were proved in [AEK14, AEK17, BEK+14, BYY14, CMS13]. The extension to uniform estimates in parameters can be done by a discretization argument: (1) discretize in $t$ and $\nu$; (2) use Weyl’s inequality to control increments over small time intervals; (3) use a maximum principle for the derivative with respect to the $\nu$ parameter (see Lemma 3.2) to control increments in small $\nu$-intervals. As a consequence, we have

**Lemma 3.1.** For any $\varepsilon > 0$, the event $\mathcal{A}_\varepsilon$ happens with overwhelming probability, i.e. for any $D > 0$, there exists $N_0(\varepsilon,D)$ such that for any $N > N_0$ we have

$$\mathbb{P}(\mathcal{A}_\varepsilon) > 1 - N^{-D}.$$ 

We consider

$$\varphi_k^{(\nu)}(t) := e^{\frac{t}{2}} \frac{d}{d\nu} s_k^{(\nu)}(t).$$

For the simplicity of notations, we omit the parameter $\nu$ if the context is clear. Then $\varphi_k$ satisfies the following non-local parabolic type equation

$$\frac{d}{dt} \varphi_k = \frac{1}{2N} \sum_{\ell \neq \pm k} \frac{\varphi_{\ell} - \varphi_k}{(s_{\ell} - s_k)^2}.$$ 

(14)

Let $\psi_k = \psi_k^{(\nu)}$ solve the same equation as $\varphi_k$ in (14) but with initial condition $\psi_k(0) = |\varphi_k(0)| = |\sigma_k(H) - \sigma_k(G)|$. Following the same arguments in [Wan19, Lemma 3.1], this equation satisfies a maximum principle.
Lemma 3.2. For all \( t \geq 0 \) and \( -N \leq k \leq N \), we have
\[
\psi_k(t) = \psi_{-k}(t), \quad \psi_k(t) \geq 0, \quad |\psi_k(t)| \leq \max_{-N \leq \ell \leq N} |\psi_{\ell}(0)|, \quad |\varphi_k(t)| \leq \psi_k(t).
\]

We consider the following weighted Stieltjes transform
\[
\mathcal{S}_t(z) := e^{-\frac{t}{2}} \sum_{-N \leq k \leq N} \frac{\varphi_k(t)}{s_k(t) - z}, \quad \mathcal{\bar{S}}_t(z) := e^{-\frac{t}{2}} \sum_{-N \leq k \leq N} \frac{\psi_k(t)}{s_k(t) - z}.
\]

Let \( S_t(z) \) and \( m_{sc}(z) \) denote the Stieltjes transforms of the empirical measure for the singular values and of the semicircle law
\[
S_t(z) = \frac{1}{2N} \sum_{-N \leq k \leq N} \frac{1}{s_k - z}, \quad m_{sc}(z) = \frac{-z + \sqrt{z^2 - 4}}{2}.
\]

A well-known result in random matrix theory is the following local semicircle law for the Stieltjes transform \( S_t(z) \). Let \( \omega > 0 \) be an arbitrarily fixed constant. Define the spectral domain
\[
\mathcal{S} = \mathcal{S}_\omega := \{ z = E + i\eta : |E| \leq \omega^{-1}, N^{-1+\omega} \leq \eta \leq \omega^{-1} \}.
\]

For any \( \omega > 0 \) and \( z \in \mathcal{S}_\omega \), it was shown in [AEK14, AEK17] that
\[
|S_t(z) - m_{sc}(z)| < \frac{1}{N\eta}.
\]

As computed in [Wan19, Lemma 3.3], a key result is that \( \mathcal{S}_t(z) \) and \( \mathcal{\bar{S}}_t(z) \) satisfy the following stochastic advection equation.

Lemma 3.3. For \( \text{Im} z \neq 0 \), we have
\[
d\mathcal{\bar{S}}_t = \left( S_t(z) + \frac{z}{2} \right) (\partial_z \mathcal{\bar{S}}_t)dt + \frac{1}{4N} (\partial_{zz} \mathcal{\bar{S}}_t)dt + \left[ \frac{e^{-\frac{t}{2}}}{2N} \sum_{-N \leq k \leq N} \frac{\psi_k}{(s_k - z)^2(s_k + z)} \right] dt
\]
\[
- \frac{e^{-\frac{t}{2}}}{\sqrt{N}} \sum_{-N \leq k \leq N} \frac{\psi_k}{(s_k - z)^2} dB_k.
\]

Based on the local semicircle law (16), we expect that this stochastic differential equation can be approximated by the deterministic advection PDE
\[
\partial_t h = \frac{\sqrt{z^2 - 4}}{2} \partial_z h.
\]

The above PDE have the following explicit characteristics
\[
z_t = \frac{e^{\frac{t}{2}} (z + \sqrt{z^2 - 4}) + e^{-\frac{t}{2}} (z - \sqrt{z^2 - 4})}{2}.
\]

This implies \( \mathcal{\bar{S}}_t(z) \approx \mathcal{\bar{S}}_0(z_t) \), and we will justify this approximation in the next subsection. Moreover, we remark that \( \mathcal{S}_t \) satisfies the same equation (17) with \( \psi_k \) replaced by \( \varphi_k \).

Before moving to the main estimates, we first collect some basic results, including the geometry of the characteristics and a rough estimate for the initial condition. These can be proved via direct computations and the details can be found in [Bou22, Section 2].
Let $\xi(z) = \min\{\lvert z - 2 \rvert, \lvert z + 2 \rvert\}$. For any $\varepsilon > 0$, we consider the curve and the domain

$$\mathcal{S}_\varepsilon = \left\{ E + i\eta : -2 + N^{-1+4\varepsilon} < E < 2 - N^{-1+4\varepsilon}, \eta = N^{-1+4\varepsilon} \xi(E)^{-1/2} \right\}, \quad \mathcal{R}_\varepsilon = \bigcup_{0 \leq t \leq 1} \{ z_t : z \in \mathcal{S}_\varepsilon \}.$$

We also define $a(z) = \text{dist}(z, [-2, 2])$ and $b(z) = \text{dist}(z, [-2, 2]^c)$.

**Lemma 3.4.** Uniformly in $0 < t < 1$ and $z = z_0$ satisfying $\eta = \text{Im} z > 0$ and $\lvert z - 2 \rvert < \frac{1}{10}$, we have

$$\text{Re}(z_t - z_0) \sim t \frac{a(z)}{\xi(z)^{1/2}} + t^2, \quad \text{Im}(z_t - z_0) \sim t \frac{b(z)}{\xi(z)^{1/2}}.$$

In particular, for $\varepsilon > 0$, if $z \in \mathcal{S}_\varepsilon$, then $z_t - z_0 \sim (tN^{-1+4\varepsilon} \xi(E)^{-1} + t^2) + i\xi(E)^{1/2}t$.

Moreover, for any $\xi > 0$, and $z = E + i\eta \in [-2 + \xi, 2 - \xi] \times [0, \xi^{-1}]$, we have $\text{Im}(z_t - z_0) \sim t$.

**Lemma 3.5.** Let $\varepsilon > 0$ be any small constant. In the set $\mathcal{S}_\varepsilon$, for any $z = E + i\eta \in \mathcal{R}_\varepsilon$, we have $\text{Im} \tilde{S}_0(z) \lesssim N^\varepsilon$ if $\eta > \max(E-2, -E-2)$, and $\text{Im} \tilde{S}_0(z) \lesssim N^\varepsilon \xi(z)^{-1}\eta$ otherwise. The same bounds also hold for $\lvert \text{Im} \tilde{S}_0 \rvert$.

A key ingredient of our proof is the following a priori estimate for $\tilde{S}_t$, whose proof is deferred to Appendix B.1.

**Proposition 3.1.** Let $\varepsilon > 0$ be any small constant. Uniformly for all $0 < t < 1$ and $z = E + i\eta \in \mathcal{S}_\varepsilon$, with overwhelming probability we have

$$\text{Im} \tilde{S}_t(z) \lesssim N^{2\varepsilon} \frac{\xi(E)^{1/2}}{(\xi(E)^{1/2} \lor t)}.$$

This estimate yields a rough control for the decay of $\varphi_k(t)$, which will be an important input for more refined estimates.

**Lemma 3.6.** For all $-N \leq k \leq N$ and $0 \leq t \leq 1$, we have

$$|\varphi_k(t)| \leq \frac{1}{N} \frac{1}{\left(\frac{N+1-|k|}{N}\right)^{1/3} \lor t}$$

(20)

Proof. By Lemma 3.2, it suffices to control $\psi_k(t)$. By the nonnegativity of $\psi_k(t)$, we have

$$\text{Im} \tilde{S}_t(z) = \sum_{-N \leq k \leq N} \frac{\psi_k(t) \text{Im} z}{|s_k(t) - z|^2} \geq \psi_k(t) \frac{\text{Im} z}{|s_k(t) - z|^2},$$

which implies

$$\psi_k(t) \leq \text{Im} \tilde{S}_t(z) \frac{|s_k(t) - z|^2}{\text{Im} z}.$$

Let $\varepsilon > 0$. For $(N + 1 - |k|) > N^{10\varepsilon}$, pick the point $z = \gamma_k + iN^{-1+4\varepsilon} \xi(\gamma_k)^{-1/2} \in \mathcal{S}_\varepsilon$. In this case we have $\xi(\gamma_k)^{1/2} \sim (\frac{N+1-|k|}{N})^{1/3}$. Therefore, in the set $\mathcal{S}_\varepsilon$, by Proposition 3.1, uniformly for all $-N + N^{10\varepsilon} \leq k \leq N - N^{10\varepsilon}$ and $0 \leq t \leq 1$, with overwhelming probability we have

$$|\psi_k(t)| \leq \frac{N^{8\varepsilon}}{N} \frac{1}{\left(\frac{N+1-|k|}{N}\right)^{1/3} \lor t}.$$

For $(N + 1 - |k|) \leq N^{10\varepsilon}$, without loss of generality we consider $N + 1 - k \leq N^{10\varepsilon}$. In this case, let $k_0 = N - N^{10\varepsilon} + 1$ and consider $z = \gamma_{k_0} + iN^{-1+4\varepsilon} \xi(\gamma_{k_0})^{-1/2}$. The same argument results in a similar bound with a larger $N^{20\varepsilon}$ factor. By the arbitrariness of $\varepsilon$, this completes the proof. \qed
3.2 Local relaxation at the hard edge

In this subsection we prove a quantitative estimate for the local relaxation flow \(11\) at the hard edge. The main estimate in this section is the following. We remark that Theorem 6 is equivalent to Theorem 1 using the rescaling \(8\).

**Theorem 6.** For \(\varepsilon_0 > 0\) arbitrarily small and any \(N^{-1+\varepsilon_0} < t < 1\), we have

\[
|\sigma_1(H, t) - \sigma_1(G, t)| < \frac{1}{N^2 t}.
\]

To estimate \(\varphi_k(t)\) near the hard edge, we introduce the following quantity to approximate it. Let \(\gamma_k = (\gamma_k)_t\) with the convention \(\gamma^t = (\gamma + i0^+)_t\), and define

\[
\tilde{\varphi}_k(t) := \frac{1}{2N \Im m_{sc}(\gamma_k)} \sum_{-N \leq j \leq N} \Im \left( \frac{1}{\gamma_j - \gamma_k} \right) (\sigma_j(H) - \sigma_j(G)).
\]

Our goal is to prove the following estimates

**Proposition 3.2.** Let \(0 < c < 1\) be a fixed small constant. For \(\varepsilon_0 > 0\) arbitrarily small with any \(N^{-1+\varepsilon_0} < t < 1\) and \(k \in [(c-1)N, (1-c)N]\), we have

\[
|\sigma_k(H, t) - \sigma_k(G, t) - \tilde{\varphi}_k(t)| \lesssim \frac{1}{N^2 t}.
\]

To obtain the optimal control for the local relaxation flow, we need to carefully estimate \(\tilde{\varphi}_k\) near the hard edge. A first step towards such estimates is given in the following lemma.

**Lemma 3.7.** Let \(\varepsilon > 0\) and \(0 < c < 1\). For any \((k, \ell) \in [(c-1)N, (1-c)N]^2\), \(|E| < 2 - c\), and \(s, t, \eta \in [N^{-1+4\varepsilon}, 1]\), in the set \(A\), for \(z = E + i\eta\) we have

\[
|\tilde{\varphi}_k(t) - \tilde{\varphi}_\ell(s)| \lesssim N^\varepsilon \left( \frac{|k - \ell|}{N^2(s \wedge t)} + \frac{|s - t|}{N(s \wedge t)} \right),
\]

\[
\left| \frac{\Im G_0(z_t)}{2N \Im S_0(z_t)} - \tilde{\varphi}_\ell(s) \right| \lesssim N^\varepsilon \left( \frac{|E - \gamma_t|}{N(s \wedge (\eta + t))} + \frac{|\eta + t - s|}{N(s \wedge (\eta + t))} \right).
\]

**Proof.** By the properties of the Stieltjes transform \(m_{sc}(z)\) (see e.g. [EY17, Section 6]) and direct computation, we have

\[
\Im m_{sc}(\gamma_k^t) \gtrsim 1, \quad \Im m_{sc}(\gamma_t^s) \gtrsim 1,
\]

and

\[
|\Im m_{sc}(\gamma_k^t) - \Im m_{sc}(\gamma_t^s)| + |\gamma_k^t - \gamma_t^s| \lesssim \frac{|k - \ell|}{N} + |s - t|.
\]

In the set \(A\), the rigidity estimates imply that

\[
|\sigma_j(H) - \sigma_j(G)| \lesssim N^{-\frac{2}{3} + \varepsilon}(N + 1 - |j|)^{-\frac{1}{3}}.
\]

Then we have

\[
\left| \frac{1}{2N \Im} \left( \sum_{-N \leq j \leq N} \frac{\sigma_j(H) - \sigma_j(G)}{\gamma_j - \gamma_k} \right) \right| \lesssim \frac{1}{2N \Im} \left| \sum_{-N \leq j \leq N} \frac{N^{-\frac{2}{3} + \varepsilon}(N + 1 - |j|)^{-\frac{1}{3}}}{(\gamma_j - \Re \gamma_k^t)^2 + (\Im \gamma_k^t)^2} \right|
\]

\[
\lesssim N^{-1+\varepsilon} \int_{-2}^{2} \frac{\xi(x)^{-\frac{1}{2}}}{(x - \Re \gamma_k^t)^2 + (\Im \gamma_k^t)^2} \frac{1}{d\rho_{sc}(x)}
\]

\[
\lesssim N^{-1+\varepsilon} \int_{-2}^{2} \frac{d\rho_{sc}(x)}{(x - \Re \gamma_k^t)^2 + (\Im \gamma_k^t)^2}
\]

\[
\lesssim N^{-1+\varepsilon}.
\]
By triangle inequality, we have

\[ |\hat{\varphi}_k(t) - \hat{\varphi}_\ell(s)| \leq \frac{1}{2N} \left| \frac{1}{\text{Im } m_{sc}(\gamma^t_k)} - \frac{1}{\text{Im } m_{sc}(\gamma^s_\ell)} \right| \text{Im} \left( \sum_{-N \leq j \leq N} \frac{\sigma_j(H) - \sigma_j(G)}{\gamma_j - \gamma^t_k} \right) + \frac{1}{2N\text{Im } m_{sc}(\gamma^s_\ell)} \sum_{-N \leq j \leq N} \left| \text{Im} \left( \frac{1}{\gamma_j - \gamma^t_k} \right) - \text{Im} \left( \frac{1}{\gamma_j - \gamma^s_\ell} \right) \right| |\sigma_j(H) - \sigma_j(G)| \]

\[ =: I_1 + I_2. \]

Using (25) and (26), we obtain

\[ I_1 \lesssim N^\varepsilon \left( \frac{|k - \ell|}{N^2} + \frac{|s - t|}{N} \right). \]

For the second term \( I_2 \), in the set \( \mathcal{A}_\varepsilon \), the rigidity and (25) imply that

\[ I_2 \lesssim N^{-1+\varepsilon} \sum_{-N \leq j \leq N} N^{-\frac{2}{3}} (N + 1 - |j|)^{-\frac{1}{3}} \left| \frac{\gamma^t_k - \gamma^s_\ell}{(\gamma_j - \gamma^t_k)(\gamma_j - \gamma^s_\ell)} \right| \]

\[ \lesssim N^{-1+\varepsilon} \left( \frac{|k - \ell|}{N} + |s - t| \right) \sum_{-N \leq j \leq N} N^{-\frac{2}{3}} (N + 1 - |j|)^{-\frac{1}{3}} \left( \frac{1}{|\gamma_j - \gamma^t_k|^2} + \frac{1}{|\gamma_j - \gamma^s_\ell|^2} \right). \]

Recall from Lemma 3.4 that \( \text{Im } \gamma^t_k \sim t \) and \( \text{Im } \gamma^s_\ell \sim s \). Using a similar argument as in (26) we obtain

\[ I_2 \lesssim N^{-1+\varepsilon} \left( \frac{|k - \ell|}{N} + |s - t| \right) \left( \frac{1}{t} + \frac{1}{s} \right) \lesssim N^\varepsilon \left( \frac{|k - \ell|}{N^2(s \wedge t)} + \frac{|s - t|}{N(s \wedge t)} \right). \]

Hence we have proved (23). For the other part (24), it can be proved via the same arguments. \( \square \)

As a consequence, we have a good control for the size of \( \hat{\varphi}_k(t) \) away from the soft edge. This is based on the symmetric structure of \( \{\hat{\varphi}_k\} \).

**Lemma 3.8.** Let \( \varepsilon > 0 \) and \( 0 < c < 1 \). For any \( t \in [N^{-1+4\varepsilon}, 1] \) and \( k \in [(c - 1)N, (1 - c)N] \), with overwhelming probability we have

\[ |\hat{\varphi}_k(t)| \lesssim N^\varepsilon \frac{k}{N^{2t}}. \]  

**Proof.** A key observation is the following

\[ \text{Re } \gamma^t_{-k} = -\text{Re } \gamma^t_k, \quad \text{Im } \gamma^t_{-k} = \text{Im } \gamma^t_k, \quad \text{Re } m_{sc}(\gamma^t_{-k}) = -\text{Re } m_{sc}(\gamma^t_k), \quad \text{Im } m_{sc}(\gamma^t_{-k}) = \text{Im } m_{sc}(\gamma^t_k). \]

Therefore, we have

\[ \hat{\varphi}_{-k}(t) = \frac{1}{2N\text{Im } m_{sc}(\gamma^t_k)} \sum_{-N \leq j \leq N} \text{Im} \left( \frac{1}{\gamma_j - \gamma^t_{-k}} \right) (\sigma_j(H) - \sigma_j(G)) \]

\[ = \frac{1}{2N\text{Im } m_{sc}(\gamma^t_k)} \sum_{-N \leq j \leq N} \frac{\text{Im } (\gamma^t_{-k})}{(\gamma_j - \text{Re } (\gamma^t_{-k}))^2 + (\text{Im } \gamma^t_{-k})^2} (\sigma_j(H) - \sigma_j(G)) \]

\[ = \frac{1}{2N\text{Im } m_{sc}(\gamma^t_k)} \sum_{-N \leq j \leq N} \frac{\text{Im } (\gamma^t_k)}{(\gamma_j + \text{Re } (\gamma^t_k))^2 + (\text{Im } \gamma^t_k)^2} (\sigma_j(H) - \sigma_j(G)) \]

14
Using the symmetrization of the singular values, we further have
\[
\hat{\varphi}_{-k}(t) = -\frac{1}{2N\text{Im } m_{sc}(\gamma_k)} \sum_{-N \leq j \leq N} \frac{\text{Im } (\gamma_j^H)}{\gamma_j - \gamma_k^H} \left( \sigma_{-j}(H) - \sigma_{-j}(G) \right)
\]
\[
= -\frac{1}{2N\text{Im } m_{sc}(\gamma_k)} \sum_{-N \leq j \leq N} \text{Im } \left( \frac{1}{\gamma_j - \gamma_k^H} \right) \left( \sigma_{-j}(H) - \sigma_{-j}(G) \right)
\]
\[
= -\hat{\varphi}_k(t)
\]
Consequently, by (23) we have
\[
|\hat{\varphi}_k(t) - \hat{\varphi}_{-k}(t)| = 2|\hat{\varphi}_k(t)| \lesssim N^\varepsilon \frac{2k}{N^2 t}.
\]
This shows the desired result.

Finally, it’s straightforward to derive Theorem 6 from Proposition 3.2 and Lemma 3.8. Thus, our primary goal is to prove Proposition 3.2.

### 3.3 Bootstrap arguments

We will prove the main technical estimate Proposition 3.2 via a bootstrap argument.

**Definition 3 (Hypothesis \( \mathcal{H}_\alpha \)).** Consider the following hypothesis: For any fixed small \( 0 < c < 1 \), the following holds for \( \varepsilon_0 > 0 \) arbitrarily small. For any \( N^{-1+\varepsilon_0} < t < 1 \), \( k \in \lfloor (c-1)N, (1-c)N \rfloor \) and \( \nu \in [0,1] \), we have
\[
| \varphi_{(\nu)}^k(t) - \hat{\varphi}_k(t) | \lesssim \frac{(Nt)\alpha}{N^2 t}.
\]

Proposition 3.2 is derived via a bootstrap of the hypothesis \( \mathcal{H}_\alpha \). Specifically, we have the following two lemmas.

**Lemma 3.9.** The hypothesis \( \mathcal{H}_1 \) is true.

**Proof.** Recall from Lemma 3.6 that (20) implies that \( \varphi_{(\nu)}^k(t) \lesssim N^{-1} \). On the other hand, from the definition (22) of \( \hat{\varphi}_k \), using the rigidity and (26) we obtain \( \hat{\varphi}_k(t) \lesssim N^{-1} \) thanks to the arbitrariness of \( \varepsilon_0 \). Therefore, the triangle inequality yields \( | \varphi_{(\nu)}^k(t) - \hat{\varphi}_k(t) | \lesssim N^{-1} \), which completes the proof.

**Lemma 3.10.** If \( \mathcal{H}_\alpha \) is true, then \( \mathcal{H}_{3\alpha/4} \) is true, i.e.
\[
| \varphi_{(\nu)}^k(t) - \hat{\varphi}_k(t) | \lesssim \frac{(Nt)\frac{3\alpha}{4}}{N^2 t}.
\]

The self-improving property of the hypothesis \( \mathcal{H}_\alpha \) stated in Lemma 3.10 is the main technical part of the proof for Proposition 3.2. We defer its proof to Appendix B.2.

Finally, the optimal control (21) for the local relaxation flow at the hard edge follows from these two lemmas together with Lemma 3.8.
Proof of Proposition 3.2. Note that

\[
\sigma_k(H,t) - \sigma_k(G,t) - \tilde{\sigma}_k(t) = \int_0^1 \varphi_k^{(\nu)}(t) \, d\nu - \tilde{\varphi}_k(t) = \int_0^1 \left( \varphi_k^{(\nu)}(t) - \tilde{\varphi}_k(t) \right) \, d\nu \quad (29)
\]

Consider an arbitrarily fixed \( \delta > 0 \), based on Lemma 3.9 and Lemma 3.10, after a finite time of iterations, with overwhelming probability we have

\[
|\varphi_k(t) - \tilde{\varphi}_k(t)| \leq \frac{N^\delta}{N^{2t}}
\]

This shows that for any fixed \( \tilde{D} \) and \( p \), and for large enough \( N \), we have

\[
\mathbb{E} \left( |\varphi_k(t) - \tilde{\varphi}_k(t)|^{2p} \right) \leq \left( \frac{N^\delta}{N^{2t}} \right)^{2p} + N^{-\tilde{D}}.
\]

By (29) we obtain

\[
\mathbb{E} \left( |\sigma_k(H,t) - \sigma_k(G,t) |^{2p} \right) \leq \int_0^1 \mathbb{E} \left( |\varphi_k(t) - \tilde{\varphi}_k(t)|^{2p} \right) \, d\nu \leq \left( \frac{N^\delta}{N^{2t}} \right)^{2p} + N^{-\tilde{D}}.
\]

We choose \( p = D/\delta \) and \( \tilde{D} = D + 100p \), and then the Markov inequality yields

\[
\mathbb{P} \left( |\sigma_k(H,t) - \sigma_k(G,t) - \tilde{\sigma}_k(t)| \leq \frac{N^2\delta}{N^{2t}} \right) > 1 - N^{-D},
\]

which completes the proof thanks to the arbitrariness of \( \delta \) and \( D \). \( \square \)

4 Quantitative Universality

4.1 Quantitative resolvent comparison

In classical random matrix theory, the spectral universality is proved by comparison of the resolvent for matrices with some moment matching conditions. To obtain a quantitative universality for the smallest singular value, we need a quantitative version of the resolvent comparison theorem.

For a fixed constant \( a \in (1,2) \), let \( \rho = \rho(N) \in [N^{-a} N^{-1}] \) be a cutoff scaling. Let \( r > 0 \) and consider two symmetric functions \( f_1(x), f_2(x) \) that are non-increasing in \( |x| \), given by

\[
f_1(x) := \begin{cases} 0 & \text{if } |x| > rN^{-1} \\ 1 & \text{if } |x| < rN^{-1} - \rho \end{cases}, \quad f_2(x) := \begin{cases} 0 & \text{if } |x| > rN^{-1} + \rho \\ 1 & \text{if } |x| < rN^{-1} \end{cases}.
\]

Also, consider a fixed non-increasing smooth function \( F \) such that \( F(x) = 1 \) for \( x \leq 0 \) and \( F(x) = 0 \) for \( x \geq 1 \).

A key observation is that the functions \( f_1, f_2 \) and \( F \) can bound the distribution of the smallest singular value \( \sigma_1(H) \). For any function \( f : \mathbb{R} \to \mathbb{R} \), we denote \( \text{Tr} f(H) := \sum_{i=-N}^N f(\sigma_i(H)) \).

Lemma 4.1. We have

\[
\mathbb{E} \left[ F \left( \text{Tr} f_1(H) \right) \right] \leq \mathbb{P} \left( \sigma_1(H) > rN^{-1} \right) \leq \mathbb{E} \left[ F \left( \text{Tr} f_2(H) \right) \right], \quad (30)
\]
Proof. For the right-hand side, assume \( \sigma_1(H) > rN^{-1} \). By definition of the function \( f_1 \), we have
\[
\sum_{i=-N}^{N} f_1(\sigma_i(H)) = 0,
\]
which implies \( F(\text{Tr} f_1(H)) = 1 \). Also note that \( F \geq 0 \). Therefore, we conclude
\[
P (\sigma_1(H) > rN^{-1}) = E \left[ \mathbf{1} \{ \sigma_1(H) > rN^{-1} \} \right] \leq E \left[ F(\text{Tr} f_1(H)) \right].
\]
The left-hand side can be proved similarly.

When estimating the distribution \( P (\sigma_1(H) > rN^{-1}) \), thanks to the rigidity of singular values, we can assume \( r < N^\varepsilon \) without loss of generality, where \( \varepsilon > 0 \) is a constant that can be arbitrarily small. Based on Lemma 4.1, to compare the distribution of the smallest singular values of different random matrices, it suffices to compare the functions \( \text{Tr} f_1 \) and \( \text{Tr} f_2 \). In the remaining part of this section, we provide a systematic treatment of such a comparison.

Pick a point \( E \in \mathbb{R} \) with \( 0 < E < N^{-1+\varepsilon} \). Let \( f(x) \) be a smooth symmetric function that is non-increasing in \( |x| \) satisfying
\[
f(x) = \begin{cases} 
0 & \text{if } |x| > E \\
1 & \text{if } |x| < E - \rho,
\end{cases} \quad \text{and} \quad \|f^{(k)}\|_\infty \lesssim \rho^{-k} \text{ for } k = 1, 2. \tag{31}
\]
For the test functions \( f \) and \( F \) defined as above, we have the following quantitative comparison of the resolvents, whose proof is deferred to Appendix C.

**Proposition 4.1.** Let \( X \) and \( Y \) be two independent random matrices satisfying (1) and (2). Assume the first three moments of the entries are identical, i.e. \( E[X_{ij}^k] = E[Y_{ij}^k] \) for all \( 1 \leq i, j \leq N \) and \( 1 \leq k \leq 3 \). Suppose also that for some parameter \( t = t(N) \) we have
\[
\left| E[(\sqrt{N}X_{ij})^4] - E[(\sqrt{N}Y_{ij})^4] \right| \leq t, \quad \text{for all } 1 \leq i, j \leq N. \tag{32}
\]
With the test functions \( f \) and \( F \) defined as above, there exists a constant \( C > 0 \) such that the following is true for any \( \varepsilon > 0 \)
\[
|E[F(\text{Tr} f(X))] - E[F(\text{Tr} f(Y))]| \leq NC\varepsilon \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a \right), \tag{33}
\]

### 4.2 Proof of Theorem 2

Using the quantitative comparison theorem (Proposition 4.1) and the smoothed analysis (Theorem 6), we now prove the quantitative universality.

For a general random matrix \( H \) satisfying Assumptions (1) and (2), there exists another matrix \( H'_0 \) that also satisfies the same assumptions such that the matrix \( H'_t := e^{-t/2}H'_0 + (1 - e^{-t})^{1/2}G \) has the same first three moments as \( H \) and the difference between the fourth moments (in the sense of (32)) is \( O(t) \). This is guaranteed by [EYY11, Lemma 3.4].

Lemma 4.1 and Proposition 4.1 yields
\[
E \left[ F(\text{Tr} f_2(H'_t)) \right] - NC\varepsilon \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a \right) \leq P (\sigma_1(H) > rN^{-1})
\]
\[
\quad \leq E \left[ F(\text{Tr} f_1(H'_t)) \right] + NC\varepsilon \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a \right). \tag{34}
\]
Using Lemma 4.1 for $H'_t$ with $f_1$ and $f_2$ shifted by $\pm \rho$, we have

$$\mathbb{P} \left( \sigma_1(H'_t) > \frac{r}{N} + \rho \right) - N^{C\varepsilon} \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t \rho N^a \right) \leq \mathbb{P} \left( \sigma_1(H) > \frac{r}{N} \right)$$

$$\leq \mathbb{P} \left( \sigma_1(H'_t) > \frac{r}{N} - \rho \right) + N^{C\varepsilon} \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t \rho N^a \right).$$ (35)

Using smoothed analysis Theorem 6, we have

$$\mathbb{P} \left( \sigma_1(G) > \frac{r}{N} + \rho + \frac{1}{N^2 t} \right) - N^{C\varepsilon} \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t \rho N^a \right) \leq \mathbb{P} \left( \sigma_1(H) > \frac{r}{N} \right)$$

$$\leq \mathbb{P} \left( \sigma_1(G) > \frac{r}{N} - \rho - \frac{1}{N^2 t} \right) + N^{C\varepsilon} \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t \rho N^a \right).$$ (36)

Taking $\rho = N^{-a}$, $t = N^{a-2}$ and setting $a = 1 + \delta$, we obtain the optimal bounds

$$\mathbb{P} \left( N \sigma_1(G) > r + N^{-\delta} \right) - N^{C\varepsilon} \left( N^{-1+\delta} \lor N^{-\frac{1}{2}} \right) \leq \mathbb{P} \left( N \sigma_1(H) > r \right)$$

$$\leq \mathbb{P} \left( N \sigma_1(G) > r - N^{-\delta} \right) + N^{C\varepsilon} \left( N^{-1+\delta} \lor N^{-\frac{1}{2}} \right).$$ (37)

Hence, thanks to the arbitrariness of $\varepsilon$, we have proved Theorem 2.

Finally, for the complex case, using the exact formula for the distribution of $\sigma_1(G_C)$, we obtain a rate of convergence to the limiting law. Recall that

$$\mathbb{P}(N \sigma_1(G_C) \leq r) = \int_0^r e^{-x} \, dx = 1 - e^{-r}.$$

**Proof of Corollary 1.** For the complex case, the previous arguments are still valid. Therefore, we still have (37). Since $N \sigma_1(G_C)$ has a bounded density, we have

$$\mathbb{P}(N \sigma_1(G_C) \leq r) - N^\varepsilon \left( N^{-\delta} + (N^{-1+\delta} \lor N^{-1/2}) \right) \leq \mathbb{P}(N \sigma_1(H_C) \leq r)$$

$$\leq \mathbb{P}(N \sigma_1(G_C) \leq r) + N^\varepsilon \left( N^{-\delta} + (N^{-1+\delta} \lor N^{-1/2}) \right)$$ (38)

Choosing $\delta = 1/2$, we obtain the optimal estimate

$$\mathbb{P}(N \sigma_1(H_C) \leq r) = 1 - e^{-r} + N^{-\frac{1}{2}+\varepsilon},$$

which proves the desired result. \(\square\)

## 5 Condition Number

### 5.1 Smoothed analysis

Note that the condition number $\kappa(H)$ is scaling invariant in the sense that $\kappa(a H) = \kappa(H)$ for any $a > 0$. Therefore, in the smoothed analysis, it suffices to consider $H_t = e^{-t/2} H + (1 - e^t)^{1/2} G$, whose singular values satisfies the stochastic differential equation (10).
Recall from Theorem 6, we have shown that
\[ |\sigma_1(H_t) - \sigma_1(G)| \lesssim \frac{1}{N^2 t}. \]
Note that in Lemma 3.6, using the same arguments as in Proposition 3.2, we can derive that
\[ |\sigma_N(H_t) - \sigma_N(G)| \lesssim \frac{1}{N t}. \]
Then for any large \( D > 0 \) and small \( \varepsilon_1, \varepsilon_2 > 0 \), there exists \( N_0(\varepsilon_1, \varepsilon_2, D) \) such that the following holds with probability at least \( 1 - N^{-D} \).
\[ \sigma_1(G) - \frac{N \varepsilon_1}{N^2 t} \leq \sigma_1(H_t) \leq \sigma_1(G) + \frac{N \varepsilon_1}{N^2 t}, \quad \sigma_N(G) - \frac{N \varepsilon_2}{N t} \leq \sigma_N(H_t) \leq \sigma_N(G) + \frac{N \varepsilon_2}{N t}. \]
Without loss of generality, we assume that \( \varepsilon_1 \sim \varepsilon_2 \sim \varepsilon \) for some \( \varepsilon > 0 \) that can be arbitrarily small. Then we have
\[ \frac{\kappa(H_t)}{N} = \frac{\sigma_N(H_t)}{N \sigma_1(H_t)} \leq \frac{\sigma_N(G) + \frac{N \varepsilon_2}{N t}}{N \sigma_1(G) - \frac{N \varepsilon_1}{N t}} \leq \frac{\kappa(G)}{N} + \frac{N C \varepsilon}{N t}, \]
where in the last inequality we use that \( N \sigma_1(G) \) is of order constant with overwhelming probability. By the arbitrariness of \( \varepsilon > 0 \), we can relabel the parameter and then obtain
\[ \kappa(H_t) \leq \kappa(G) + \frac{N \varepsilon}{t}. \]
Similarly, we can also prove a lower bound and conclude that
\[ |\kappa(H_t) - \kappa(G)| \lesssim \frac{1}{t}. \]
For the matrix \( H + \lambda G \), we can write the matrix as
\[ H + \lambda G = \sqrt{1 + \lambda^2} \left( \frac{1}{\sqrt{1 + \lambda^2}} H + \frac{\lambda}{\sqrt{1 + \lambda^2}} G \right) = \sqrt{1 + \lambda^2} H_{\log(1+\lambda^2)}. \]
Therefore we have that \( \kappa(H + \lambda G) = \kappa(H_{\log(1+\lambda^2)}) \). As a consequence, we obtain that
\[ |\kappa(H + \lambda G) - \kappa(G)| \lesssim \frac{1}{\log(1+\lambda^2)}, \]
which completes the proof for Theorem 3.

5.2 Quantitative universality

In this section, we prove Theorem 4, which establishes the quantitative universality for the condition number. We will use the universality for both the smallest singular value and the largest singular values. In particular, since the smallest singular values is typically of order \( O(N^{-1}) \), a quantitative control is necessary. The largest singular value is of order constant, and therefore it is easier to deal with. Specifically, for the largest singular value, due to the Tracy-Widom limit for the largest eigenvalue of the sample covariance matrix, we know that \( |\sigma_N(H) - 2| \leq N^{-2/3+\varepsilon} \) for any \( \varepsilon > 0 \) with overwhelming probability.
Let \( \varepsilon > 0 \) be an arbitrarily small constant. For any large \( D > 0 \) and sufficiently large \( N \), we have
\[
P \left( \frac{\kappa(H)}{N} > r \right) = P \left( N \sigma_1(H) < r^{-1} \sigma_N(H) \right) \leq P \left( N \sigma_1(H) < r^{-1} \left( 2 + N^{-\frac{2}{3} + \varepsilon} \right) \right) + N^{-D}
\]
Using Theorem 2, we obtain
\[
P \left( N^{-1} \kappa(H) > r \right) \leq P \left( r N \sigma_1(G) < 2 \right) + N^{-\frac{1}{3} - \varepsilon} + N^{-D}
\]
\[
\leq P \left( r N \sigma_1(G) < \sigma_N(G) + N^{-\frac{4}{3} + \varepsilon} \right) + N^{-\frac{1}{3} - \varepsilon} + N^{-D}
\]
\[
\leq P \left( \frac{\kappa(G)}{N} > r - \frac{N^{\varepsilon/2}}{N^{2/3} \sigma_1(G)} \right) + N^{-\frac{1}{3} - \varepsilon} + N^{-D}
\]
where the third inequality follows from that \( N \sigma_1(G) \) is of order constant with overwhelming probability. Taking large enough \( D \), we have
\[
P \left( \frac{\kappa(H)}{N} > r \right) \leq P \left( \frac{\kappa(G)}{N} > r - N^{-\frac{2}{3} + \varepsilon} \right) + O \left( N^{-\frac{1}{3} - \varepsilon} \right).
\]
This yields the upper bound in (7), and the lower bound can be proved similarly. Hence, we have proved Theorem 4.

6 Beyond Strictly Square Matrices

As mentioned in the Introduction, the optimal local law for an \( M \times N \) random matrix with general \( \lim N/M = 1 \) is a notoriously hard problem. In particular, the optimal rigidity estimates Lemma 3.1 is unknown unless we restrict \( M \equiv N \). In this section, we discuss a slight extension of the strictly-square case. We show that in the regime \( M = N + O(N^{o(1)}) \), all of our theorems still hold.

This claim is based on the following important observation. All proofs of our paper only relies on the local law (as well as its consequences), and therefore it suffices to show that such a local law is still valid for a general \( M \times N \) matrix. More specifically, the main task is to show that the optimal local semicircle law (16) still holds for the Girko symmetrization of an \( M \times N \) random matrix. Modulo the optimal local law, the optimal rigidity will still be valid as a by-product via standard approaches in random matrix theory.

For an \( M \times N \) matrix \( H \), we consider the augmented matrix \( H_A \), which is an \( M \times M \) matrix by adding \( M - N \) columns to \( H \) with independent entries satisfying (1) and (2). Without loss of generality, we may assume that the added columns are the first \( M - N \) columns in \( H_A \). Since \( H_A \) is a square matrix, the local semicircle law (see [AEK14, Theorem 1.1]) is still true. Specifically, for any fixed \( \omega > 0 \), define the spectral domain
\[
S = S_\omega := \{ z = x + iy : |x| \leq \omega^{-1}, M^{-1+\omega} \leq y \leq \omega^{-1} \}.
\]
Then for any \( z = x + iy \in S \), define the resolvent \( G^A(z) := (\tilde{H}_A - z)^{-1} \) and we have
\[
\left| \frac{1}{2M} \text{Tr} G^A(z) - m_{sc}(z) \right| < \frac{1}{M y}.
\]
and
\[
\max_{i,j} |G^A_{ij}(z) - \delta_{ij}m_{sc}(z)| < \left( \frac{1}{My} + \sqrt{\frac{\text{Im} m_{sc}(z)}{My}} \right).
\]

For an \( n \times n \) matrix \( X \) and a subset \( 1 \leq k \leq n \), we define \( X^{(k)} \) as the \((n-k) \times (n-k)\) matrix
\[
X^{(k)} = (X_{ij})_{k+1 \leq i,j \leq n}.
\]

Recall the Girko symmetrization \( \tilde{H} \) of a matrix \( H \). Then we have \( \tilde{H} = \tilde{H}_A^{(M-N)} \).

For \( z = x + iy \) with \( y > 0 \), let \( G^{(k)}(z) := (\tilde{H}_A^{(k)} - z)^{-1} \). In particular we have \( G^{(0)}(z) = G^A(z) \) and \( G^{(M-N)}(z) = G(z) := (\tilde{H} - z)^{-1} \). Then we have the following resolvent identity (see e.g. [BGK17, Lemma 3.5])
\[
G^A_{ij}(z) = G^{(1)}_{ij}(z) + \frac{G^A_{11}(z)G^A_{ij}(z)}{G^A_{11}(z)}, \text{ for all } 2 \leq i, j \leq 2M.
\]

From the local semicircle law for the square matrix \( H_A \), with overwhelming probability we have
\[
|G^A_{1,1}(z)|, |G^A_{1,j}(z)| \leq M^\varepsilon \left( \frac{1}{My} + \sqrt{\frac{\text{Im} m_{sc}(z)}{My}} \right), \text{ and } |G^A_{1,1}(z)| \sim 1.
\]

Using local law for \( G^A_{ij} \) again, this implies that
\[
|G^{(1)}_{ij}(z) - G^A_{ij}(z)| \lesssim \frac{M^{2\varepsilon}}{My}.
\]

More generally, we have
\[
G^{(k)}_{ij}(z) = G^{(k+1)}_{ij}(z) + \frac{G^{(k)}_{i,k+1}(z)G^{(k)}_{k+1,j}(z)}{G^{(k)}_{k+1,k+1}(z)}, \text{ for all } 2 \leq i, j \leq 2M.
\]

By the same arguments as above, for any fixed \( k \), we can derive that
\[
|G^{(k+1)}_{ij}(z) - G^{(k)}_{ij}(z)| \lesssim \frac{M^{2\varepsilon}}{My}.
\]

By a telescoping summation, we derive
\[
|G_{ij}(z) - \delta_{ij}m_{sc}(z)| \lesssim \sum_{k=0}^{M-N-1} |G^{(k+1)}_{ij}(z) - G^{(k)}_{ij}(z)| + |G^A_{ij}(z) - \delta_{ij}m_{sc}(z)|
\]
\[
\lesssim M^\varepsilon \left( \frac{1}{My} + \sqrt{\frac{\text{Im} m_{sc}(z)}{My}} \right) + (M - N) \frac{M^{2\varepsilon}}{My}
\]

Since \( M - N = N^{o(1)} \), the second term can be absorb into the first term with a larger factor \( N^{3\varepsilon} \).

Thanks to the arbitrariness of \( \varepsilon \), we have
\[
|G_{ij}(z) - \delta_{ij}m_{sc}(z)| \lesssim \left( \frac{1}{Ny} + \sqrt{\frac{\text{Im} m_{sc}(z)}{Ny}} \right).
\]
Moreover, the resolvent identity also yields

$$\sum_{i=k+2}^{2M} G^{(k)}_{ii}(z) = \text{Tr} G^{(k+1)}(z) + \frac{1}{G^{(k)}_{k+1,k+1}(z)} \sum_{i=k+2}^{2M} G^{(k)}_{i,k+1}(z) G^{(k)}_{k+1,i}(z).$$

This yields

$$\text{Tr} G^{(k)}(z) - \text{Tr} G^{(k+1)}(z) = \frac{1}{G^{(k)}_{k+1,k+1}(z)} \sum_{i=k+1}^{2M} G^{(k)}_{i,k+1}(z) G^{(k)}_{k+1,i}(z).$$

Using the Ward identity, this implies that

$$\left| \text{Tr} G^{(k)}(z) - \text{Tr} G^{(k+1)}(z) \right| \leq \frac{1}{|G^{(k)}_{k+1,k+1}(z)|} \sum_{i=k+1}^{2M} \left| G^{(k)}_{i,k+1}(z) \right|^2 \leq \frac{1}{|G^{(k)}_{k+1,k+1}(z)|} \frac{\text{Im} G^{(k)}_{k+1,k+1}(z)}{y} \leq \frac{1}{y}.$$

From the local law for $G^A$, with overwhelming probability we have

$$\left| \frac{1}{2M} \text{Tr} G^A(z) - m_{sc}(z) \right| \leq \frac{M^\varepsilon}{M y}.$$

Again, using $M - N = N^{o(1)}$, the telescoping sum yields

$$\left| \frac{1}{M + N} \text{Tr} G(z) - m_{sc}(z) \right| \leq \frac{N^\varepsilon}{N y} + \frac{N^{o(1)}}{N y}.$$

The second term can be absorbed into the first term for any fixed $\varepsilon > 0$. The arbitrariness of $\varepsilon$ concludes that

$$\left| \frac{1}{M + N} \text{Tr} G(z) - m_{sc}(z) \right| < \frac{1}{Ny}. \quad (40)$$

Hence, (39) and (40) have shown that the local law also holds for $H$. The rigidity estimates Lemma 3.1 also follow from a classical argument in random matrix theory (see e.g. [BGK17]). As a consequence, Theorem 1, Theorem 2, Theorem 3 and Theorem 4 are all still valid.

Acknowledgment

The author thanks Paul Bourgade for suggesting this problem and for insightful comments on an early version of this manuscript.

A Auxiliary Results

To make this paper self-contained, we collect some well-known results that are used in the paper.

The first result is about controlling the size of a martingale, which is used in Proposition 3.1 to bound the martingale term in the stochastic dynamics (42), and also in Lemma B.1 to bound (45). This is from [SW09, Appendix B.6, Equation (18)].
Lemma A.1. For any continuous martingale $M$ and any $\lambda, \mu > 0$, we have
\[ \mathbb{P}\left( \sup_{0 \leq u \leq t} |M_u| \geq \lambda, \langle M \rangle_t \leq \mu \right) \leq 2e^{-\frac{\lambda^2}{2\mu}}. \]

The second result is the Helffer-Sjöstrand formula, which is a classical result in functional calculus. This formula is used in Proposition 4.1 to compute the trace of functions via the Stieltjes transform. We are using the version in [EY17, Section 11.2].

Lemma A.2 (Helffer-Sjöstrand formula). Let $f \in C^1(\mathbb{R})$ with compact support and let $\chi(y)$ be a smooth cutoff function with support in $[-1,1]$, with $\chi(y) = 1$ for $|y| \leq \frac{1}{2}$ and with bounded derivatives. Then
\[ f(\lambda) = \frac{1}{\pi} \int_{\mathbb{R}^2} \frac{iy f''(x) \chi(y) + i(f(x) + if'(x))\chi'(y)}{\lambda - x - iy} \, dx \, dy. \]

We also have the following resolvent expansion identity. This is a well-known result in linear algebra, and it is used in Proposition 4.1 to compare the resolvents of two matrices.

Lemma A.3 (Resolvent expansion). For any two matrices $A$ and $B$, we have
\[ (A + B)^{-1} = A^{-1} - (A + B)^{-1}BA^{-1} \]
provided that all the matrix inverses exist.

Finally, we have some estimates for the Stieltjes transform of the semicircle law. For $z = E + i\eta$ with $\eta > 0$, recall that $m_{\text{sc}}(z)$ denotes the Stieltjes transform of the semicircle distribution. The following estimates are well known in random matrix theory (see e.g. [EY17, Lemma 6.2]).

Lemma A.4. We have for all $z = E + i\eta$ with $\eta$ that
\[ |m_{\text{sc}}(z)| = |m_{\text{sc}}(z) + z|^{-1} \leq 1. \]
Furthermore, there is a constant $c > 0$ such that for $E \in [-10, 10]$ and $\eta \in (0, 10]$ we have
\[ c \leq |m_{\text{sc}}(z)| \leq 1 - c\eta, \quad |1 - m_{\text{sc}}^2(z)| \sim \sqrt{\xi(E) + \eta}, \]
as well as
\[ \operatorname{Im} m_{\text{sc}}(z) \sim \begin{cases} \sqrt{\xi(E) + \eta} & \text{if } |E| \leq 2, \\ \frac{e}{\sqrt{\xi(E) + \eta}} & \text{if } |E| \geq 2, \end{cases} \]
where $\xi(E) := |E| - 2$ is the distance of $E$ to the spectral edge.

B Proofs for Smoothed Analysis

B.1 Proof of Proposition 3.1

The proof is essentially the same as [Wan19, Proposition 3.8], and we briefly describe the key steps here for completeness. For any $1 \leq \ell, m \leq N^{10}$, we define $t_{\ell} = \ell N^{-10}$ and $z^{(m)} = E_m + i\eta_m = E_m + iN^{-1+4\varepsilon}\xi(E_m)^{-1/2}$, where $\int_{-\infty}^{E_m} d\rho_{\text{sc}} = mN^{-10}$. Consider the stopping times
\[ \tau_0 = \inf \left\{ 0 \leq t \leq 1 : \exists - N \leq k < N \text{ s.t. } |s_k(t) - \gamma_k| > N^{-\frac{2}{3} + \varepsilon} (N^1 - |k|)^{-\frac{1}{3}} \right\}, \]
\[ \tau_{\ell,m} = \inf \left\{ 0 \leq s \leq t_{\ell} : \operatorname{Im} \tilde{\Omega}_s \left( z^{(m)}_{t_{\ell} - s} \right) > \frac{N^2 e}{2} \frac{\xi(E_m)^{1/2}}{(\xi(E_m)^{1/2} \vee \ell)}, \right\}, \]
\[ \tau = \min \left\{ \tau_0, \tau_{\ell,m} : 0 \leq \ell, m \leq N^{10}, \xi(E_m) > N^{-\frac{2}{3} + 4\varepsilon} \right\}, \]

23
with the convention \( \inf \emptyset = \infty \). We claim that it suffices to show that \( \tau = \infty \) with overwhelming probability.

To prove this claim, for any \( z \in \mathcal{F}_\epsilon \) and \( 0 \leq t \leq 1 \), we pick \( t_\ell \leq t \leq t_{\ell+1} \) and \( |z - z^{(m)}| < N^{-5} \). Note that the maximum principle Lemma 3.2 implies \( |\psi_k(t)| \leq 1 \) for all \( k \) and \( t \). Then we have \( |\widetilde{G}_t(z) - \widetilde{G}_t(z^{(m)})| \lesssim N^{-2} \). Also, note that for \( z = E + i\eta \) we have \( |S_t(z)| \leq \eta^{-1} \), and

\[
\left| \partial_z \widetilde{G}_t(z) \right| \lesssim N \max_k |\psi_k(0)| \eta^{-2} \lesssim N \eta^{-2}, \quad \left| \partial_{zz} \widetilde{G}_t(z) \right| \lesssim N \eta^{-3}.
\]

Consider the events

\[
\mathcal{E}_{\ell,m,k} := \left\{ \sup_{t_\ell \leq u \leq t_{\ell+1}} \left| \int_{t_\ell}^u \frac{e^{-\frac{u}{2}} \psi_k(v)}{(s_k(v) - z^{(m)})^2} dB_k(v) \right| < N^{-4} \right\}.
\]

On the event \( \bigcap_k \mathcal{E}_{\ell,m,k} \), the above estimates imply that \( |\widetilde{G}_t(z^{(m)}) - \widetilde{G}_t(z^{(m)})| < N^{-2} \). It further shows that

\[
|\widetilde{G}_t(z) - \widetilde{G}_t(z^{(m)})| < N^{-2}.
\]

Since this holds for all \( z \) and \( t \), we have shown that

\[
\{ \tau = \infty \} \bigcap \left( \bigcap_{1 \leq \ell \leq N^{10}, -N \leq k \leq N} \mathcal{E}_{\ell,m,k} \right) \subset \bigcap_{t \in \mathcal{A}_\epsilon, 0 \leq t \leq 1} \left\{ \text{Im} \widetilde{G}_t(z) \lesssim N^{2\epsilon} \frac{\xi(E)^{1/2}}{(\xi(E)^{1/2} \lor t)} \right\}.
\]

Moreover, note that

\[
\left\langle \int_{t_\ell}^{t_{\ell+1}} \frac{e^{-\frac{u}{2}} \psi_k(v)}{(s_k(v) - z^{(m)})^2} dB_k(v) \right\rangle_{t_{\ell+1}} \leq \int_{t_\ell}^{t_{\ell+1}} \frac{e^{-v} |\psi_k(v)|^2}{(s_k(v) - z^{(m)})^4} dv \lesssim N^{-10} \left( N^{-1+4\epsilon} \right)^{-4} \left( \max_k |\psi_k(0)| \right)^2 \lesssim N^{-6+16\epsilon}.
\]

Using Lemma A.1, we conclude that the event \( \mathcal{E}_{\ell,m,k} \) happens with overwhelming probability. By a union bound, we further have that \( \bigcap_{\ell,m,k} \mathcal{E}_{\ell,m,k} \) happens with overwhelming probability. Together with the set inclusion (41), we conclude that the claim is true, i.e. it suffices to prove \( \tau = \infty \) with overwhelming probability.

To prove \( \tau = \infty \) with overwhelming probability, consider some fixed \( t = t_\ell \) and \( z = z^{(m)} = E + i\eta \), and define the function \( f_u(z) := \widetilde{G}_u(z_{t-u}) \). By Lemma 3.5, the initial condition is well controlled \( \text{Im} \widetilde{G}_0(z) \lesssim N^{2\epsilon} \frac{\xi(E_m)^{1/2}}{(\xi(E_m)^{1/2} \lor \tau)} \). To bound the increments, note that the dynamics (17) yields

\[
df_{u \land \tau}(z) = \epsilon_u(z_{t-u}) \text{d}(u \land \tau) - \frac{e^{-\frac{u}{2}}}{\sqrt{N}} \sum_{-N \leq k \leq N} \frac{\psi_k(u)}{(z_{t-u} - s_k(u))^2} dB_k(u \land \tau),
\]

where

\[
\epsilon_u(z) := (S_u(z) - m_{\text{sc}}(z)) \partial_z \widetilde{G}_u + \frac{1}{4N} (\partial_{zz} \widetilde{G}_u) + \frac{e^{-\frac{u}{2}}}{2N} \sum_{-N \leq k \leq N} \frac{\psi_k(u)}{(s_k - z)^2(s_k + z)}.
\]
By the local semicircle law (16), we have

\[
\begin{align*}
\sup_{0 \leq s \leq t} \left| \int_0^s (S_u(z_{t-u}) - m_{sc}(z_{t-u})) \partial_z \tilde{E}_u(z_{t-u}) d(u \wedge \tau) \right| &
\leq \int_0^t \frac{N^\varepsilon}{N \text{Im} (z_{t-u})} \sum_{-N \leq k \leq N} |s_k(u) - z_{t-u}|^2 d(u \wedge \tau) \\
&\leq \int_0^t \frac{N^\varepsilon \text{Im} \tilde{E}_u(z_{t-u})}{N (\text{Im} z_{t-u})^2} d(u \wedge \tau) \\
&\leq \int_0^t \frac{N^\varepsilon d\psi_z u}{N (\eta + t - u) \left( \frac{b(z)}{\xi(z)^{1/2}} \right)^2} \left( \frac{\xi(E)^{\frac{1}{2}}}{(\xi(E)^{\frac{1}{2}} \vee t)} \right) \\
&\lesssim \frac{\xi(E)^{\frac{1}{2}}}{(\xi(E)^{\frac{1}{2}} \vee t)}.
\end{align*}
\]

Also, we have

\[
\begin{align*}
\sup_{0 \leq s \leq t} \left| \int_0^s \frac{1}{4N} \partial_{zz} \tilde{E}_u(z_{t-u}) d(u \wedge \tau) \right| &\lesssim \int_0^t \frac{\text{Im} \tilde{E}_u(z_{t-u})}{N (\text{Im} z_{t-u})^2} d(u \wedge \tau) \lesssim N^{-\varepsilon} \frac{\xi(E)^{\frac{1}{2}}}{(\xi(E)^{\frac{1}{2}} \vee t)},
\end{align*}
\]

and

\[
\begin{align*}
\sup_{0 \leq s \leq t} \left| \int_0^s \frac{e^{-\frac{u}{2}}}{2N} \sum_{-N \leq k \leq N} \frac{\psi_k(u)}{(s_k - z_{t-u})^2} d(u \wedge \tau) \right| &\lesssim \frac{1}{N} \int_0^t \frac{1}{\text{Im} (z_{t-u})} \sum_{-N \leq k \leq N} \frac{\psi_k(u)}{|s_k(u) - z_{t-u}|^2} d(u \wedge \tau) \\
&\lesssim \int_0^t \frac{\text{Im} f_u(z_{t-u})}{N (\text{Im} z_{t-u})^2} d(u \wedge \tau) \lesssim N^{-\varepsilon} \frac{\xi(E)^{\frac{1}{2}}}{(\xi(E)^{\frac{1}{2}} \vee t)}.
\end{align*}
\]

For the martingale part

\[
M_k := \int_0^s \frac{e^{-\frac{u}{2}}}{\sqrt{N}} \sum_{-N \leq k \leq N} \frac{\psi_k(u)}{(z_{t-u} - s_k(u))^2} dB_k(u \wedge \tau),
\]

using the rigidity of singular values (Lemma 3.1), with overwhelming probability we have

\[
\sup_{0 \leq s \leq t} |M_k|^2 \lesssim N^{\varepsilon} \int_0^t \frac{1}{N} \sum_{-N \leq k \leq N} \left| \frac{\psi_k(u)}{|z_{t-u} - \gamma_k|^4} \right|^2 d(u \wedge \tau).
\]

To estimate this integral, we chop the interval \([-N, N]\) into \(2N^{1-4\varepsilon}\) subintervals \(I_j = [k_j, k_{j+1}]\) where \(k_j = -N + jN^{4\varepsilon}\). We can bound the summation in the integral in the following way

\[
\frac{1}{N} \sum_{-N \leq k \leq N} \left| \frac{\psi_k(u)}{|z_{t-u} - \gamma_k|^4} \right|^2 \leq \frac{1}{N} \sum_{0 \leq j \leq 2N^{1-4\varepsilon}} \left( \max_{k \in I_j} \psi_k(u) \right) \left( \max_{k \in I_j} \frac{1}{|z_{t-u} - \gamma_k|^4} \right) \left( \sum_{k \in I_j} \psi_k(u) \right).
\]

Using similar discretization arguments as above, we can derive

\[
\max_{k \in I_j} \psi_k(u) \leq \sum_{k \in I_j} \psi_k(u) \leq \frac{N^{6\varepsilon}}{N (\xi(\gamma_k)^{1/2} \vee u)},
\]

\[
\int_0^t \frac{1}{N} \sum_{-N \leq k \leq N} \left| \frac{\psi_k(u)}{|z_{t-u} - \gamma_k|^4} \right|^2 d(u \wedge \tau) \lesssim \frac{N^{8\varepsilon}}{N (\xi(\gamma_k)^{1/2} \vee u)}.
\]
is a delicate task. The key part of the proof is a careful analysis of the estimation.

An observation is that

\[ \text{B.2 Proof of Lemma 3.10} \]

The proof of Lemma 3.10 is a delicate task. The main part of the proof is a careful analysis of the dynamics. The main idea is to approximate the dynamics with a short-range version, which will be easier to control. To do this, we show the finite speed of propagation estimate for the short-range kernel of the parabolic-type equation (14) satisfied by \( \varphi_k \). Then we prove a short-range approximation of the original dynamics and introduce a regularized equation. Finally, we show that, with a well-behaved initial condition, the regularized equation gives us the desired good approximation.

To begin with, the core input of the bootstrap argument is the following technical lemma, which states that the estimate of the local average will improve along with the induction hypothesis \( H_0 \).

**Lemma B.1.** Assume \( H_0 \). Let \( \xi > 0 \) be any fixed small constant. For any \( 0 < t < 1 \), any \( \varepsilon_0 > 0 \) arbitrarily small and \( z = E + i\eta \) satisfying \( N^{-1 + \varepsilon_0} < \eta < 1 \), \( |E| < 2 - \xi \), we have

\[
\left| \text{Im} \mathcal{G}_t(z) - e^{-\frac{t}{2}} \frac{\text{Im} S_t(z)}{\text{Im} S_0(z_t)} \text{Im} \mathcal{G}_0(z_t) \right| \lesssim \left( \frac{(Nt)^{\alpha}}{N^2 \eta} + \frac{1}{Nt} \right)
\]

**Proof.** Fix \( t \) and consider the function

\[
g_u(z) := \mathcal{G}_u(z_{t-u}) - e^{-\frac{t}{2}} \frac{\text{Im} \mathcal{G}_0(z_t)}{\text{Im} S_0(z_t)} S_u(z_{t-u}), \quad 0 \leq u \leq t.
\]

An observation is that \( e^{-u/2} S_u(z) \) satisfy the same stochastic advection equation (17) with \( \varphi_k \) replaced by \( \frac{1}{2N} \). Therefore, we have

\[
dg_u = \left( S_u(z_{t-u}) - m_{sc}(z_{t-u}) \right) \left( \partial_z \mathcal{G}_u(z_{t-u}) - e^{-\frac{t}{2}} \frac{\text{Im} \mathcal{G}_0(z_t)}{\text{Im} S_0(z_t)} \partial_z S_u(z_{t-u}) \right) du
\]

\[ + \frac{1}{4N} \left( \partial_z \mathcal{G}_u(z_{t-u}) - e^{-\frac{t}{2}} \frac{\text{Im} \mathcal{G}_0(z_t)}{\text{Im} S_0(z_t)} \partial_{zz} S_u(z_{t-u}) \right) du
\]

\[ + \frac{e^{-\frac{t}{2}}}{2N} \sum_{-N \leq k \leq N} \theta_k(u)\left( s_k(u) - z_{t-u} \right)^2 dK_u
\]

\[ - \frac{e^{-\frac{t}{2}}}{\sqrt{N}} \sum_{-N \leq k \leq N} \theta_k(u)\left( s_k(u) - z_{t-u} \right)^2 dB_k.
\]
where
\[ \theta_k(u) = \varphi_k(u) - \frac{\text{Im} \mathcal{G}_0(z_t)}{2\text{Im} S_0(z_t)}. \]

Similarly as in the proof of Proposition 3.1, for \( \varepsilon > 0 \) and \( 0 \leq \ell, m, p \leq N^{10} \), define \( t_\ell = \ell N^{-10} \) and \( z^{(m,p)} = E_m + i\eta_p \) where \( f^{E_m}_{\ell} \text{d}\rho = mN^{-10} \) and \( \eta_p = N^{-1+\varepsilon} + pN^{-10} \). We also pick \( c > 0 \) such that \( |(1-c)N| = \arg\min_k \gamma_k - (2 - \frac{\xi}{2}) \). Assuming \( H_\alpha \), let \( \varepsilon_0 > 0 \) be the arbitrarily small scale in the hypothesis. Let \( C > 0 \) be some suitably large constant. Recall the stopping times
\[
\tau_0 = \inf \left\{ 0 \leq u \leq 1 : \exists - N \leq k \leq N \text{ s.t. } |s_k(u) - \gamma_k| > N^{-\frac{2}{3}+\varepsilon}(N + 1 - |k|)^{-\frac{1}{2}} \right\},
\]
\[
\tau_1 = \inf \left\{ N^{-1+\varepsilon_0} \leq u \leq 1 : \exists - N \leq k \leq N \text{ s.t. } |\varphi_k(u)| > \frac{NC_\varepsilon}{N} \left( \frac{1}{(N+1-|k|)^{1/3} \vee u} \right) \right\},
\]
and consider the new stopping times
\[
\tau_2 = \inf \left\{ N^{-1+\varepsilon_0} \leq u \leq 1 : \exists k \in [(c-1)N, (1-c)N] \text{ s.t. } |\varphi_k(u) - \tilde{\varphi}_k(u)| > N^{C_\varepsilon} \left( \frac{(Nt)\alpha}{N^2 t\eta} + \frac{1}{N^t} \right) \right\},
\]
\[
\tau_{\ell,m,p} = \inf \left\{ 0 \leq u \leq t_\ell : \left| \text{Im} g_u^{(t_\ell)}(z^{(m,p)}) \right| > N^{C_\varepsilon} \left( \frac{(Nt)\alpha}{N^2 t\eta} + \frac{1}{N^t} \right) \right\}
\]
\[
\tau = \min\{\tau_0, \tau_1, \tau_2, \tau_{\ell,m,p} : 0 \leq \ell, m, p \leq N^{10}, |E_m| < 2 - \xi\}.
\]

Recall the convention \( \inf \emptyset = \infty \). As shown in the proof of Proposition 3.1, it suffices to show that \( \tau = \infty \) with overwhelming probability.

A key ingredient for the analysis of the dynamics of \( g_u \) is the following estimates on \( \theta_k(u) \). To do this, we fix some \( t = t_\ell \) and \( z = z^{(m,p)} \) with \( |E_m| < 2 - \xi \), and let \( N^{-1+\varepsilon_0} \leq u \leq t \land \tau \) and \( k \in [(c-1)N, (1-c)N] \).

On the one hand, we have a direct a priori estimate. Since \( u \leq \tau_1 \), we have
\[ |\varphi_k(u)| \lesssim N^{-\frac{2}{3}+C_\varepsilon} (N + 1 - |k|)^{-\frac{1}{2}}. \]

Moreover, note that for \( z = E + i\eta \) with \( E \) in the bulk and \( t < 1 \), uniformly we have \( \text{Im} S_0(z_t) \gtrsim 1 \). By Lemma 3.5, this shows
\[ \frac{\text{Im} \mathcal{G}_0(z_t)}{2\text{Im} S_0(z_t)} \lesssim N^{-1+\varepsilon} \lesssim N^{-\frac{2}{3}+C_\varepsilon} (N + 1 - |k|)^{-\frac{1}{2}}. \]

As a consequence, we have
\[ |\theta_k(u)| \lesssim N^{-\frac{2}{3}+C_\varepsilon} (N + 1 - |k|)^{-\frac{1}{2}}. \quad (46) \]

On the other hand, the estimate can also be obtained via approximation
\[ |\theta_k(u)| \lesssim |\varphi_k(u) - \tilde{\varphi}_k(u)| + |\tilde{\varphi}_k(u) - \tilde{\varphi}_j(t)| + |\tilde{\varphi}_j(t) - \frac{\text{Im} f_0(z_t)}{2\text{Im} S_0(z_t)}|. \]

For the first term, since \( u \leq \tau_2 \) we have
\[ |\varphi_k(u) - \tilde{\varphi}_k(u)| \lesssim N^{C_\varepsilon} \frac{(Nu)^q}{N^2 u}. \]
Choosing $|\gamma_j - E| \leq N^{-1+2\varepsilon}$, the remaining two terms are controlled by Lemma 3.8

$$|\hat{\varphi}_k(u) - \hat{\varphi}_j(t)| + \left| \hat{\varphi}_j(t) - \frac{\text{Im} f_0(z_t)}{2N \text{Im} S_0(z_t)} \right| \lesssim N^\varepsilon \left( \frac{|k - j|}{N^2 u} + \frac{|t - u|}{N u} + \frac{|E - \gamma_j|}{N t} + \frac{\eta}{N t} \right) \lesssim N^{C\varepsilon} \left( \frac{|\gamma_k - E|}{N u} + \frac{t - u + \eta}{N u} + \frac{1}{N t} \right).$$

Together, we decompose the error terms into two parts and obtain the following

$$|\theta_k(u)| \leq N^{C\varepsilon} \left( \frac{|\gamma_k - E|}{N u} + \frac{t - u + \eta}{N u} + \frac{(Nu)^\alpha}{N^2 u} \right) =: \varphi^G \left( \frac{|\gamma_k - E|}{N u} + \Lambda(a, N, t, \eta, u) \right) \quad (47)$$

With the above control on $\theta_k(u)$, the dynamics (45) can be used to bound $\text{Im} (g_t - g_0)$ similarly as in Proposition 3.1. For the first term, we have

$$\int_0^{t_{\tau\tau}} \frac{N^{C\varepsilon}}{N \text{Im} (z_{t-u})} \sum_{|k| \leq N} |\theta_k(u)| (s_k(u) - z_{t-u})^2 \text{du} \leq \int_0^{t_{\tau\tau}} \frac{N^{C\varepsilon}}{N \text{Im} (z_{t-u})} \left( \sum_{|k| \geq (1-c)N} \frac{|\theta_k(u)|}{\gamma_k - z_{t-u}} + \sum_{|k| < (1-c)N} \frac{|\theta_k(u)|}{\gamma_k - z_{t-u}} \right) \text{du} \leq: I_1 + I_2.$$  

For the soft edge part $|k| \geq (1-c)N$, using (46) we obtain

$$I_1 \leq N^{C\varepsilon} \int_0^{t_{\tau\tau}} \frac{1}{N \text{Im} (z_{t-u})} \sum_{|k| \geq (1-c)N} N^{-\frac{2}{3}+C\varepsilon} (N + 1 - |k|)^{-\frac{1}{3}} \text{du} \leq N^{C\varepsilon} \log(1 + N\tau) \frac{1}{N}. \quad (49)$$

For $I_2$, note that

$$\sum_{|k| < (1-c)N} \frac{|\theta_k(u)|}{\gamma_k - z_{t-u}} \leq N^{C\varepsilon} \left( \frac{1}{N u} \sum_{|k| < (1-c)N} \frac{|\gamma_k - E|}{\gamma_k - z_{t-u}} + \Lambda \sum_{|k| < (1-c)N} \frac{1}{\gamma_k - z_{t-u}} \right) \leq N^{C\varepsilon} \left( \frac{1}{u} + N \frac{\Lambda}{\eta + t - u} \right)$$

This yields

$$\int_{N^{-1+\varepsilon_0}}^{t_{\tau\tau}} \frac{N^{C\varepsilon}}{N \text{Im} (z_{t-u})} \sum_{|k| < (1-c)N} \frac{|\theta_k(u)|}{\gamma_k - z_{t-u}}^2 \text{du} \leq N^{C\varepsilon} \int_{N^{-1+\varepsilon_0}}^{t_{\tau\tau}} \frac{1}{N(\eta + t - u)} \left( \frac{1}{u} + N \frac{\Lambda}{\eta + t - u} \right) \text{du} \leq N^{C\varepsilon} \int_{N^{-1+\varepsilon_0}}^{t_{\tau\tau}} \frac{1}{N(\eta + t - u)} \left( \frac{1}{u} + N \frac{1}{\eta + t - u} \left( \frac{t - u + \eta}{N u} + \frac{1}{N t} + \frac{(Nu)^\alpha}{N^2 u} \right) \right) \text{du} \leq N^{C\varepsilon} \left( \frac{(Nu)^\alpha}{N^2 t u} + \frac{1}{N t} \right). \quad (50)$$
Moreover, without loss of generality, we may assume that \( \varepsilon_0 \sim \varepsilon \). Then, using (46) we obtain
\[
\int_0^{N^{-1+\varepsilon_0}} \frac{N^{C\varepsilon}}{\Im (z_{t-u})} \sum_{|k|<(1-\alpha)N} \left| \frac{\theta_k(u)}{|\gamma_k - z_{t-u}|^2} \right| du \lesssim \frac{N^{C\varepsilon}}{N^2(\eta + t)^2}.
\]  
(51)

Together with previous estimates, this shows
\[
\int_0^{\Upsilon \wedge \tau} |S_u(z_{t-u}) - m_{\mathcal{S}c}(z_{t-u})| \left| \partial_z^2 \mathcal{S}u(z_{t-u}) - e^{-\frac{i}{2}} \frac{\Im \mathcal{S}0(z_t)}{\Im S_0(z_t)} \partial_z^2 S_u(z_{t-u}) \right| du \lesssim N^{C\varepsilon} \left( \frac{(N\bar{\eta})^\alpha}{N^2t\eta} + \frac{1}{Nt} \right),
\]  
(52)

Similarly, we have
\[
\int_0^{\Upsilon \wedge \tau} \frac{1}{4N} \left| \partial_{zz} \mathcal{S}u(z_{t-u}) - e^{-\frac{i}{2}} \frac{\Im \mathcal{S}0(z_t)}{\Im S_0(z_t)} \partial_{zz} S_u(z_{t-u}) \right| du \lesssim N^{C\varepsilon} \left( \frac{(N\bar{\eta})^\alpha}{N^2t\eta} + \frac{1}{Nt} \right)
\]  
and
\[
\int_0^{\Upsilon \wedge \tau} \left| e^{-\frac{i}{2}} \frac{\theta_k(u)}{2N} \sum_{-N \leq k \leq N} \frac{\theta_k(u)}{(s_k(u) - z_{t-u})^2(s_k(u) + z_{t-u})} \right| du \lesssim N^{C\varepsilon} \left( \frac{(N\bar{\eta})^\alpha}{N^2t\eta} + \frac{1}{Nt} \right)
\]
It suffices to bound the martingale term
\[
M_s := \int_0^s e^{-\frac{i}{2}} \sum_{-N \leq k \leq N} \frac{\theta_k(u)}{(s_k(u) - z_{t-u})^2} dB_k.
\]

Again we decompose the integral into two parts
\[
\langle M \rangle_{t \wedge \tau} \lesssim \frac{1}{N} \int_0^{\Upsilon \wedge \tau} \sum_{|k| \geq (1-c)N} \left| \frac{\theta_k(u)}{|\gamma_k - z_{t-u}|^2} \right|^2 du + \frac{1}{N} \int_0^{\Upsilon \wedge \tau} \sum_{|k|<(1-c)N} \left| \frac{\theta_k(u)}{|\gamma_k - z_{t-u}|^2} \right| du
\]
\[=: J_1 + J_2.
\]

The contribution from the soft edge is easy to control
\[
J_1 \lesssim \frac{N^{C\varepsilon}}{N} \int_0^{\Upsilon \wedge \tau} \sum_{|k| \geq (1-c)N} \left( N^{-\frac{2c}{3}} (N + 1 - |k|)^{-\frac{1}{3}} \right)^2 du \lesssim N^{C\varepsilon} \frac{t}{N^2}
\]

For the other term, we use both (46) and (47)
\[
J_2 \lesssim \frac{N^{C\varepsilon}}{N} \int_0^{\Upsilon \wedge \tau} \frac{1}{N^2} \sum_{|k|<(1-c)N} \left[ \frac{1}{|\gamma_k - z_{t-u}|^4} \right] du
\]
\[+ \frac{N^{C\varepsilon}}{N} \int_0^{\Upsilon \wedge \tau} \sum_{|k|<(1-c)N} \left[ \frac{1}{|\gamma_k - z_{t-u}|^4} \left( \frac{|\gamma_k - E|^2}{N^2u^2} + \Lambda^2 \right)^{\frac{1}{2}} \right] du.
\]

Note that
\[
\frac{N^{C\varepsilon}}{N} \int_0^{\Upsilon \wedge \tau} \frac{1}{N^2} \sum_{|k|<(1-c)N} \left[ \frac{1}{|\gamma_k - z_{t-u}|^4} \right] du \lesssim \frac{N^{C\varepsilon}}{N} \int_0^{\Upsilon \wedge \tau} \frac{1}{N(\eta + t - u)^3} du \lesssim \frac{N^{C\varepsilon}}{N^2t^2}.
\]

29
For the other term, without loss of generality we may assume $\eta < t$. For the $\frac{|\gamma_k - E|^2}{N^2u^2}$ term, we have

$$\frac{NC\varepsilon}{N} \int_\eta^t \sum_{|k| < (1-c)N} \frac{1}{|\gamma_k - z_{t-u}|} \left( \frac{|\gamma_k - E|^2}{N^2u^2} \wedge \frac{1}{N^2} \right) du \leq \frac{NC\varepsilon}{N^2t^2} + \frac{NC\varepsilon}{N} \int_\eta^t \sum_{|k|: |\gamma_k - E| \leq u} \frac{1}{|\gamma_k - z_{t-u}|} \frac{|\gamma_k - E|^2}{N^2u^2} du \leq \frac{NC\varepsilon}{N^2t^2} + \frac{NC\varepsilon}{N^2t^2} \int_\eta^t \frac{1}{N^2u^2} \left( \int_{-u}^u x^2 + (\eta + t - u) x \right) dx \leq \frac{NC\varepsilon}{N^2t^2}.$$

For the contribution of $\Lambda(a,N,t,\eta,u)$, we have

$$\frac{NC\varepsilon}{N} \int_\eta^t \sum_{|k| < (1-c)N} \frac{1}{|\gamma_k - z_{t-u}|} \left( \Lambda^2 \wedge \frac{1}{N^2} \right) du \leq \frac{NC\varepsilon}{N} \int_\eta^t \sum_{|k| < (1-c)N} \frac{1}{|\gamma_k - z_{t-u}|} \left[ \left( \frac{(t-u)^2}{N^2u^2} + \frac{(Nu)^{2\alpha}}{N^4u^2} + \frac{\eta^2}{N^2t^2} \right) \wedge \frac{1}{N^2} \right] du \leq \frac{NC\varepsilon}{N} \int_\eta^t \frac{1}{(\eta + t - u)^3} \left[ \frac{(Nu)^{2\alpha}}{N^4u^2} + \frac{\eta^2}{N^2t^2} + \left( \frac{(t-u)^2}{N^2u^2} \wedge \frac{1}{N^2} \right) \right] du.$$

The first two terms in the bracket give us

$$\int_\eta^t \frac{1}{(\eta + t - u)^3} \left( \frac{(Nu)^{2\alpha}}{N^4u^2} + \frac{\eta^2}{N^2t^2} \right) du \leq \frac{(Nt)^{2\alpha}}{N^4t^2\eta^2} + \frac{1}{N^2t^2}.$$

For the remaining term, we have

$$\int_\eta^t \frac{1}{(\eta + t - u)^3} \left( \frac{(t-u)^2}{N^2u^2} \wedge \frac{1}{N^2} \right) du \leq \int_\eta^t \frac{1}{(\eta + t - u)^2} du + \int_\frac{t}{2}^t \frac{1}{(\eta + t - u)^3} \frac{(t-u)^2}{N^2u^2} du \leq \frac{\log N}{N^2t^2}.$$

Combining these results shows

$$\langle M \rangle \leq \frac{(Nt)^{2\alpha}}{N^4t^2\eta^2} + \frac{1}{N^2t^2}.$$

Using Lemma A.1 and a union bound, for any fixed large $D > 0$ and sufficiently large $N > N_0(\varepsilon,D)$, we have

$$\mathbb{P} \left( \sup_{\ell,m,p,0 \leq s \leq t, |E_m| < 2 - \xi} |M_s| \leq \frac{(Nt)^{\alpha}}{N^2t\eta} + \frac{1}{Nt} \right) > 1 - N^{-D}.$$

Together with previous estimates, with overwhelming probability we have

$$\sup_{\ell,m,p,0 \leq s \leq \tau, |E_m| < 2 - \xi} |g_s(z)| \leq \frac{(Nt)^{\alpha}}{N^2t\eta} + \frac{1}{Nt}.$$

This implies $\min_{\ell,m,p} \{ \tau_{\ell,m,p} \} = \infty$ with overwhelming probability. Moreover, we have shown in Lemma 3.1, Lemma 3.6 that $\tau_0 = \infty$ and $\tau_1 = \infty$ with overwhelming probability. Assuming the hypothesis $H_\alpha$, we also have $\tau_2 = \infty$ with overwhelming probability. These imply that $\tau = \infty$ with overwhelming probability. Hence we complete the proof. \qed
Now we move on to the short-range approximation of the dynamics. Recall that \( \{ \varphi_k \} \) satisfies the parabolic equation (14), and we rewrite it as

\[
\frac{d}{dt} \varphi_k = (\mathcal{P} \varphi)_k
\]

where the time-dependent operator \( \mathcal{P} \) is defined in the following way: For \( f : \mathbb{R} \to \mathbb{R}^{2N} \),

\[
(\mathcal{P}f)_k := \sum_{j \neq \pm k} c_{jk}(t)(f_j(t) - f_k(t)), \quad c_{jk}(t) := \frac{1}{2N(s_j(t) - s_k(t))^2}.
\]

Consider some parameter \( l = l(N, \alpha) \) which will be determined later, we decompose the operator \( \mathcal{P} \) into two parts \( \mathcal{P} = \mathcal{P}_{\text{short}} + \mathcal{P}_{\text{long}} \). The operators \( \mathcal{P}_{\text{short}} \) and \( \mathcal{P}_{\text{long}} \) represent the short-range interactions and long-range interactions respectively and are defined as follows

\[
(\mathcal{P}_{\text{short}}f)_k = \sum_{|j-k| \leq l} c_{jk}(t)(f_j(t) - f_k(t)),
\]

\[
(\mathcal{P}_{\text{long}}f)_k = \sum_{|j-k| > l} c_{jk}(t)(f_j(t) - f_k(t)).
\]

Note that the operators \( \mathcal{P}_{\text{short}}, \mathcal{P}_{\text{long}} \) are also time dependent. Let \( \mathcal{F}_{\text{short}}(s,t) \) denote the semigroup associated with the operator \( \mathcal{P}_{\text{short}} \) in the sense

\[
\partial_t \mathcal{F}_{\text{short}}(s,t) = \mathcal{P}_{\text{short}}(t) \mathcal{F}_{\text{short}}(s,t), \quad \mathcal{F}_{\text{short}}(s,s) = \text{Id}.
\]

Also, let \( \mathcal{F} \) denote the semigroup associated with \( \mathcal{P} \).

To prove the short-range approximation, we need the following finite speed of propagation estimate for the semigroup. Such estimates were proved in [CL19] with minor changes.

**Lemma B.2.** For any fixed small \( c > 0 \) and large \( D > 0 \), there exists \( N_0(c,D) \) such that the following holds with probability at least \( 1 - N^{-D} \). For any \( \varepsilon > 0 \), \( N > N_0 \), \( 0 < u < v < 1 \), \( l \geq N|u - v| \), \( |k| \leq (1 - c)N \) and \( -N \leq j \leq N \) such that \( |k - j| > N^\varepsilon l \), we have

\[
(\mathcal{F}_{\text{short}}(u,v)\delta_k)(j) < N^{-D}.
\]

With Lemma B.2, we have the following short-range approximation estimate. In particular, this short-range approximation can be improved based on the hypothesis \( \mathcal{H}_\alpha \).

**Lemma B.3.** Assume \( \mathcal{H}_\alpha \). Let \( c > 0 \) be any fixed small constant. There exists a constant \( C > 0 \) such that for any \( \varepsilon > 0 \), \( N^{-1+C\varepsilon} < t < 1, \frac{1}{2} \leq u < v \leq t, l \geq N^\varepsilon \) and \( |k| \leq (1 - c)N \), we have

\[
|((\mathcal{F}(u,v) - \mathcal{F}_{\text{short}}(u,v)) \varphi(u))_k| \prec (v-u) \left( \frac{N(Nt)^\alpha}{l} \frac{1}{N^2t} + \frac{1}{Nt} \right).
\]

**Proof.** The Duhamel’s principle implies

\[
((\mathcal{F}(u,v) - \mathcal{F}_{\text{short}}(u,v)) \varphi(u))_k = \int_u^v (\mathcal{F}_{\text{short}}(s,v)\mathcal{P}_{\text{long}} \varphi)(s))_k ds.
\]

On the event that Lemma B.2 holds, for \( |k| \leq (1 - 3c)N \), the finite speed of propagation yields

\[
(\mathcal{F}_{\text{short}}(s,v)\mathcal{P}_{\text{long}} \varphi)(s))_k = (\mathcal{F}_{\text{short}}(s,v)\mathcal{P}_{\text{long}} \varphi \mathbb{1}_{\{2c-1\}N, (1-2c)N}\)(s))_k + N^{-D},
\]

31
where \((\varphi_1 \mathbb{1}_{(2c-1)N,(1-2c)N})_j = \varphi_j \mathbb{1}_{(2c-1)N,(1-2c)N}) (j)\). Moreover, using the property that \(\mathcal{T}_{\text{short}}\) is an \(L^\infty\) contraction, we have
\[
|((\mathcal{T}(u, v) - \mathcal{T}_{\text{short}}(u, v)) \varphi(u))_k| \leq |u - v| \sup_{|j| < (1 - 2c)N, u < s < v} \left| (P_{\text{long}} \varphi)_j (s) \right| + N^{-D}. \tag{55}
\]
For \(|i| \leq (1 - c)N\), assuming \(\mathcal{H}_\alpha\) and on the event that Lemma 3.7 holds, there exists a constant \(C > 0\) so that
\[
|\varphi_i(s) - \varphi_j(s)| \leq |\varphi_i(s) - \hat{\varphi}_i(s)| + |\varphi_j(s) - \hat{\varphi}_j(s)| + |\hat{\varphi}_i(s) - \hat{\varphi}_j(s)| \lesssim N^{C\varepsilon} \left( \frac{(Nt)^\alpha}{N^2 t} + \frac{|i - j|}{N^2 t} \right).
\]
For \(|i| > (1 - c)N\), Lemma 3.6 yields
\[
|\varphi_i(s) - \varphi_j(s)| \leq |\varphi_i(s)| + |\varphi_j(s)| \lesssim N^{-\frac{\alpha}{2} + C\varepsilon} (N + 1 - |i|)^{\frac{1}{2}}.
\]
Therefore, using rigidity of singular values, we have
\[
(P_{\text{long}} \varphi)_j (s) = \sum_{|i-j| > |j|} \frac{\varphi_i(s) - \varphi_j(s)}{2N(s_i(s) - s_j(s))^2}
\lesssim N^{1+C\varepsilon} \sum_{|i-j| > |j|} \frac{1}{(i - j)^2} \left( \frac{(Nt)^\alpha}{N^2 t} + \frac{|i - j|}{N^2 t} \right) + N^{-1+C\varepsilon} \sum_{|i| > (1 - c)N} N^{-\frac{\alpha}{2}} (N + 1 - |i|)^{\frac{1}{2}}
\lesssim N^{C\varepsilon} \left( \frac{N}{T} \frac{(Nt)^\alpha}{N^2 t} + \frac{1}{Nt} \right).
\]
Combined with (55), this implies the desired claim with \(|k| \leq (1 - 3c)N\). Note that \(c > 0\) is arbitrary, and thus it concludes the proof. \(\square\)

Further, we will show that we have nice control for a regularization of the short-range dynamics with a well-behaved initial data. To do this, we follow the techniques developed in [BY17]. Consider some fixed times \(u < t\), the short-range parameter \(l\), and define an averaging space window scale \(r\). Throughout the remaining parts of this section, for any fixed arbitrarily small \(\varepsilon > 0\), we make the following assumption on these parameters
\[
N^{30\varepsilon} (t - u) < N^{20\varepsilon} \frac{1}{N} < N^{10\varepsilon} r < t. \tag{56}
\]
For a fixed index \(k\), as in [BY17], we define the flattening operator with parameter \(a > 0\) by
\[
(F_\alpha f)_j (v) := \begin{cases} f_j(v) & \text{if } |j - k| \leq a \\ \hat{\varphi}_k(t) & \text{if } |j - k| > a \end{cases} \quad \text{for } u \leq v \leq t,
\]
and the averaging operator
\[
(Af)_j := \frac{1}{|[Nr, 2Nr]|} \sum_{a \in [Nr, 2Nr]} (F_\alpha f)_j.
\]
As shown in [BY17, Equation (7.4)], the averaging operator can also be represented as a combination of Lipschitz function, i.e. there exists a Lipschitz function \(h\) with \(|h_i - h_j| \leq \frac{|i - j|}{N r}\) such that
\[
(Af)_j = h_j f_j + (1 - h_j) \hat{\varphi}_k(t). \tag{57}
\]
Finally, for \( u < v < t \), consider the regularized dynamics
\[
\begin{align*}
\frac{d}{dv} \Gamma_j(v) &= (\mathcal{P}_{\text{short}}(v) \Gamma)_j, \\
\Gamma(u) &= \mathcal{A} \varphi(u)
\end{align*}
\]

The following lemma shows that the averaging the regularized dynamics gives good approximation for \( \hat{\varphi}_k \).

**Lemma B.4.** Assume \( \mathcal{H}_\alpha \). Let \( c > 0 \) be any fixed small constant. There exists a constant \( C > 0 \) such that for any \( \epsilon > 0 \), \( N^{-1+c\epsilon} < \eta, t < 1, \frac{T}{2} \leq u < v \leq t, l > N^\epsilon \), \( j,k \in [(2c-1)N, (1 - 2c)N] \) such that \( |\gamma_j - \gamma_k| < 10r \), and \( z = \gamma_j + i\eta \), we have
\[
\left| \frac{1}{2N} \text{Im} \sum_{[i,j] < t} \frac{\Gamma_i(v)}{s_i(v) - z} - \left( \frac{1}{2N} \text{Im} \sum_{[i,j] < t} \frac{1}{s_i(v) - z} \right) \hat{\varphi}_k(v) \right| < \left( \frac{r}{Nt} + \eta \frac{Nt}{N^2} + \left( \frac{l}{Nt} + \frac{N(t-u)}{l} + \frac{1}{N\eta} \right) \right) . \tag{58}
\]

**Proof.** We decompose the upper line of (58) into three parts
\[
\frac{1}{2N} \text{Im} \sum_{[i,j] < t} \frac{\Gamma_i(v)}{s_i(v) - z} - \left( \frac{1}{2N} \text{Im} \sum_{[i,j] < t} \frac{1}{s_i(v) - z} \right) \hat{\varphi}_k(v) =: I_1 + I_2 + I_3,
\]
where
\[
I_1 = \frac{1}{2N} \text{Im} \sum_{[i,j] < t} \frac{(\mathcal{F}_{\text{short}}(u,v) \mathcal{A} \varphi(u) - A \mathcal{F}_{\text{short}}(u,v) \varphi(u))_i}{s_i(v) - z},
\]
\[
I_2 = \frac{1}{2N} \text{Im} \sum_{[i,j] < t} \frac{(A \mathcal{F}_{\text{short}}(u,v) \varphi(u) - A \mathcal{F}(u,v) \varphi(u))_i}{s_i(v) - z},
\]
\[
I_3 = \frac{1}{2N} \text{Im} \sum_{[i,j] < t} \frac{(A \mathcal{F}(u,v) \varphi(u))_i - \hat{\varphi}_k(v)}{s_i(v) - z}.
\]

For the first term \( I_1 \), Note that
\[
(\mathcal{F}_{\text{short}}(u,v) \mathcal{A} \varphi(u) - A \mathcal{F}_{\text{short}}(u,v) \varphi(u))_i = \frac{1}{\lceil [Nr, 2Nr] \rceil} \sum_{a \in [Nr, 2Nr]} (\mathcal{F}_{\text{short}}(u,v) \mathcal{F}_a \varphi(u) - \mathcal{F}_a \mathcal{F}_{\text{short}}(u,v) \varphi(u))_i.
\]

When \( |i - k| < a - N^\epsilon l \), we have
\[
(\mathcal{F}_a \mathcal{F}_{\text{short}}(u,v) \varphi(u))_i = (\mathcal{F}_{\text{short}}(u,v) \varphi(u))_i,
\]
and the finite speed of propagation (53) yields
\[
(\mathcal{F}_{\text{short}}(u,v) \mathcal{F}_a \varphi(u))_i = (\mathcal{F}_{\text{short}}(u,v) \varphi(u))_i + N^{-D}.
\]
This gives
\[
(\mathcal{F}_{\text{short}}(u,v) \mathcal{F}_a \varphi(u) - \mathcal{F}_a \mathcal{F}_{\text{short}}(u,v) \varphi(u))_i < N^{-D}
\]
Similarly, this bound also holds in the case $|i-k| > a + N^\varepsilon l$. Now suppose $a - N^\varepsilon l \leq |i-k| \leq a + N^\varepsilon l$. Applying (23) and Hypothesis $\mathcal{H}_\alpha$, we obtain

$$
|\langle \mathcal{I}_{\text{short}}(u,v) \mathcal{F}_a \varphi(u) - \mathcal{F}_a \mathcal{I}_{\text{short}}(u,v) \varphi(u) \rangle_i| 
\leq \max_{m: |m-k| - |a| \leq 2N^\varepsilon t} |\varphi_m(v) - \bar{\varphi}_k(t)| + N^{-D}
\leq \max_{m: |m-k| - |a| \leq 2N^\varepsilon t} |\varphi_m(v) - \bar{\varphi}_m(v)| + \max_{m: |m-k| - |a| \leq 2N^\varepsilon t} |\bar{\varphi}_m(v) - \bar{\varphi}_k(t)| + N^{-D}
\leq N^C \varepsilon \left( \frac{(Nt)^{\alpha}}{N^2 t} + \frac{r + (t-u)}{N t} \right).
$$

Combined with the estimate above, this implies

$$
I_1 \leq N^C \varepsilon \frac{l}{N^r} \left( \frac{(Nt)^{\alpha}}{N^2 t} + \frac{r + (t-u)}{N t} \right). 
$$

(59)

For the term $I_2$, note that $|i-j| < l$ implies $|i| \leq (1-c)N$. Therefore, using the Lipschitz representation of the averaging operator (57), the short-range approximation (54) gives us

$$
|\langle \mathcal{A} \mathcal{I}_{\text{short}}(u,v) \varphi(u) - \mathcal{A} \mathcal{F}(u,v) \varphi(u) \rangle_i| 
\leq |\langle \mathcal{I}_{\text{short}}(u,v) \varphi(u) - \mathcal{F}(u,v) \varphi(u) \rangle_i| \leq N^C \varepsilon (t-u) \left( \frac{N (Nt)^{\alpha}}{l N^2 t} + \frac{1}{N t} \right).
$$

This shows

$$
I_2 \leq N^C \varepsilon (t-u) \left( \frac{N (Nt)^{\alpha}}{l N^2 t} + \frac{1}{N t} \right).
$$

(60)

Finally, for the term $I_3$, by the Lipschitz representation of the averaging operator (57), it can be rewritten in the following way,

$$
I_3 = \frac{1}{2N} \text{Im} \sum_{|i-j| \leq l} \frac{h_j(v_i(v) - \bar{\varphi}_k(v))}{s_i - z} - \frac{1}{2N} \text{Im} \sum_{|i-j| > l} \frac{h_j(\varphi_i(v) - \bar{\varphi}_k(v))}{s_i - z}
+ \frac{1}{2N} \text{Im} \sum_{|i-j| < l} \frac{(h_i - h_j)(\varphi_i(v) - \bar{\varphi}_k(v))}{s_i - z}
+ \frac{1}{2N} \text{Im} \sum_{|i-j| < l} \frac{(1-h_i)(\bar{\varphi}_k(t) - \bar{\varphi}_k(v))}{s_i - z}
=: J_1 + J_2 + J_3 + J_4.
$$

Using (24) and (44), we control $J_1$ in the following way

$$
J_1 \leq \frac{e^\eta}{2N} \left( \text{Im} \mathcal{G}_v(z) - e^{-\eta} \frac{\text{Im} S_v(z)}{\text{Im} S_0(z_v)} \text{Im} \mathcal{G}_0(z_v) \right) + \text{Im} S_v(z) \left( \frac{\text{Im} \mathcal{G}_0(z_v)}{N \text{Im} S_0(z_v)} - \bar{\varphi}_k(v) \right)
\leq \frac{N^C \varepsilon}{N^3 t \eta} + \frac{1}{N^2 t} + \frac{\eta + r}{N t}.
$$

Applying (23) to estimate $J_2$, we obtain

$$
J_2 \leq \frac{1}{2N} \sum_{|i-j| > l} \frac{\eta}{(\gamma_i - \gamma_j)^2 + \eta^2} (|\varphi_i(v) - \bar{\varphi}_i(v)| + |\bar{\varphi}_i(v) - \bar{\varphi}_k(v)|)
\leq \frac{N^C \varepsilon}{2N} \sum_{|i-j| > l} \frac{\eta}{(\gamma_i - \gamma_j)^2 + \eta^2} \left( \frac{(Nt)^{\alpha}}{N^2 t} + \frac{|i-j|}{N^2 t} + \frac{N r}{N^2 t} \right)
\leq \frac{N^C \varepsilon}{N^2 t} \left( \frac{(Nt)^{\alpha} N \eta}{l} + \frac{\eta}{N t} + \frac{r}{N t} \right).
$$
Similarly, by the Lipschitz property of \( \{h_i\} \), we estimate \( J_3 \) as follows

\[
J_3 \leq \frac{1}{2N} \sum_{|i-j|<\ell} \frac{\eta}{(\gamma_i - \gamma_j)^2 + \eta^2} |i-j| \left( \frac{(Nt)^\alpha}{N^2 t} + \frac{|i-j|}{N^2 t} + \frac{Nr}{N^2 t} \right)
\]

\[
\leq Nc_\varepsilon \frac{l}{Nr} \left( \frac{(Nt)^\alpha}{N^2 t} + \frac{r}{Nt} \right) \leq Nc_\varepsilon \left( \frac{(Nt)^\alpha}{N^2 t} + \frac{r}{Nt} \right).
\]

Using the same arguments, by (23), we have

\[
J_4 \leq \frac{Nc_\varepsilon}{2N} \sum_{|i-j|<\ell} \frac{\eta}{(\gamma_i - \gamma_j)^2 + \eta^2} \frac{t-v}{Nt} \leq Nc_\varepsilon \frac{r}{Nt}.
\]

Together with the previous estimates, this leads to

\[
I_3 \leq Nc_\varepsilon \left( \frac{r}{Nt} + \frac{\eta}{Nt} + \frac{(Nt)^\alpha}{N^2 t} \left( \frac{l}{Nt} + \frac{N\eta}{l} + \frac{1}{Nt} \right) \right)
\]

Combined with (59) and (60), we obtain the desired result.\[\Box\]

Finally, we have all the tools to prove Lemma 3.10.

**Proof of Lemma 3.10.** We fix some small \( c > 0 \) and consider an arbitrarily small \( \varepsilon > 0 \). Throughout the whole proof, we do all estimates on the overwhelming probability event where Lemma B.2, Lemma B.3 and Lemma B.4 hold. For a fixed index \( k \in \mathbb{I}((2c-1)N,(1-2c)N) \), we have

\[
|\varphi_k(t) - \Gamma_k(t)| \leq |(\mathcal{T}(u,t) - \mathcal{T}_{\text{short}}(u,t))\varphi(u)_k| + |(\mathcal{T}_{\text{short}}(u,t)(\varphi(u) - \Gamma(u)))_k|.
\]

By the definition of the averaging operator, we know that \( \Gamma(u) = A\varphi(u) = \varphi(u) \) on the set \( \{j : |j-k| \leq Nr\} \). Therefore, combined with the finite speed of propagation estimate (53) for the second term and the short-range approximation (54) for the first term, we obtain

\[
|\varphi_k(t) - \Gamma_k(t)| \leq Nc_\varepsilon (t-u) \left( \frac{(Nt)^\alpha N}{N^2 t} + \frac{1}{Nt} \right) + N^{-2022}. \tag{61}
\]

It suffices to estimate \( |\Gamma_k(t) - \hat{\varphi}_k(t)| \). Consider the function

\[
M(v) := \max_{-N \leq i \leq N} (\Gamma_i(v) - \hat{\varphi}_k(t)).
\]

Similarly as in Lemma 3.2, we can show a parabolic maximum principle for \( M \) and consequently \( M \) decreases in time. Moreover, note that \( \Gamma_i(u) = \hat{\varphi}_k(t) \) if \( |i-k| \geq 2Nr \).

Let \( j = j(v) \) to denote the index that attains the maximum. If there exists a time \( u \leq v \leq t \) such that \( |j-k| > 3Nr \), then in this case the finite speed propagation (53) gives us

\[
M(t) \leq M(v) = \Gamma_j(v) - \hat{\varphi}_k(t) \leq N^{-2022}. \tag{62}
\]

On the other hand, now we assume that \( |j(v) - k| < 3Nr \) for all \( u \leq v \leq t \). In this case, we have

\[
\frac{d}{dv} (\Gamma_j(v) - \hat{\varphi}_k(t)) = \sum_{|i-j|<\ell} \frac{\Gamma_i(v) - \Gamma_j(v)}{2N(s_i(v) - s_j(v))^2} \leq \frac{1}{2N} \sum_{|i-j|<\ell} \frac{\Gamma_i(v) - \Gamma_j(v)}{(s_i(v) - s_j(v))^2 + \eta^2}.
\]

35
This gives us
\[
\frac{d}{dv} (\Gamma_j(v) - \hat{\varphi}_k(t)) \leq \frac{1}{2N\eta} \text{Im} \sum_{|i-j|<l} \frac{\Gamma_i(v)}{s_i(v) - z} - \left( \frac{1}{2N\eta} \text{Im} \sum_{|i-j|<l} \frac{1}{s_i(v) - z} \right) \Gamma_j(v)
\]
and therefore
\[
\frac{d}{dv} (\Gamma_j(v) - \hat{\varphi}_k(t)) \leq \frac{1}{\eta} \left( \frac{1}{2N} \text{Im} \sum_{|i-j|<l} \frac{\Gamma_i(v)}{s_i(v) - z} - \left( \frac{1}{2N} \text{Im} \sum_{|i-j|<l} \frac{1}{s_i(v) - z} \right) \hat{\varphi}_k(v) \right)
+ \frac{1}{\eta} \left( \frac{1}{2N} \text{Im} \sum_{|i-j|<l} \frac{1}{s_i(v) - z} \right) (\hat{\varphi}_k(v) - \hat{\varphi}_k(t))
+ \frac{1}{\eta} \left( \frac{1}{2N} \text{Im} \sum_{|i-j|<l} \frac{1}{s_i(v) - z} \right) (\hat{\varphi}_k(t) - \Gamma_j(v)).
\]
Applying Lemma B.4 and Lemma 3.7 yields
\[
\frac{d}{dv} M(v^+) \lesssim -\frac{1}{\eta} M(v) + \frac{N \epsilon}{\eta} \left( \frac{r}{Nt} + \frac{(Nt)^\alpha}{N^2t} \left( \frac{l}{N} + \frac{N\eta}{l} + \frac{N(t-u)}{l} + \frac{1}{N\eta} \right) \right),
\]
where the left-hand side represents the right derivative of \( M \) at time \( v \). Let \( \eta = \frac{(t-u)}{N^\alpha} \), then above inequality leads to
\[
M(t) \leq N \epsilon \left( \frac{r}{Nt} + \frac{(Nt)^\alpha}{N^2t} \left( \frac{l}{N} + \frac{N(t-u)}{l} + \frac{1}{N(t-u)} \right) \right)
\]
Choosing
\[
\begin{align*}
  r &= \frac{(Nt)^{\frac{3\alpha}{2}}}{N}, \\
  l &= (Nt)^{\frac{\alpha}{2}}, \\
  t-u &= \frac{(Nt)^{\frac{3\alpha}{2}}}{N},
\end{align*}
\]
then we have
\[
M(t) \leq N \epsilon \left( \frac{(Nt)^{3\alpha}}{N^{2t}} \right)
\]
Similarly, this bound also holds for \(-\max_{-N \leq i \leq N} (\Gamma_i(s) - \hat{\varphi}_k(t))\). Combined with (61) and (62), this completes the proof. \(\Box\)

C Proofs for Quantitative Universality

In this section, we prove the quantitative resolvent comparison Proposition 4.1.

The key idea is based on the Lindeberg exchange method (for a detailed introduction we refer to the monograph [EY17, VH14]). We first fix an ordering map of the indices \( \phi : \{(i,j) : 1 \leq i, j \leq N\} \to [N^2] \). For \( 0 \leq k \leq N^2 \), let \( H_k \) be the random matrix defined as
\[
(H_k)_{ij} = \begin{cases} 
X_{ij} & \text{if } \phi(i,j) \leq k \\
Y_{ij} & \text{otherwise}
\end{cases}
\]

so that $H_0 = Y$ and $H_{N^2} = X$. By telescoping summation, it suffices to show the following is true uniformly in $1 \leq k \leq N^2$,

$$\left| \mathbb{E}[F(\text{Tr } f(H_k))] - \mathbb{E}[F(\text{Tr } f(H_{k-1})]] \right| \leq \frac{NC\varepsilon}{N^2} \left( \frac{1}{\rho N^2} + \frac{(\rho N^a)^5}{\sqrt{N}} + t\rho N^a \right) \tag{64}$$

To prove (64), we use the Helffer-Sjöstrand formula. Let $\chi$ be a smooth symmetric cutoff function such that $\chi(y) = 1$ if $|y| < N^{-a}$ and $\chi(y) = 0$ if $|y| > 2N^{-a}$, with $\|\chi\|_\infty \leq N^a$. For any matrix $H \in \{H_k\}_{k=0}^{N^2}$, let $\widetilde{H}$ denote its Girko symmetrization

$$\widetilde{H} = \begin{pmatrix} 0 & H^\top \\ H & 0 \end{pmatrix}$$

Recall that the symmetrized singular values $\{\sigma_i(H)\}_{i=-N}^N$ are the eigenvalues of $\widetilde{H}$. With the cutoff function $\chi$, applying Lemma A.2 to $\widetilde{H}$ yields

$$\text{Tr } f(H) = \int_{\mathbb{C}} g(z) \text{Tr}(\widetilde{H} - z)^{-1} d^2z, \tag{65}$$

where $d^2z$ is the Lebesgue measure on $\mathbb{C}$ and

$$g(z) := \frac{1}{\pi} [iyf''(x)\chi(y) + i(f(x) + iyf'(x))\chi'(y)], \quad z = x + iy.$$

The analysis of the comparison can be proceeded in the following steps:

**Step 1: Approximation of $\text{Tr } f(H)$**. We first truncate the integral in (65) and define

$$\mathcal{T}(H) := \int_{|y| > N^{-2}} g(z) \text{Tr}(\widetilde{H} - z)^{-1} d^2z.$$ 

The approximation error can be bounded by

$$|\text{Tr } f(H) - \mathcal{T}(H)| \lesssim \int_{|y| < N^{-2}, E < |x| < E + \rho} |f''(x)| \sum_{N \leq |k| \leq N} \frac{y^2}{|\sigma_k - (x + iy)|^2} dx \, dy \lesssim \int_{E < |x| < E + \rho} \frac{1}{\rho^2 N^2} \left( \frac{1}{N^2} \sum_{N \leq |k| \leq N} \frac{1}{|\sigma_k - (x + \frac{1}{N})|^2} \right) dx.$$ 

For singular values near the origin, i.e. $|k| \leq NC\varepsilon$, we have

$$\int_{E < |x| < E + \rho} \frac{1}{|\sigma_k - (x + \frac{1}{N})|^2} dx \leq \int_{E < |x| < E + \rho} N^2 dx \lesssim \rho N^2.$$ 

On the other hand, for $|k| > NC\varepsilon$, by the rigidity of singular values, we have the following overwhelming probability bound

$$\int_{E < |x| < E + \rho} \frac{1}{|\sigma_k - (x + \frac{1}{N})|^2} dx \leq \frac{\rho}{|E - \gamma_k|^2}.$$ 

Combining the above two bounds together, we obtain

$$|\text{Tr } f(H) - \mathcal{T}(H)| \lesssim \frac{NC\varepsilon}{\rho N^2} + \frac{1}{\rho N^4} \sum_{|k| > NC\varepsilon} \frac{1}{|E - \gamma_k|^2} \lesssim \frac{NC\varepsilon}{\rho N^2} + \frac{1}{\rho N^3} \left( \frac{1}{N} \sum_{k=1} \frac{1}{(k/N)^2} \right) \lesssim \frac{NC\varepsilon}{\rho N^2}.$$
with overwhelming probability.

**Step 2: Expansions and moment matching.** With the approximation by $T(H)$, it suffices to show that
\[
|\mathbb{E}[F(T(H_k))] - \mathbb{E}[F(T(H_{k-1})]| \lesssim N^{C_\varepsilon} \left( \frac{\rho N^a}{\sqrt{N}} + \rho N^a \right) \tag{66}
\]
uniformly for all $1 \leq k \leq N^2$. Now consider a fixed $1 \leq \omega \leq N^2$ corresponding to the index $(i,j)$, i.e. $\phi(i,j) = \omega$. We rewrite the matrices $H_\omega$ and $H_{\omega-1}$ in the following way
\[
H_\omega = W + \frac{1}{\sqrt{N}} U, \quad H_{\omega-1} = W + \frac{1}{\sqrt{N}} V,
\]
where the matrix $W$ coincides with $H_{\omega-1}$ and $H_\omega$ except on the $(i,j)$ entry with $W_{ij} = 0$. Then note that the matrices $U, V$ satisfy $U_{ij} = \sqrt{N} X_{ij}$ and $V_{ij} = \sqrt{N} Y_{ij}$ and all other entries are zero. Recall the notation $\tilde{H}$ for the Girko symmetrization. Consider the resolvents of the matrices $\tilde{W}$ and $\tilde{H}_\omega$
\[
R := (\tilde{W} - z)^{-1}, \quad S := (\tilde{H}_\omega - z)^{-1}.
\]
The Taylor expansion yields
\[
\mathbb{E}[F(T(H_\omega))] - \mathbb{E}[F(T(H_{\omega-1})] = \sum_{k=1}^{4} \mathbb{E} \left[ \frac{F^{(k)}(T(W))}{k!} \left( (T(H_\omega) - T(W))^k - (T(H_{\omega-1}) - T(W))^k \right) \right] + O(\|F^{(5)}\|_\infty) \mathbb{E} \left[ (T(H_\omega) - T(W))^5 + (T(H_{\omega-1}) - T(W))^5 \right]. \tag{67}
\]
We first control the term corresponding to the fifth derivative. By Lemma A.3, the first order resolvent expansion gives us
\[
\frac{1}{N} \text{Tr} S - \frac{1}{N} \text{Tr} R = \frac{1}{\sqrt{N}} \text{Tr} S \tilde{U} R.
\]
Consequently,
\[
|T(H_\omega) - T(W)| \leq \frac{1}{\sqrt{N}} \int_{|y| > N^{-2}} |g(z)| \left| \text{Tr} S(z) \tilde{U} R(z) \right| d^2z.
\]
We can restrict the integral on the domain $\{ z = x + yi : N^{-2} < |y| < 2N^{-a}, E < |x| < E + \rho \}$ as the contribution outside this region is negligible. Moreover, a key observation is that the matrix $\tilde{U}$ only has two non-zero entries. Thus,
\[
|T(H_\omega) - T(W)| \lesssim N^{\frac{1}{2} + C_\varepsilon} \int_{N^{-2} < |y| < 2N^{-a}, E < |x| < E + \rho} |g(z)| \left( \max_{k \neq \ell} |S_{k\ell}(z)| \right) \left( \max_{k \neq \ell} |R_{k\ell}(z)| \right) d^2z
\]
\[
+ N^{-\frac{1}{2} + C_\varepsilon} \int_{N^{-2} < |y| < 2N^{-a}, E < |x| < E + \rho} |g(z)| \left( \max_k |S_{kk}(z)| \right) \left( \max_k |R_{kk}(z)| \right) d^2z.
\]
Note that in this integral domain, the scale of $|y|$ is smaller than the natural size of the local law. Therefore, we will use a suboptimal version of the local semicircle law for a larger spectral domain, which was discussed in [EKYY13, LS18]. For $z = x + iy$ in this integral domain, with overwhelming probability we have
\[
\max_{k, \ell} |S_{k\ell}(z) - \delta_{k\ell} m_{sc}(z)| \leq N^{C_\varepsilon} \left( \frac{1}{\sqrt{N}} + \Psi(z) \right), \quad \Psi(z) = \frac{1}{N y} + \sqrt{\frac{\text{Im} m_{sc}(z)}{N y}}.
\]
The same result also holds for $R_{k\ell}(z)$. By Lemma A.4, we have $\Psi(z) \lesssim \frac{1}{\sqrt{Ny}}$ for $z$ in the integral domain. Note that the contribution of the diagonal resolvent entries is negligible. Therefore, with overwhelming probability we have

$$|T(H_\omega) - T(W)| \lesssim N^{C_\varepsilon} N^{-1/2} \int_{N^{-2} < |y| < 2N^{-a}, E < |x| < E + \rho} \frac{|g(z)|}{Ny} d^2z \lesssim N^{C_\varepsilon} N^{-1/2} \rho N^a.$$ 

Similarly, this bound also holds for $|T(H_{\omega-1}) - T(W)|$, and we obtain

$$\mathbb{E}[(T(H_\omega) - T(W))^5] \lesssim \frac{NC_\varepsilon}{N^2} \left(N^{-\frac{1}{2}}(\rho N^a)^5\right), \quad \mathbb{E}[(T(H_{\omega-1}) - T(W))^5] \lesssim \frac{NC_\varepsilon}{N^2} \left(N^{-\frac{1}{2}}(\rho N^a)^5\right).$$

Hence the fifth order term in (67) is bounded by

$$O(\|F^{(5)}\|_\infty) \mathbb{E}[(T(H_\omega) - T(W))^5 + (T(H_{\omega-1}) - T(W))^5] \lesssim \frac{NC_\varepsilon}{N^2} \left(\rho N^a\right)^5.$$ (68)

Now we consider the first term $k = 1$ in the Taylor expansion (67). Denote

$$\tilde{R} := \frac{1}{N} \text{Tr} R, \quad \tilde{R}^{(m)}_X := \frac{(-1)^m}{N} \text{Tr}(R \tilde{U})^m R, \quad \Omega_X := -\frac{1}{N} \text{Tr}(R \tilde{U})^5 S,$$

and also define

$$\tilde{R}^{(m)}_Y := \frac{(-1)^m}{N} \text{Tr}(R \tilde{V})^m R, \quad \Omega_Y := -\frac{1}{N} \text{Tr}(R \tilde{V})^5 (\tilde{H}_{\omega-1} - z)^{-1}.$$ 

Using the resolvent expansion (Lemma A.3) up to the fifth order, we obtain

$$\frac{1}{N} \text{Tr} S = \tilde{R} + \sum_{m=1}^{4} N^{-\frac{m}{2}} \tilde{R}^{(m)}_X + N^{-\frac{3}{2}} \Omega_X.$$ 

A similar expansion also holds for $(\tilde{H}_{\omega-1} - z)^{-1}$. Then we have

$$\mathbb{E} \left[F'(T(W)) (T(H_\omega) - T(H_{\omega-1}))\right] = \mathbb{E} \left[F'(T(W)) \int_{|y| > N^{-2}} g(z) \left(\sum_{m=1}^{4} N^{-\frac{m}{2}+1} (\tilde{R}^{(m)}_X - \tilde{R}^{(m)}_Y) + N^{-\frac{3}{2}} (\Omega_X - \Omega_Y)\right) d^2z\right]$$ (69)

A key observation is that for $1 \leq m \leq 3$, the terms $\tilde{R}^{(m)}_X$ and $\tilde{R}^{(m)}_Y$ only depend on the first three moments of $X_{ij}$ and $Y_{ij}$. Recall that the first three moments of $X_{ij}$ and $Y_{ij}$ are identical. Therefore, the terms corresponding to $1 \leq m \leq 3$ in (69) makes no contribution.

**Step 3: Higher order error.** For the $m = 4$ term in (69), note that

$$\text{Tr}(R \tilde{U})^4 R = \sum_{1 \leq \ell \leq 2N} \sum_{\{\alpha_k, \beta_k\} = \{i+j\}} R_{\ell \alpha_1 \beta_1 \alpha_2} \tilde{U}_{\alpha_1 \beta_1 \beta_2} \tilde{U}_{\alpha_2 \beta_2 \beta_3} \tilde{U}_{\alpha_3 \beta_3 \beta_4} \tilde{U}_{\alpha_4 \beta_4} R_{\beta_4 \ell}.$$ (70)

A similar formula is also true for $\text{Tr}(R \tilde{V})^4 R$. Note that typically we have $\ell \neq \alpha_1$ and $\ell \neq \beta_1$, but we may have $\beta_1 = \alpha_2$, $\beta_2 = \alpha_3$, $\beta_3 = \alpha_4$. Moreover, the terms with either $\ell = 1 + N$ or $\ell = j$ are combinatorially negligible in the summation and therefore we can ignore these terms in the
following computations. Recall that the difference between the fourth moments of $X_{ij}$ and $Y_{ij}$ is bounded by $t$. Thus, we have

$$
\mathbb{E} \left[ N(\hat{R}^{(4)}_X - \hat{R}^{(4)}_Y) \right] = \mathbb{E} \left[ \operatorname{Tr}(\hat{R}\hat{U})^4R - \operatorname{Tr}(\hat{R}\hat{V})^4R \right] \lesssim Nt \left( \max_{k \neq \ell} |R_{k\ell}| \right)^2 \left( \max_k |R_{kk}| \right)^3.
$$

As mentioned above, for the integral in (69) we can restrict the integral domain to $N^{-2} < |y| < 2N^{-a}$ and $E < |x| < E + \rho$. In this region, the entries of the resolvent are bound by $\max_{k \neq \ell} |R_{k\ell}(z)| \lesssim \frac{NC_\varepsilon}{\sqrt{N}y}$ and $\max_k |R_{kk}(z)| \lesssim NC_\varepsilon$. As a consequence,

$$
\left| \mathbb{E} \left[ F'(T(W)) \right|_{|y|>N^{-2}} g(z)N^{-\frac{3}{2}} \Omega_X d^2z \right| 
\lesssim \frac{NC_\varepsilon t}{N} \int_{N^{-2} < |y| < 2N^{-a}, E < |x| < E + \rho} \frac{|g(z)|}{N^2} \frac{d^2z}{N^2 \sqrt{N}} \lesssim \frac{NC_\varepsilon}{N^2 \sqrt{N}}. \quad (71)
$$

For the term $\Omega_X - \Omega_Y$, since these terms involve the higher moments of $X_{ij}$ and $Y_{ij}$, we simply bound it by the size of $\Omega_X$ and $\Omega_Y$. By a similar expansion as in (70) and the local law, we have $|\Omega_X|, |\Omega_Y| \lesssim \frac{NC_\varepsilon}{\sqrt{N}y}$. Therefore,

$$
\left| \mathbb{E} \left[ F'(T(W)) \right|_{|y|>N^{-2}} g(z)N^{-\frac{3}{2}} \Omega_X d^2z \right| 
\lesssim \frac{NC_\varepsilon N^{-\frac{3}{2}}}{N} \int_{N^{-2} < |y| < 2N^{-a}, E < |x| < E + \rho} \frac{|g(z)|}{N^2} \frac{d^2z}{N^2 \sqrt{N}} \lesssim \frac{NC_\varepsilon}{N^2 \sqrt{N}} \left( \frac{\rho N^a}{N^2} \right)^5. \quad (72)
$$

The same bound also holds for $\Omega_Y$.

Finally, as explained in classical literature of random matrix theory (see e.g. [EY17, Theorem 17.4]), the contributions of higher order terms in the Taylor expansion (67) are of smaller order. Consequently, combining (68), (71) and (72) yields the claim (66), which implies the desired result (33).

References

[AEK14] Oskari Ajanki, László Erdős, and Torben Krüger. Local semicircle law with imprimitive variance matrix. *Electron. Commun. Probab.*, 19:no. 33, 9, 2014.

[AEK17] Johannes Alt, László Erdős, and Torben Krüger. Local law for random Gram matrices. *Electron. J. Probab.*, 22:Paper No. 25, 41, 2017.

[AMR11] David Arthur, Bodo Manthey, and Heiko Röglin. Smoothed analysis of the k-means method. *Journal of the ACM (JACM)*, 58(5):1–31, 2011.

[BC10] Peter Bürgisser and Felipe Cucker. Smoothed analysis of moore–penrose inversion. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2769–2783, 2010.

[BCMV14] Aditya Bhaskara, Moses Charikar, Ankur Moitra, and Aravindan Vijayaraghavan. Smoothed analysis of tensor decompositions. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 594–603, 2014.
[BEK+14] Alex Bloemendal, László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19:no. 33, 53, 2014.

[BGK17] Florent Benaych-Georges and Antti Knowles. Local semicircle law for Wigner matrices. In *Advanced topics in random matrices*, volume 53 of *Panor. Synthèses*, pages 1–90. Soc. Math. France, Paris, 2017.

[Bou22] Paul Bourgade. Extreme gaps between eigenvalues of Wigner matrices. *Journal of the European Mathematical Society*, 24(8):2823–2873, 2022.

[BY17] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicity. *Comm. Math. Phys.*, 350(1):231–278, 2017.

[BYY14] Paul Bourgade, Horng-Tzer Yau, and Jun Yin. Local circular law for random matrices. *Probability Theory and Related Fields*, 159(3-4):545–595, 2014.

[CL19] Ziliang Che and Patrick Lopatto. Universality of the least singular value for sparse random matrices. *Electron. J. Probab.*, 24:Paper No. 9, 53, 2019.

[CMS13] Claudio Cacciapuoti, Anna Maltsev, and Benjamin Schlein. Local Marchenko-Pastur law at the hard edge of sample covariance matrices. *J. Math. Phys.*, 54(4):043302, 13, 2013.

[Ede88] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[EKYY13] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of erdős–rényi graphs i: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.

[ESYY12] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(1):1–46, 2012.

[EY17] László Erdős and Horng-Tzer Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.

[EYY11] László Erdős, Horng-Tzer Yau, and Jun Yin. Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.*, 2(1):15–81, 2011.

[FV16] Brendan Farrell and Roman Vershynin. Smoothed analysis of symmetric random matrices with continuous distributions. *Proceedings of the American Mathematical Society*, 144(5):2257–2261, 2016.

[GPV21] Mehrdad Ghadiri, Richard Peng, and Santosh S Vempala. Sparse regression faster than $d^2$. *arXiv preprint arXiv:2109.11537*, 2021.

[Hig02] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002.

[LS18] Ji Oon Lee and Kevin Schnelli. Local law and tracy–widom limit for sparse random matrices. *Probability Theory and Related Fields*, 171(1):543–616, 2018.
[LSY19] Benjamin Landon, Philippe Sosoe, and Horng-Tzer Yau. Fixed energy universality of Dyson Brownian motion. *Adv. Math.*, 346:1137–1332, 2019.

[MT16] Govind Menon and Thomas Trogdon. Smoothed analysis for the conjugate gradient algorithm. *SIGMA. Symmetry, Integrability and Geometry: Methods and Applications*, 12:109, 2016.

[Nie22] Zipei Nie. Matrix anti-concentration inequalities with applications. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 568–581, 2022.

[PV21] Richard Peng and Santosh Vempala. Solving sparse linear systems faster than matrix multiplication. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 504–521. SIAM, 2021.

[Sma85] Steve Smale. On the efficiency of algorithms of analysis. *Bulletin of the American Mathematical Society*, 13(2):87–121, 1985.

[SS20] Rikhav Shah and Sandeep Silwal. Smoothed analysis of the condition number under low-rank perturbations. *arXiv preprint arXiv:2009.01986*, 2020.

[SST06] Arvind Sankar, Daniel A Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis and Applications*, 28(2):446–476, 2006.

[ST04] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[SW09] Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics*, volume 59 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. Reprint of the 1986 original [MR0838963].

[SX21] Kevin Schnelli and Yuanyuan Xu. Convergence rate to the Tracy-Widom laws for the largest eigenvalue of sample covariance matrices. *arXiv preprint arXiv:2108.02728*, 2021.

[SX22a] Kevin Schnelli and Yuanyuan Xu. Convergence rate to the Tracy-Widom laws for the largest eigenvalue of Wigner matrices. *Comm. Math. Phys.*, 393(2):839–907, 2022.

[SX22b] Kevin Schnelli and Yuanyuan Xu. Quantitative tracy-widom laws for the largest eigenvalue of generalized Wigner matrices. *arXiv preprint arXiv:2207.00546*, 2022.

[TB97] Lloyd N. Trefethen and David Bau, III. *Numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[TV10a] Terence Tao and Van Vu. Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.*, 20(1):260–297, 2010.

[TV10b] Terence Tao and Van Vu. Smooth analysis of the condition number and the least singular value. *Mathematics of computation*, 79(272):2333–2352, 2010.

[VH14] Ramon Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.
[Wan19] Haoyu Wang. Quantitative universality for the largest eigenvalue of sample covariance matrices. *arXiv preprint arXiv:1912.05473*, 2019.

[Wsc04] Mario Wschebor. Smoothed analysis of $\kappa(a)$. *Journal of Complexity*, 20(1):97–107, 2004.