Electronic Journal of Probability

Electron. J. Probab. **24** (2019), no. 9, 1–53.
ISSN: 1083-6489 https://doi.org/10.1214/19-EJP269

Universality of the least singular value for sparse random matrices*

Ziliang Che[†] Patrick Lopatto[‡]

Abstract

We study the distribution of the least singular value associated to an ensemble of sparse random matrices. Our motivating example is the ensemble of $N \times N$ matrices whose entries are chosen independently from a Bernoulli distribution with parameter p. These matrices represent the adjacency matrices of random Erdős–Rényi digraphs and are sparse when $p \ll 1$. We prove that in the regime $pN \gg 1$, the distribution of the least singular value is universal in the sense that it is independent of p and equal to the distribution of the least singular value of a Gaussian matrix ensemble. We also prove the universality of the joint distribution of multiple small singular values. Our methods extend to matrix ensembles whose entries are chosen from arbitrary distributions that may be correlated, complex valued, and have unequal variances.

Keywords: random matrix theory; sparse; universality, singular value.
AMS MSC 2010: 60B20.
Submitted to EJP on April 16, 2018, final version accepted on January 22, 2019.
Supersedes arXiv:1711.00580.

Contents

*Z.C. is partially supported by NSF grant DMS-1607871. P.L. is partially supported by the NSF Graduate Research Fellowship Program under Grant DGE-1144152.
[†]Harvard University, United States of America. E-mail: zche@math.harvard.edu
[‡]Harvard University, United States of America. E-mail: lopatto@math.harvard.edu

1 Introduction

Random real symmetric and complex Hermitian matrices have been intensely studied since Wigner's discovery that, in the large N limit, their eigenvalue densities are universal and follow the semicircle distribution. Recent investigations have culminated in a proof of the Wigner–Dyson–Mehta conjecture, which asserts the universality of the local eigenvalue statistics in the limit [9, 16, 20]. [1]

In the case of non-Hermitian matrices, there has been a similar study of the singular values. Given a $N \times N$ matrix M, its singular values are the eigenvalues of $\sqrt{M^\dagger M}$, which we label

$$0 \leq \lambda_1 \leq \cdots \leq \lambda_N. \tag{1.1}$$

Traditionally, one studies the squares of the singular values, the eigenvalues of $M^\dagger M$. In the bulk, the limiting distribution of the squares of the singular values is universal under fairly general hypotheses and follows the Marchenko–Pastur law [1, 4, 32]. Averaged-energy universality for the local correlation functions around a fixed energy in the bulk was shown in [22, 33], and [33] also showed universality for the largest singular value λ_N. However, the methods in [33] do not suffice to prove universality of the least singular value. The analysis of this case is more subtle because the Marchenko–Pastur distribution has a density with a singularity at the origin.

Early work on the least singular value of random matrices was motivated by the analysis of algorithms in computer science. An important recent development in this area is the method of smoothed analysis, which estimates the practical performance of algorithms [35]. In these applications, it is important to estimate the probability that λ_1 is small for various random matrix models. Since the inverse of the least singular value of a matrix is equal to the operator norm of its inverse, these estimates control the probability that the inverse has large norm. We refer the reader to [34] for an introduction to this line of research. Recent results include [36] on Bernoulli matrices, [14] on structured random matrices, [38] on matrices with i.i.d. entries shifted by a deterministic matrix, and [5, 31] on sparse matrices.

[1] Z.C. is partially supported by NSF grant DMS-1607871. P.L. is partially supported by the NSF Graduate Research Fellowship Program under Grant DGE-1144152.

The universality of the least singular value distribution was considered in [37] from a viewpoint inspired by the method of property testing in computer science and combinatorics. The main result is that if ξ is a real random variable with $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = 1$, and such that $\mathbb{E}|\xi|^C < \infty$ for some sufficiently large absolute constant C, then the distribution of the least singular value of the ensemble of $N \times N$ random matrices M_N with entries chosen i.i.d. with distribution $\xi/\sqrt{N}$ satisfies

$$\mathbb{P}\left(N\lambda_1(M_N) \le r\right) = 1 - e^{-r^2/2-r} + O(N^{-c}), \tag{1.2}$$

for some $c > 0$. Also given in [37] are analogous results for complex matrices and for the joint distribution of multiple smallest singular values. However, these results require that the entries are independent and have equal variances.

This paper studies the universality of the least singular value from the same dynamical viewpoint that was used to prove the Wigner–Dyson–Mehta conjecture. Our motivating example is the ensemble of $N \times N$ matrices whose entries are chosen independently from a Bernoulli distribution with parameter p. These matrices represent the adjacency matrices of random Erdős–Rényi digraphs, which are directed graphs on N vertices where each possible directed edge is present with probability p. Such matrices are sparse when $p \ll 1$, and our result implies that the distribution of the least singular value is universal in the regime $pN \gg 1$. We also apply our method to prove universality of the least singular value for matrices whose entries have unequal variances and weak correlations.

An important feature of our proof is that we consider an analogue of Dyson Brownian motion where the particles move in the Weyl chamber corresponding to the hyperoctahedral group. This is in contrast to the literature on the universality of eigenvalue statistics, which studies the traditional Dyson Brownian motion with particles restricted to the Weyl chamber corresponding to the symmetric group. For background on Brownian motion in a Weyl chamber, we refer the reader to [24]. An essential technical input is showing that these dynamics, which govern the evolution of the singular value distribution, reach local equilibrium after short times $t \gg N^{-1}$.

Acknowledgments. The authors thank H.-T. Yau for suggesting the problem, useful discussions, and helpful comments on our preliminary draft. P.L. thanks A. Aggarwal, B. Landon, and P. Sosoe for useful discussions. Finally, the authors thank the referee for their comments and suggestions, which significantly improved the paper.

2 Overview and main result

In this section we state and prove our main theorem on the universality of the least singular value for sparse matrices, invoking several preliminary results proved in the forthcoming sections. We adapt the three-step method [20–23] and take advantage of recent technical improvements [26–28]. Step 1 is to obtain local control of the singular values of a deformed ensemble, and in particular to prove their rigidity, which is necessary for the stochastic analysis in the second step. Step 2 is to obtain short time universality. We prove that for a sparse matrix, the least singular value is universal after time $t = N^{-1+\varepsilon}$ when the singular values are evolved according to the singular value analogue of Dyson Brownian motion. Step 3 is to remove the time evolution and prove universality for the original ensemble.

Section 3 contains the main estimate of this paper, the analysis of the time evolution of the singular values, which is necessary for Step 2. In Section 4 we carry out Step 1. Step 3 is accomplished in Section 5 by a Green function comparison theorem. In Section 6 we discuss how to extend our methods to matrices with correlated entries and unequal variances, and in Appendix A we verify the SDE governing the evolution of the singular values is well-posed.

We now define the primary matrix model we study in this work. To motivate this definition, consider an $N \times N$ matrix X of independent Bernoulli random variables $\{x_{ij}\}$ that take the value 1 with probability p and 0 with probability $1-p$ for some parameter p, which may depend on N. Heuristically, in order to place the bulk eigenvalues on an interval of constant order, we must normalize the matrix by the average ℓ^2 sum of a row, which in this case is $\sqrt{pN}$. Setting $q = \sqrt{pN}$, $M = X/q$ and extracting the mean f of the resulting entries, we may decompose the normalized matrix as

$$M = B + f \left| w \right\rangle \left\langle w \right|$$

where $f = \sqrt{p}$. The parameter q also plays a role in B, where the k-th moment of each entry is $O(N^{-1}q^{2-k})$.

We generalize this setting and come to the following definition.

Definition 2.1. We say a sequence of matrices $(M_N)_{N=1}^{\infty}$ is a sparse random matrix ensemble with sparsity parameter q and mean f if, for every N, M_N is a $N \times N$ matrix of the form

$$M_N = B_N + f \left| w \right\rangle \left\langle w \right| \tag{2.1}$$

where $w = N^{-1/2}(1, \ldots, 1)^{\dagger}$, f is a parameter (which may depend on N) such that $0 \le f \le N^{1/2}$, and B_N is a real matrix with independent entries b_{ij} such that

$$\mathbb{E}[b_{ij}] = 0, \quad \mathbb{E}[b_{ij}^2] = s_{ij}^{(N)}, \quad \mathbb{E}[|b_{ij}|^k] \le \frac{C^k}{Nq^{k-2}} \tag{2.2}$$

for all k. We assume there exists $\alpha > 0$ such that q satisfies

$$N^{\alpha} \le q \le N^{1/2}. \tag{2.3}$$

We also assume the variance matrices S_N are doubly stochastic with elements $s_{ij}^{(N)}$ satisfying

$$\frac{c}{N} \le s_{ij}^{(N)} \le \frac{C}{N} \tag{2.4}$$

for some universal constants c and C.

Remark. We have chosen to state our results for the model in Definition 2.1 to simplify the exposition. However, several generalizations are possible.

First, the assumption that the variance matrices S_N are doubly stochastic is made for convenience, so that we are in the usual case where the limiting global distribution of the singular values is a quarter circle. We prove universality in the technically more involved case of matrices with correlated entries in Section 6, and our work there subsumes the case of matrices with independent entries and a non-stochastic variance matrix.

Second, the condition (2.2) may be relaxed considerably to require bounds on only a finite number of moments. These moment bounds are used to prove certain stochastic domination estimates in Section 5 by applying Markov's inequality to large moments of the entries (see Definition 4.2). However, while Definition 4.2 requires a certain estimate to hold for all D and $\varepsilon > 0$, our proof requires this estimate only for a fixed large D and fixed small $\varepsilon > 0$ (independent of the size of the matrix N). Hence, the bound (2.2) is needed only for some large but fixed number of moments, and the condition can be weakened to requiring that

$$\mathbb{E}[|b_{ij}|^k] \le \frac{C}{Nq^{k-2}} \tag{2.5}$$

for some constant C for all $k \le K_0$, where K_0 is some large constant. In particular, our result extends to prove universality for random matrices with entries of the form $b_{ij}\xi_{ij}$, where $\{b_{ij}\}$ are independent Bernoulli random variables with parameter $p = N^{-1+\delta}$ for any $\delta > 0$ and and $\{\xi_{ij}\}$ are i.i.d. random variables with all moments finite.

Finally, the results of [19] suggest universality of the least singular value should hold for matrices with sparsity parameter as small as $q = (\log N)^C$ for some large constant C (and even this is probably not optimal, in light of some recent results on sparse matrices [15, 25] and the relaxation time of Dyson Brownian motion in the Hermitian case [13]). Our approach would generalize to this case if we could show short time universality for $t = (\log N)^C/N$. However, we show this only for $t = N^{-1+\varepsilon}$. It is plausible that a careful examination of our proof would yield the stronger result, but we do not take this up here.

Our main result shows that the least singular value of a sparse random matrix ensemble is universal in the large N limit. The Wigner ensemble case of this result was proved in Theorem 1.3 of [37].

Theorem 2.2. Let $(M_N)_{N=1}^{\infty}$ be a sparse matrix ensemble with least singular values $\lambda_1(M_N)$. For all $r \geq 0$, we have

$$\mathbb{P}\left(N\lambda_1(M_N) \leq r\right) = 1 - e^{-r^2/2-r} + O(N^{-c}) \tag{2.6}$$

where $c > 0$ is an absolute constant uniform in r.

Since our short time universality estimate Theorem 3.2 holds not just for the smallest singular value, but also for the smallest k singular values when k is at most a small power of N, we also obtain universality of the joint distribution of the smallest k singular values for fixed k. This is the content of the following theorem. While we do not state the universal distribution explicitly, an exact expression can be derived, and we refer the reader to Section 6 of [37] for details.

Theorem 2.3. Fix a positive integer k. Let $(M_N)_{N=1}^{\infty}$ be a sparse ensemble and $(G_N)_{N=1}^{\infty}$ be an ensemble with independent entries of distribution $\mathcal{N}(0, N^{-1})$. For any matrix A, define

$$\Lambda_k(A) = (N\lambda_1(A), \ldots, N\lambda_k(A)), \tag{2.7}$$

and for any choice of energies $\widehat{E} = (E_i) \in \mathbb{R}^k$, define

$$\Omega(\widehat{E}) = \left\{x \in \mathbb{R}^k \colon x_i \leq E_i \text{ for all } i \leq k\right\}. \tag{2.8}$$

Let $\widehat{E} \pm N^{-c}$ denote the vector $(E_i \pm N^{-c}) \in \mathbb{R}^k$. Then

$$\begin{aligned}\mathbb{P}\left(\Lambda_k(G_N) \in \Omega(\widehat{E} - N^{-c})\right) - N^{-c} &\leq \mathbb{P}\left(\Lambda_k(M_N) \in \Omega(\widehat{E})\right) \\ &\leq \mathbb{P}\left(\Lambda_k(G_N) \in \Omega(\widehat{E} + N^{-c})\right) + N^{-c},\end{aligned}$$

uniformly in all choices of $\widehat{E} \in \mathbb{R}^k$, for some $c > 0$ and large enough N.

Finally, we mention that analogues of our results hold when the matrix entries are complex valued. The proofs are essentially identical.

2.1 Proof of Theorem 2.2

Here we present the proof of our main result, Theorem 2.2. The proof of Theorem 2.3 is analogous.

We suppose for the remainder of this paper that the entry distributions for all random matrices considered are absolutely continuous with respect to Lebesgue measure, so that their singular values are distinct almost surely and can be strictly ordered. The case of a general matrix H is dealt with by considering $H(\varepsilon) = H + \varepsilon V$, where V is a Gaussian matrix. The entires of $H(\varepsilon)$ have distributions that are absolutely continuous for $\varepsilon > 0$, since their laws are convolutions with a Gaussian distribution. All results may be extended to H by taking the limit as ε goes to 0 and using Weyl's inequality for singular values.

We recall the following definitions used in the proof.

Definition 2.4. We say that an event $\mathcal{F}$ holds with overwhelming probability if for any $D > 0$ we have $\mathbb{P}(\mathcal{F}^c) \le N^{-D}$ for large enough N. For a family of events $\{\mathcal{F}(u)\}$, we say $\{\mathcal{F}(u)\}$ holds with overwhelming probability if $\sup_u \mathbb{P}(\mathcal{F}(u)^c) \le N^{-D}$ for large enough N.

Definition 2.5. Given a random matrix with eigenvalues $\{\lambda_i\}$ and $E_1 \le E_2$, define the eigenvalue counting function

$$n(E_1, E_2) = |\{E_1 < \lambda_i < E_2\}|. \tag{2.9}$$

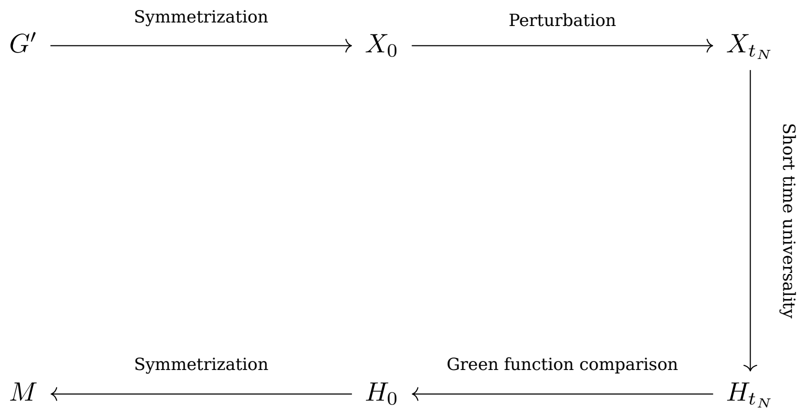

Figure 1: Diagram of the proof of Theorem 2.2. The sparse ensemble M and a Gaussian ensemble G' are both symmetrized in order to access their singular values through the eigenvalues of the symmetric matrices H_0 and X_0. The distribution of the eigenvalue counting function near zero for the perturbed Gaussian matrix X_t is transferred to H_t via a short time universality result. Finally, this distribution is pulled back to M using a Green function comparison argument.

Proof. For any N, let G_N be a $N \times N$ matrix with independent entries of distribution $\mathcal{N}(0, N^{-1})$. We set $M_t = M_N + \sqrt{t} G_N$ and let $\lambda_1^\circ(t)$ be the least singular value of M_t, suppressing dependence on N. Let G_N' and G_N'' be matrices independent from G_N with the same entry distributions, and let $\tilde{\lambda}_1^\circ(t)$ be the least singular value of $G_t = G_N' + \sqrt{t} G_N''$. Define the block matrices

$$H_t = \begin{bmatrix} 0 & M_t \\ M_t^\dagger & 0 \end{bmatrix}, \quad X_t = \begin{bmatrix} 0 & G_t \\ G_t^\dagger & 0 \end{bmatrix}. \tag{2.10}$$

Note that the eigenvalues of H_t are precisely the singular values of M_t and their negatives, and same relation holds between X_t and G_t.

For any $\varepsilon > 0$, Lemma 5.10 on short time universality shows that there exists a coupling $(\lambda_i(t), \tilde{\lambda}_i(t))$ of the $\lambda_i^\circ(t)$ and $\tilde{\lambda}_i^\circ(t)$ such that with overwhelming probability we have

$$|\lambda_1(t_N) - \tilde{\lambda}_1(t_N)| \le \frac{1}{N^{1+\sigma}} \tag{2.11}$$

for some N-dependent parameter t_N satisfying $t_N \le N^{-1+\varepsilon}$ and some $\sigma > 0$.

We conclude by removing the time evolution. This is accomplished by comparing smoothed versions of the eigenvalue counting functions using results in Section 5. Let n_t be the eigenvalue counting function for the $\lambda_i(t)$ and $\tilde{n}_t$ be the counting function for the $\tilde{\lambda}_i(t)$, as in Definition 2.5. Set

$$E = r/N, \quad y = N^{-1-\sigma}, \quad \eta = yN^{-32\sigma}. \tag{2.12}$$

Recasting (2.11) in terms of these counting functions gives that, for any $D > 0$,

$$\mathbb{P}(\tilde{n}_{t_N}(-E-y, E+y) = 0) - N^{-D} \le \mathbb{P}(n_{t_N}(-E, E) = 0) \le \mathbb{P}(\tilde{n}_{t_N}(-E+y, E-y) = 0) + N^{-D}. \tag{2.13}$$

The conclusion of Lemma 5.13 is that

$$\mathbb{E}q(\operatorname{Tr}\chi_{E+y} \star \theta_\eta(H_{t_N})) - N^{-D} \le \mathbb{P}(n_{t_N}(-E, E) = 0) \le \mathbb{E}q(\operatorname{Tr}\chi_{E-y} \star \theta_\eta(H_{t_N})) + N^{-D}, \tag{2.14}$$

where q is a smooth cutoff function defined in Section 5, and

$$\theta_\eta = \frac{\eta}{\pi(x^2 + \eta^2)} = \frac{1}{\pi} \operatorname{Im} \frac{1}{x - i\eta}. \tag{2.15}$$

Further, Lemma 5.15 shows that there exists $\varepsilon > 0$ and $c > 0$ such that

$$\mathbb{E}q(\operatorname{Tr}\chi_{E-y} \star \theta_\eta(H_t)) \le \mathbb{E}q(\operatorname{Tr}\chi_{E-y} \star \theta_\eta(H_0)) + N^{-c} \le \mathbb{P}(n_0(-E+2y, E-2y) = 0) + N^{-c} \tag{2.16}$$

for all $t \le N^\varepsilon/N$, where the last inequality holds by another application of Lemma 5.13.

Fix this ε and corresponding $t_N \le N^\varepsilon/N$ for the rest of the proof. After adjusting c downward and combining this display with the previous one, we have for large enough N that

$$\mathbb{P}(n_{t_N}(-E, E) = 0) \le \mathbb{P}(n_0(-E+2y, E-2y) = 0) + N^{-c}. \tag{2.17}$$

Similar reasoning, interchanging the roles of n_0 and n_t, gives

$$\mathbb{P}(n_0(-E-2y, E+2y) = 0) \le \mathbb{P}(n_{t_N}(-E, E) = 0) + N^{-c}. \tag{2.18}$$

We conclude, after setting $\tilde{y} = 2y$, that there exists $c > 0$ such that, for large enough N,

$$\mathbb{P}(\tilde{n}_0(-E-\tilde{y}, E+\tilde{y}) = 0) - N^{-c} \le \mathbb{P}(n_0(-E, E) = 0) \le \mathbb{P}(\tilde{n}_0(-E+\tilde{y}, E-\tilde{y}) = 0) + N^{-c}. \tag{2.19}$$

Rephrased in the language of cumulative distribution functions, this is

$$\mathbb{P}\left(\tilde{\lambda}_1(0) \le E - \tilde{y}\right) - N^{-c} \le \mathbb{P}\left(\lambda_1(0) \le E\right) \le \mathbb{P}\left(\tilde{\lambda}_1(0) \le E + \tilde{y}\right) + N^{-c}. \tag{2.20}$$

Since $\tilde{\lambda}_1$ is the least singular value of a matrix with i.i.d. entries, we may use Theorem 1.3 in [37] to control its distribution. After changing the variable of integration, we have for some small $c > 0$ that

$$\mathbb{P}\left(\tilde{\lambda}_1(0) \le \frac{r}{N}\right) = 1 - e^{-r^2/2-r} + O(N^{-c}). \tag{2.21}$$

Note that (2.21) shows

$$\mathbb{P}\left(\tilde{\lambda}_1(0) \in (E-\tilde{y}, E+\tilde{y})\right) \le N^{-c}. \tag{2.22}$$

Together with (2.20), this completes the proof. □

3 Short time universality

In order to state the main result of this section, we introduce the following notation. We fix $\delta_1 > 0$ and let g and G be N-dependent parameters such that

$$N^{-1+\delta_1} \le g \le N^{-\delta_1}, \quad G \le N^{-\delta_1}. \tag{3.1}$$

We consider a deterministic matrix V of initial data and let $B_t = \{B_{ij}(t)\}_{1\le i,j\le N}$ be a matrix of i.i.d. real Brownian motions. We define

$$M_t = V + \frac{1}{\sqrt{N}} B_t, \quad H_t = \begin{bmatrix} 0 & M_t \\ M_t^\dagger & 0 \end{bmatrix}. \tag{3.2}$$

Let $\{s_i(t)\}_{i=-N}^{N}$ (omitting the zero index) be the eigenvalues of H_t, which are the singular values of M_t along with their negatives. We set

$$m_V(z) = \sum_{i=-N}^{N} \frac{1}{s_i(0) - z}. \tag{3.3}$$

Definition 3.1. With g and G defined as above, we say V is (g, G)-regular if

$$c \le \operatorname{Im} m_V(E + i\eta) \le C \tag{3.4}$$

for $|E| \le G$ and $\eta \in [g, 10]$ for large enough N, and if there exists a constant C_V such that $|v_i| \le N^{C_V}$ for all v_i.

We now state the main result of this section. Let W be a random matrix whose entries are i.i.d. $\mathcal{N}(0, N^{-1})$ variables and let $\tilde{B}_t = \{\tilde{B}_{ij}(t)\}_{1\le i,j\le N}$ be a matrix of i.i.d. real Brownian motions. Define $W_t = W + N^{-1/2}\tilde{B}_t$. Recall $\{s_i(t)\}_{i=1}^{N}$ are the singular values of M_t, and let $\{r_i(t)\}_{i=1}^{N}$ be the singular values of W_t.

Theorem 3.2. Fix $\sigma > 0$, and let V be a (g, G)-regular deterministic matrix. Let M_t, W_t, $\{s_i(t)\}$, and $\{r_i(t)\}$ be defined as above. Then there exists a coupling of the processes $\{s_i(t)\}$ and $\{r_i(t)\}$ such that the following holds. Given parameters $0 < \omega_1 < \omega_0$ and times $t_0 = N^{-1+\omega_0}$, $t_1 = N^{-1+\omega_1}$, with the restrictions that

$$g N^{\sigma} \le t_0 \le N^{-\sigma} G^2, \quad 2\omega_1 < \omega_0, \tag{3.5}$$

there exist $\omega, \delta > 0$ such that

$$|s_i(t_a)) - r_i(t_a)| < N^{-1-\delta} \tag{3.6}$$

for $i < N^{\omega}$ and $t_a = t_0 + t_1$.

3.1 Preliminaries

We prove Theorem 3.2 by analyzing the SDE that governs the time evolution of the singular values of M_t. Given initial data $(s_i(0))_{i=1}^{N}$, the SDE for the perturbed singular values is

$$ds_i(t) = \frac{dB_i}{\sqrt{N}} + \frac{1}{2N} \sum_{j\ne i} \left(\frac{1}{s_i(t) - s_j(t)} + \frac{1}{s_i(t) + s_j(t)} \right) dt. \tag{3.7}$$

To be precise, we mean that the stochastic process $(s_i(t))_{i=1}^{N}$ and the stochastic process given by the singular values of M_t are equal in distribution.

Technical information about this SDE is contained in Appendix A, where we show existence and uniqueness of strong solutions and verify that it represents the claimed evolution of the singular values (in distribution). The arguments used in the appendix may also be used to show existence and uniqueness of strong solutions for the other SDEs considered in this section.

Our main idea is to analyze a symmetrized version of (3.7). Define $s_{-k}(t)$ and $B_{-k}(t)$ for $1 \le k \le N$ by

$$s_k(t) = -s_{-k}(t), \quad B_k(t) = -B_{-k}(t), \quad t \ge 0.$$

Then the SDE (3.7) is equivalent to

$$ds_i = \frac{dB_i}{\sqrt{N}} + \frac{1}{2N} \sum_{\substack{-N \le j \le N \\ j \ne \pm i, 0}} \frac{dt}{s_i - s_j}. \tag{3.8}$$

Note we have labeled the particles from -1 to $-N$ and 1 to N, so that the zero index is omitted. Unless otherwise stated, this will be our convention throughout this section, and we will no longer note this omission explicitly.

Note also that (3.8) is the same as a standard DBM, except it is missing the repulsion term for $j = -i$. In particular, for the least singular value $i = 1$, the force deflecting this particle from the origin is weaker than in the usual DBM. This complicates the analysis of (3.8), since we are now missing an important regularizing effect, and level repulsion estimates such as the one used in [28] to study the short time behavior of DBM seem out of reach. The key point of this section is that the smallest singular values reach equilibrium on short time scales even without this regularizing term.

We begin by studying the differences $s_i(t) - r_i(t)$. By demanding that the SDEs for the r_i and s_i are driven by the same Brownian motions, we force the Brownian motion term to vanish in the equation for the differences, which renders it considerably easier to work with. This coupling approach was developed in [9, 28]. Here we use a refinement of this idea, introduced in [27], and construct a continuous interpolation. Next, we obtain precise control over this difference equation by comparing its semigroup elements to the solution kernel of an appropriate integral equation, which represents the homogenized version of the DBM. The solution kernel is approximately the Poisson kernel, which corresponds to the fact that the DBM difference equation is a discrete version of the square root of the Laplacian (see Section 7.11 of [30] for details in the continuous case). Theorem 3.2 then follows from a short calculation in Subsection 3.6 exploiting cancellation in the kernel coming from the symmetrization of the SDE.

Except for the final calculation, this plan follows closely [27], whose methods are fundamental to our approach. The work [27] proved short time universality, at level of particle gaps, for DBM with deterministic initial data. This result, called “fixed energy universality,” is of great importance because it enables the proof of bulk universality for a vast array of matrix models. Our result is similarly flexible and we have stated it in more generality than needed for this work in order to facilitate future applications.

Remark. In this section we deal only with the real variable case, where we apply a real Gaussian perturbation. If we consider complex matrices, the appropriate SDE is

$$ds_k = \sqrt{\frac{1}{2N}}\, dB_k + \frac{1}{2N} \sum_{j \ne \pm k} \frac{dt}{s_k - s_j}. \tag{3.9}$$

The only difference here is that the diffusion term loses a factor of $\sqrt{2}$. The argument in the complex case is the same as the real case, and the main result of this section Theorem 3.2 holds with no difference.

3.1.1 Definitions

Given a probability measure $\rho(E)\,dE$ and some even number $2N$, we define the classical particle locations γ_i as follows, suppressing dependence on N. For $i \geq 1$, we set

$$\gamma_i = \inf\left\{x\colon \int_{-\infty}^{x} \rho(E)\,dE \geq \frac{N+i-1}{2N}\right\}. \tag{3.10}$$

For $i \leq -1$, we set

$$\gamma_i = \inf\left\{x\colon \int_{-\infty}^{x} \rho(E)\,dE \geq \frac{N+i}{2N}\right\}. \tag{3.11}$$

In accordance with our labeling convention, we do not define γ_0. Our labeling is chosen so that $\gamma_1 = 0$ for a symmetric distribution centered at 0.

We recall the Stieltjes transform of a $N \times N$ matrix M with eigenvalues λ_i is given by

$$m_N(z) = \frac{1}{N}\sum_i \frac{1}{\lambda_i - z}. \tag{3.12}$$

3.2 Interpolation

Our goal is to compare solutions to the SDE (3.8) with initial data $s_i(0)$, a (g,G)-regular deterministic set of singular values, and $r_i(0)$, the singular values of the Gaussian ensemble G. We introduce times $t_0 = N^{-1+\omega_0}, t_1 = N^{-1+\omega_1}$ with parameters $0 < \omega_1 < \omega_0$ and such that $gN^{\sigma} \leq t_0 \leq N^{-\sigma}G^2$ and $2\omega_1 \leq \omega_0$.

For $\alpha \in [0,1]$ we introduce a continuous interpolation between $s_i(t)$ and $r_i(t)$:

$$dz_i(t,\alpha) = \frac{dB_i}{\sqrt{N}} + \frac{1}{2N}\sum_{j\neq \pm i,0} \frac{dt}{z_i(t,\alpha) - z_j(t,\alpha)}. \tag{3.13}$$

The initial values are given by

$$z_i(0,\alpha) = (1-\alpha)r_i(t_0) + \alpha s_i(t_0), \tag{3.14}$$

and the processes $z(t,0)$ and $z(t,1)$ are shown below to provide the desired coupling of the singular values of M_t and W_t.

More precisely, we construct solutions of (3.13) for a countable dense set of $\alpha \in [0,1]$ and use the continuity of the map from initial data to solution paths to construct solutions for the remaining α. This approach avoids the possibility that uncountably many sets of measure zero accumulate. See Appendix A.2 for details.

In particular, $z_i(t,0) = r_i(t_0 + t)$ and $z_i(t,1) = s_i(t_0 + t)$. The time shift is enforced to simplify the notation. Using the deformed law Theorem 4.5, we will establish rigidity of the particles after the short time t_0. Hence, the time shift means the $z_i(t)$ are rigid for times $t \geq 0$, which is notationally convenient. See Lemma 3.5 for the precise statement.

We define the $\gamma_i(t)$ to be the classical locations associated to the eigenvalue density of the symmetrization of the deformed matrix $V + \sqrt{t}W$, which has eigenvalues $\{s_i(t)\}_{i=-N}^{N}$. We let γ_i^{sc} denote the classical locations for the semicircle law.

3.2.1 Interpolating Measures

We would like to introduce a family of measures $\rho(E,t,\alpha)\,dE$ which will interpolate between the densities of the initial data. One approach is to take the free convolution of $z_i(0,\alpha)$ with the semicircle law. However, the resulting measures are not regular enough for our purposes. The following two lemmas assert that we can construct interpolating measures with better regularity properties. The first concerns the regularity of the

$\rho(E,t,\alpha)$, and the second verifies these measures interpolate between the densities of the two initial ensembles. Here $m(z,t,\alpha)$ is the Stieltjes transform of $\rho(E,t,\alpha)\,dE$, and we let $\gamma_i(t,\alpha)$ denote the classical locations for $\rho(E,T,\alpha)\,dE$.

The set $\mathcal{G}_\alpha$ used in the lemma is defined as follows. We fix $q^* \in (0,1)$ and set k_0 to be the largest index such that

$$|\gamma_{k_0}(t_0)| \leq q^* G, \quad |\gamma_{-k_0}(t_0)| \leq q^* G, \quad |\gamma^{\rm sc}_{k_0}| = |\gamma^{\rm sc}_{-k_0}| \leq q^* G. \tag{3.15}$$

Then $\mathcal{G}_\alpha$ is defined as

$$\mathcal{G}_\alpha = [\alpha\gamma_{-k_0}(t_0) + (1-\alpha)\gamma^{\rm sc}_{-k_0}, \alpha\gamma_{k_0}(t_0) + (1-\alpha)\gamma^{\rm sc}_{k_0}]. \tag{3.16}$$

Lemma 3.3. A family of interpolating measures $\rho(E,t,\alpha)\,dE$ exists (in a sense made precise by the following lemma) such that the following holds. Let $\delta > 0$. For $|E| \leq N^{-\delta}t_0$, $t \leq N^{-\delta}t_0$, and $N^{-1+\delta} \leq \eta \leq 10$ all of the following estimates are true with overwhelming probability.

$$|\partial_E \rho(E,t,\alpha)| \leq \frac{C}{t_0}, \quad \rho(0,0,\alpha) = \rho(0,0,0) = \rho_{\rm sc}(0) \tag{3.17}$$

$$|\rho(E,t,\alpha) - \rho_{\rm sc}(0)| \leq C\left(\frac{t\log(N)}{t_0} + \frac{|E|}{t_0}\right) \tag{3.18}$$

$$|\rho(E,t,\alpha) - \rho(0,t,\alpha) - (\rho(E,t,0) - \rho(0,t,0))| \leq C\left(\frac{t\log(N)}{t_0} + \frac{|E|}{t_0}\right) \tag{3.19}$$

Further, for $q \in (0,1)$ and $E \in q\mathcal{G}_\alpha$, $N^{-1+\delta} \leq \eta \leq 10$, with overwhelming probability

$$|m(z,t,\alpha)| \leq C\log(N), \quad c \leq \operatorname{Im} m(z,t,\alpha) \leq C. \tag{3.20}$$

For $q \in (0,1), 0 \leq t \leq t_1$, and $E \in q\mathcal{G}_\alpha$, with overwhelming probability

$$|\partial_z m(z,t,\alpha)| \leq \frac{C}{t_0 + \eta}. \tag{3.21}$$

Proof. We indicate where such claims are shown in [27], where the proofs are identical. The estimates on $\rho(E,t,\alpha)$ are in Lemma A.6. The first bound in (3.20) comes from the construction of $\rho(E,t,\alpha)$ and the reasoning in (7.12) of [28], and the second is Lemma A.4. The last claim is Lemma A.5. □

Define $d(i,j) = |i-j|$ if $ij > 0$ and $d(i,j) = |i-j| - 1$ if $ij < 0$. This is just to define an appropriate distance for our indexing, since no element is indexed with 0. The proof of the following lemma is the same as Lemma 3.4 in [27].

Lemma 3.4. The following estimates hold for the $\rho(E,t,\alpha)\,dE$ constructed in the previous lemma with overwhelming probability. We have for $\varepsilon > 0$ and $\omega_1 < \omega_0/2$,

$$\sup_{0\leq t\leq 10t_1} |\gamma_1(t,1) - \gamma_1(t_0+t)| \leq N^{-1-\omega_0/2+\omega_1+\varepsilon}, \tag{3.22}$$

$$\sup_{0\leq t\leq 10t_1} |\gamma_1(t,0) - 0| \leq N^{-1-\omega_0/2+\omega_1+\varepsilon}. \tag{3.23}$$

For $|j|,|k| \leq N^{\omega_0/2}$, and any choice of $t \leq 10t_1$, $\omega_1 \leq \omega_0/2$, $\alpha \in [0,1]$,

$$\gamma_k(t,\alpha) - \gamma_j(t,\alpha) = \frac{d(k,j)}{\rho_{\rm sc}(0)} + O(N^{-1}). \tag{3.24}$$

We further have an analogue of Lemma 3.5 in [27], giving rigidity and a local law. Here we choose a centered interval $\hat{C}_q$ of indices, asymptotically of size qGN, so particles with indices in $\hat{C}_q$ can be controlled uniformly in α. Precisely, we define k_1 to be the largest integer such that

$$\bigcup_{0\le\alpha\le1}[\alpha\gamma_{-k_1}(t_0)+(1-\alpha)\gamma^{\rm sc}_{-k_1},\alpha\gamma_{k_1}(t_0)+(1-\alpha)\gamma^{\rm sc}_{k_1}]\subset\bigcap_{0\le\alpha\le1}\mathcal{G}_\alpha\cap\{-\mathcal{G}_\alpha\},\qquad(3.25)$$

and set for $q\in(0,1)$

$$\hat{\mathcal{C}}_q=\{j\colon|j|\le qk_1\}.\qquad(3.26)$$

We let $m_N(z,t,\alpha)$ be the Stieltjes transform of the $z_i(t,\alpha)$.

Lemma 3.5. Fix $\varepsilon,\delta,\delta_1,D>0$ and $q\in(0,1)$. Then the following two estimates hold.

$$\mathbb{P}\left[\sup_{0\le t\le N^{-\delta_1}t_0}\sup_{i\in\hat{C}_q}\sup_{0\le\alpha\le1}|z_i(t,\alpha)-\gamma_i(t,\alpha)|\ge\frac{N^\varepsilon}{N}\right]\le N^{-D}\qquad(3.27)$$

$$\mathbb{P}\left[\sup_{N^{-1+\delta}\le\eta\le10}\sup_{0\le t\le N^{-\delta_1}t_0}\sup_{0\le\alpha\le1}\sup_{E\in q\mathcal{G}_\alpha}|m_N(z,t,\alpha)-m(z,t,\alpha)|\ge\frac{N^\varepsilon}{N\eta}\right]\le N^{-D}\qquad(3.28)$$

The key input to the proof of this lemma is a deformed local law for small times t. In the eigenvalue context this law was shown in [28]. We prove the needed singular value version of this result in Section 4. Excepting this change, the proof is identical to the one described in Appendices A and B of [27].

3.3 Short range equation

Define a centered process

$$\tilde{z}_i(t,\alpha)=z_i(t,\alpha)-\gamma_1(t,\alpha).\qquad(3.29)$$

Note this differs from (3.35) of [27] because we use the classical location γ_1 to center the particles, as we have no γ_0. The new classical locations are

$$\tilde{\gamma}_i(t,\alpha)=\gamma_i(t,\alpha)-\gamma_1(t,\alpha).\qquad(3.30)$$

We recall from [29] that

$$\partial_t\gamma_i(t,\alpha)=-\operatorname{Re}[m(\gamma_i(t,\alpha),t,\alpha)].\qquad(3.31)$$

Then the SDE that governs the $\tilde{z}_i$ is

$$d\tilde{z}_i(t,\alpha)=\frac{dB_i}{\sqrt{N}}+\left(\frac{1}{2N}\sum_{j\ne\pm i}\frac{1}{\tilde{z}_i(t,\alpha)-\tilde{z}_j(t,\alpha)}+\operatorname{Re}[m(\gamma_1(t,\alpha),t,\alpha)]\right)dt.\qquad(3.32)$$

Because we do not have good control over the extremal particles, and because we are interested only in particles near the origin, it is convenient to introduce a short range cutoff. We fix $q_*\in(0,1)$, and $\omega_l,\omega_A>0$ such that

$$0<\omega_1<\omega_\ell<\omega_A<\omega_0/2.\qquad(3.33)$$

We choose the parameters in this way so the error term in the forthcoming Lemma 3.6 is $o(1/N)$. Given $q\in(0,1)$, we define

$$A_q=\{(i,j)\colon|i-j|\le N^{\omega_l}\text{ or }ij>0,i\notin\hat{C}_q,j\notin\hat{C}_q\},\qquad(3.34)$$

and we let $A_{q_*,(i)}$ be the indices j such that $(i,j) \in A_{q_*}$ and $A^c_{q_*,(i)}$ be the indices j such that $(i,j) \notin A_{q_*}$.

Define $\hat{z}_i(t,\alpha)$ as the solution to, for $|i| \leq N^{\omega_A}$,

$$d\hat{z}_i(t,\alpha) = \frac{dB_i}{\sqrt{N}} + \frac{1}{2N} \sum_j^{A_{q_*,(i)}} \frac{1}{\hat{z}_i(t,\alpha) - \hat{z}_j(t,\alpha)}\, dt. \tag{3.35}$$

For $|i| > N^{\omega_A}$,

$$d\hat{z}_i(t,\alpha) = \frac{dB_i}{\sqrt{N}} + \frac{1}{2N} \sum_j^{A_{q_*,(i)}} \frac{1}{\hat{z}_i(t,\alpha) - \hat{z}_j(t,\alpha)} dt + \frac{1}{2N} \sum_j^{A^c_{q_*}} \frac{dt}{\tilde{z}_i - \tilde{z}_j} + \mathrm{Re}[m(\gamma_1(t,\alpha),t,\alpha)]\, dt. \tag{3.36}$$

Here the initial condition is $\hat{z}_i(0) = \tilde{z}_i(0)$.

The $\hat{z}_i$ are good approximations to the $\tilde{z}_i$. This is the content of the following lemma, whose proof is identical to the proof of Lemma 3.7 in [27].

Lemma 3.6. Fix $\varepsilon, D > 0$. Then, for large N,

$$\mathbb{P}\left[\sup_{0\leq t\leq t_1} \sup_i \sup_{0\leq\alpha\leq 1} |\hat{z}_i(t,\alpha) - \tilde{z}_i(t,\alpha)| \geq N^{\varepsilon} t_1 \left(\frac{N^{\omega_A}}{N^{\omega_0}} + \frac{1}{N^{\omega_\ell}} + \frac{1}{\sqrt{NG}} \right)\right] \leq N^{-D}. \tag{3.37}$$

3.3.1 Short Range Kernel

Here we introduce a coupled parabolic equation with no Brownian motion term. We define $u_i = \partial_\alpha \hat{z}_i(t,\alpha)$, where the differentiation is justified in Appendix A.2.

We have

$$\partial_t u_i = \sum_j^{A_{q_*,(i)}} B_{ij}(u_j - u_i) + \xi_i = -(\mathcal{B}u)_i + \xi_i, \tag{3.38}$$

where

$$B_{ij} = \frac{\mathbb{1}_{i\neq \pm j}}{2N(\hat{z}_i - \hat{z}_j)^2}, \tag{3.39}$$

the operator $\mathcal{B}$ is implicitly defined by the above equation, and ξ_i is an error term that vanishes for $|i| \leq N^{\omega_A}$. Because it vanishes for small values of i, we will show its effect on particles near the origin is negligible.

We let $\mathcal{U}$ be the semigroup associated to $\mathcal{B}$. Its elements $\mathcal{U}_{ij}(s,t)$ are defined so that, if $v(t)$ is any solution of the system $\partial_t v_i = -(\mathcal{B}v)_i$, then for any times $t, s \geq 0$,

$$v_i(t) = \sum_{j=-N}^{N} \mathcal{U}_{ij}(s,t) v_j(s). \tag{3.40}$$

Note that the $\mathcal{U}_{ij}$ are random.

3.3.2 Finite speed estimates

We state two results on the decay of the semigroup elements $\mathcal{U}_{ij}$. Their proofs are straightforward adaptations of those given in Section 4 of [27] for the kernel $\mathcal{U}^{(B)}$ in that reference.

Lemma 3.7. Let $0 \leq s \leq t \leq t_1$. Fix $0 < q_1 < q_2 < q_*$ and $D, \varepsilon > 0$. For every α there exists an event $\mathcal{F}_\alpha$ such that $\mathbb{P}(\mathcal{F}_\alpha) \geq 1 - N^{-D}$ and on which the following estimates hold. If $i \in \hat{\mathcal{C}}_{q_2}$ and $0 \leq s \leq t \leq 10t_1$, then

$$|\mathcal{U}_{ji}(s,t,\alpha)| \leq N^{-D}, \quad |i-j| > N^{\omega_\ell + \varepsilon}. \tag{3.41}$$

If $i \notin \hat{\mathcal{C}}_{q_2}$, $j \in \hat{\mathcal{C}}_{q_1}$, and $0 \le s \le t \le 10t_1$, then

$$|\mathcal{U}_{ji}(s,t,\alpha)| \le N^{-D}. \tag{3.42}$$

Lemma 3.8. Let $0 < q_1 < q_*$ and $D, \varepsilon > 0$. For every α there exists an event $\mathcal{F}_\alpha$ such that $\mathbb{P}(\mathcal{F}_\alpha) \ge 1 - N^{-D}$ and on which he following bound holds. For $i, j \in \hat{\mathcal{C}}_{q_1}$ and $0 \le s \le t \le 10t_1$,

$$\mathcal{U}_{ij}(s,t) \le \frac{N^\varepsilon}{N} \frac{(t-s) \vee N^{-1}}{\left(\frac{i-j}{N}\right)^2 + ((t-s) \vee N^{-1})^2}. \tag{3.43}$$

3.4 *A priori* estimates

We first make some definitions necessary for the homogenization argument. We fix a constant $\varepsilon_B > 0$ such that $\omega_A - \varepsilon_B > \omega_l$, and fix an integer a such that $0 < |a| \le N^{\omega_A - \varepsilon_B}$. We also define the deterministic particle locations $\gamma_j^{\mathrm{f}} = j(2N\rho_{\mathrm{sc}}(0))^{-1}$.

We consider a particular solution w of the short range equation that will play a key role:

$$\partial_t w_i = -(\mathcal{B}w)_i, \quad w_i(0) = 2N\delta_a(i). \tag{3.44}$$

Define the cutoff $\eta_l = N^{\omega_\ell}(2N\rho_{\mathrm{sc}}(0))^{-1}$. We will need the kernel $p_t(x,y)$ of the following equation

$$\partial_t f(x) = \int_{|x-y| \le \eta_\ell} \frac{f(y) - f(x)}{(x-y)^2} \rho_{\mathrm{sc}}(0)\, dy. \tag{3.45}$$

The main result of this subsection is that this kernel is a good approximation to elements of the short range semigroup. We choose parameters s_0 and s_1 such that

$$N^{-1} \ll s_0 \ll s_1 \ll t_1 \ll t_0 \tag{3.46}$$

and introduce

$$f(x,t) = \frac{1}{2N} \sum_{j \ne 0} p_{s_0+t-s_1}(x, \gamma_j^{\mathrm{f}}) w_j(s_1), \quad f_i(t) = f(\hat{z}_i(t,\alpha), t). \tag{3.47}$$

We now collect some bounds on these objects. We let $p_t^{(k)}(x,y)$ denote the derivative in x. The following lemma is Lemma 3.11 in [27].

Lemma 3.9. Fix $\varepsilon_1, \varepsilon_2, D_1 > 0$. For t such that $N^{-D_1} \le t \le N^{-\varepsilon_1}\eta_\ell$ and any $D_2 > 0$,

$$p_t(x,y) \le C\frac{t}{(x-y)^2 + t^2}, \quad p_t^{(k)}(x,y) \le \frac{C}{t^k} p_t(x,y) + N^{-D_2}, \quad |\partial_t p_t(x,y)| \le \frac{C}{x^2+y^2} + N^{-D_2}. \tag{3.48}$$

If additionally $|x-y| > N^{\varepsilon_2}\eta_l$,

$$p_t(x,y) \le N^{-D_2}, \quad p_t^{(k)}(x,y) \le N^{-D_2}. \tag{3.49}$$

Lemma 3.10. Fix $\varepsilon > 0$ and $t > 0$. The following estimates hold with overwhelming probability.

$$\|w(t)\|_1 \le 2N, \quad \|w(t)\|_\infty \le N^\varepsilon t^{-1}, \quad \|w(t)\|_2^2 \le N^\varepsilon t^{-1}. \tag{3.50}$$

Additionally, if $|i| \ge |a| + N^{\omega_l + \varepsilon_1}$ then for any $D > 0$ we have with overwhelming probability

$$|w_i(t)| \le N^{-D}. \tag{3.51}$$

Proof. We see from the definition of $\mathcal{B}$ and examining $\min w_i$ that all $w_i(t)$ are non-negative for $t \ge 0$. Hence

$$\|w(t)\|_1 = \|w(0)\|_1 = 2N. \tag{3.52}$$

The second estimate is a consequence of Lemma 3.8 and the third estimate is (3.109) in [27]. The last estimate follows from applying Lemma 3.7. □

Lemma 3.11. We have

$$f(t,\hat{z}_i) \le \frac{C}{t+s_1}, \tag{3.53}$$

and for any $\varepsilon_1 > 0$, if $|i| \ge |a| + N^{\omega_\ell + \varepsilon_1}$,

$$|f_i| \le N^{-D}. \tag{3.54}$$

For $k \ge 1$,

$$f^{(k)}(t,\hat{z}_i) \le \frac{C}{t-s_1+s_0} f(t,\hat{z}_i) + N^{-D}. \tag{3.55}$$

Proof. See the beginning of the proof of Lemma 3.13 in [27]. □

3.5 Semigroup estimate

In the lemma below, we temporarily break with our convention of omitting the zero index in order to match the notation of [27]. The strange indexing is due to the fact that we will need to apply it in the context of an even number of particles labeled $-N$ to -1 and 1 to N. We think of shifting the positive indices of w and f down 1, so we will define for $i \ge 0$,

$$w_i^{\mathrm{new}} = w_{i+1} \tag{3.56}$$

and for $i \le -1$,

$$w_i^{\mathrm{new}} = w_i \tag{3.57}$$

Then the largest positive index in the following sum will not have a matching negative index, but we will see this makes no difference to our application below.

Lemma 3.12. For any $\varepsilon_1, \varepsilon_2, D > 0$ and α, there exists an event $\mathcal{F}_\alpha$ such that $\mathbb{P}(\mathcal{F}_\alpha) \ge 1 - N^{-D}$ and on which the following estimate holds.

$$\|w^{\mathrm{new}}(s_1) - f^{\mathrm{new}}(s_1)\|_2^2 \le Cs_0 \sum_{|i| \le N^{\omega_A - \varepsilon_B} + N^{\omega_\ell + \varepsilon_2}} \sum_{|i-j| \le \ell, j \ne i, -(i+1)} \frac{(w_i^{\mathrm{new}}(s_1) - w_j^{\mathrm{new}}(s_1))^2}{(i-j)^2} \tag{3.58}$$

$$+ \frac{N^{\varepsilon_1}}{s_1}\left(\frac{1}{(Ns_0)^2} + \frac{(Ns_0)^2}{\ell^2}\right) \tag{3.59}$$

Proof. We omit the superscript for this proof. Lemma 3.12 in [27] gives

$$\|w(s_1) - f(s_1)\|_2^2 \le Cs_0 \sum_{|i| \le N^{\omega_A - \varepsilon_B} + N^{\omega_\ell + \varepsilon_2}} \sum_{|i-j| \le \ell, j \ne i} \frac{(w_i(s_1) - w_j(s_1))^2}{(i-j)^2}$$
$$+ \frac{N^{\varepsilon_1}}{s_1}\left(\frac{1}{(Ns_0)^2} + \frac{(Ns_0)^2}{\ell^2}\right). \tag{3.60}$$

Note the additional terms in the sum where $j = -(i+1)$. We separate them out and bound them.

$$s_0 \sum_{|i| \le N^{\omega_A - \varepsilon_B} + N^{\omega_\ell + \varepsilon_2}} \sum_{j=-(i+1)} \frac{(w_i(s_1) - w_j(s_1))^2}{(i-j)^2} \le s_0 \sum_i (w_i(s_1) - w_{-i+1}(s_1))^2$$
$$\le s_0 C\|w(s_1)\|_2^2 \le N^{\varepsilon} s_0 s_1^{-1}. \tag{3.61}$$

In the last inequality we used Lemma 3.10, and $\varepsilon > 0$ is arbitrary. Now we are done, because

$$\frac{s_0}{s_1} \ll \frac{1}{s_1}\frac{1}{(Ns_0)^2}. \tag{3.62}$$

□

The next lemma is the key estimate for the homogenization. It is here we deal with the missing repulsion term in the symmetrized equation (3.8).

Lemma 3.13. For $t \geq s_1$, the Itô differential of $\|w(t) - f(t)\|_2^2$ takes the form

$$d\frac{1}{N}\sum_i (w_i - f_i)^2 = -\langle (w(t) - f(t)), \mathcal{B}(w(t) - f(t))\rangle \, dt + X_t \, dt + dM_t, \tag{3.63}$$

where M_t is a martingale and X_t is a process defined by the above equality.

Fix $\varepsilon, D > 0$. Then additionally for every α there is an event $\mathcal{F}_\alpha$ such that $\mathbb{P}(F_\alpha) \geq 1 - N^{-D}$ and on which the following bounds hold. For $s_1 \leq t \leq 9t_1$,

$$|X_t| \leq \frac{1}{5}\langle (w(t) - f(t)), \mathcal{B}(w(t) - f(t))\rangle + \frac{C}{t + s_1}\frac{N^\varepsilon}{t - s + s_0}\left(\frac{1}{\sqrt{N(t - s_1 + s_0)}}\right). \tag{3.64}$$

For any u_1 and u_2 with $9t_1 > u_2 > u_1 \geq s_1$,

$$\left|\int_{u_1}^{u_2} dM_t\right| \leq \frac{N^\varepsilon}{N}\frac{1}{(u_1 + s_1)^{3/2}(u_1 - s_1 + s_0)^{1/2}}. \tag{3.65}$$

Proof. This result is the analogue of Lemma 3.13 in [27], and most of the proof goes through with only notational changes. We comment on one important difference. The term (3.137) becomes, using the notation of that reference,

$$\left(\frac{1}{2N}\sum_{j \neq \pm i}^{A_{q_*}\setminus A_{2,(i)}} \frac{f_j - f_i}{(\hat{z}_i - \hat{z}_j)^2} - \int_{\eta_{\ell,2} \leq |y - \hat{z}_i| \leq \eta_\ell} \frac{f(y) - f(\hat{z}_i)}{(\hat{z}_i - y)^2}\rho_{\mathrm{sc}}(0)\, dy\right) \tag{3.66}$$

The difference is that if $-i \in A_{q_*} \setminus A_{2,(i)}$ we omit the $j = -i$ term. We will show in this case that the term is negligible, so that it can be reinserted and the proof can be completed in the same way.

By the mean value theorem, there exists $\xi \in \mathbb{R}$ such that

$$\frac{1}{2N}\frac{f_{-i} - f_i}{(\hat{z}_{-i} - \hat{z}_i)^2} = \frac{1}{2N}\left(\frac{f'_{-i}}{(\hat{z}_{-i} - \hat{z}_i)} + \frac{f''_{-i}(\xi)}{2}\right). \tag{3.67}$$

Then, using Lemma 3.5, Lemma 3.11, and the hypothesis on i, we have the bound

$$\left|\frac{1}{2N}\frac{f_{-i} - f_i}{(\hat{z}_{-i} - \hat{z}_i)^2}\right| \leq \frac{1}{2N}\left(N\frac{N^\varepsilon}{N^{\omega_{\ell,2}}}\frac{1}{t - s_1 + s_0}\frac{1}{t + s_1} + \frac{1}{(t - s_1 + s_0)^2}\frac{1}{t + s_1}\right). \tag{3.68}$$

This is the same size as the error obtained for (3.137) and (3.138) in the bound (3.142) of [27]. □

We obtain the following lemma by integrating.

Corollary 3.14. Fix $\varepsilon, D > 0$. For each α there exists an event $\mathcal{F}_\alpha$ such that $\mathbb{P}(\mathcal{F}_\alpha) \geq 1 - N^{-D}$ and on which the following estimate holds.

$$\int_{s_1}^{2t_1} \langle (w - f), \mathcal{B}(w - f)\, dx\rangle \leq C\|(w(s_1) - f(s_2)\|_2^2 + \frac{N^\varepsilon}{s_1}\frac{1}{\sqrt{Ns_0}} \tag{3.69}$$

We now obtain a time-averaged version of our desired result.

Theorem 3.15. Fix a and i such that $|a| \leq N^{\omega_A - \varepsilon_B}$ and $|i - a| \leq \ell/10$. Fix also $\varepsilon, D > 0$. For every α there is an event $\mathcal{F}_\alpha$ such that $\mathbb{P}(\mathcal{F}_\alpha) \geq 1 - N^{-D}$ and on which

$$\frac{1}{t_1}\int_0^{t_1}\left(U_{t_1+u}(i,a) - \frac{1}{N}p_{t_1+u}(\gamma_i^{\mathrm{f}}, \gamma_a^{\mathrm{f}})\right)^2 du \leq \frac{N^\varepsilon}{(Nt_1)^2}\left(\frac{(Nt_1)^4}{\ell^4} + \frac{s_1^2}{t_1^2} + \frac{t_1}{s_1}\left(\frac{1}{\sqrt{Ns_0}} + \frac{s_0}{s_1}\right)\right) \tag{3.70}$$

for $0 \leq u \leq t_1$.

Proof. We follow the proof of Theorem 3.15 in [27]. There are two changes. First, we must modify the bound in display (3.165) to use our kernel $\mathcal{B}$. Recall that we omit the index $i = 0$ in w and f. As in our discussion of Lemma 3.12, we shift the positive indices down one in order to apply a Sobolev inequality in Appendix D of [27]. As in the proof of Theorem 3.15 of that reference, we apply the Sobolev inequality to obtain

$$\frac{1}{4N^2}(w_{t_1+u}(i) - f_{t_1+u}(i))^2 \le C\left(\frac{1}{N\ell}\sum_{|j-i|\le\ell} w_{t_1+u}(j) - \frac{1}{N\ell}\sum_{|j-i|\le\ell} f_{t_1+u}(j)\right)^2 \tag{3.71}$$

$$+ C\log(N)N^{-2}\left(\sum_{|i|,|j|\le l, j\neq i,-(i+1)} \frac{[(w-f)_i(t_1+u) - (w-f)_j(t_1+u)]^2}{(i-j)^2}\right) \tag{3.72}$$

$$+ C\log(N)N^{-2}\sum_i [(w-f)_i(t_1+u) - (w-f)_{-i}(t_1+u)]^2. \tag{3.73}$$

The first two terms are bounded as in the proof in [27]. In particular, using rigidity, we see the second is bounded by

$$\frac{N^\varepsilon}{N^2}\langle (w-f)(t_1+u), \mathcal{B}(w-f)(t_1+u)\rangle. \tag{3.74}$$

The third term is, using Lemma 3.10 and the analogous bound for $\|f\|^2$, bounded by

$$\frac{C\log N}{N^2}\|(w-f)(t_1+u)\|^2 \le \frac{C\log N}{N^2}\left(\|w(t_1+u)\|_2^2 + \|f(t_1+u)\|_2^2\right) \le \frac{N^\varepsilon}{N^2}\left(\frac{1}{t_1} + \frac{1}{s_1}\right). \tag{3.75}$$

This error is bounded by $N^{-1+\delta}$ for some small $\delta > 0$, which smaller than the errors obtained when bounding (3.165) in [27]. Hence replacing the old kernel with the new one in the proof is permissible.

Second, in (3.170) we apply our Lemma 3.12 and change (3.171) to use our $\mathcal{B}$. □

The removal of the time average as in Theorem 3.16 of [27] goes through without change, since it only uses the abstract properties of $\mathcal{U}$ and the bounds in Lemma 3.9.

Theorem 3.16. Fix a and i such that

$$|a| \le \frac{N^{\omega_A - \varepsilon_B}}{2}, \quad |i-a| \le \frac{\ell}{20}. \tag{3.76}$$

Fix also $\varepsilon, D > 0$. There exists an event $\mathcal{F}_\alpha$ such that $\mathbb{P}(F_\alpha) \ge 1 - N^{-D}$ and on which the following bound holds.

$$\left|\mathcal{U}_{t_1+2t_2}(i,a) - \frac{1}{N}p_{t_1}(\gamma_i^{\mathrm{f}}, \gamma_a^{\mathrm{f}})\right| \le CN^{\varepsilon+\varepsilon_2-1}t_1^{-1}\left(\frac{s_1^2}{t_1^2} + \frac{(Nt_1)^4}{\ell^4} + \frac{t_1}{s_1}\left[\frac{1}{\sqrt{Ns_0}} + \frac{s_0}{s_1}\right]\right)^{1/2}$$
$$+ N^{\varepsilon+\varepsilon_2/2-1}t_1^{-1} \tag{3.77}$$

Then as in the proof of Theorem 3.10 in [27], choosing s_0 and s_1 in Theorem 3.16 such that $Ns_0 = (Ns_1)^{2/3}$ and $Ns_1 = (Nt_1)^{9/10}$ yields the final homogenization result.

Theorem 3.17. Fix a and i such that $|a| \le N^{\omega_A - \varepsilon_B}$ and $|i-a| \le \ell/10$. With t_1 as above and $t_2 = N^{-\varepsilon_2}t_1$ (for ε_2 satisfying $\omega_1 - \varepsilon_2 > 0$), and fixed $\varepsilon, D > 0$, there exists for every α an event $\mathcal{F}_\alpha$ with $\mathbb{P}(\mathcal{F}_\alpha) \ge 1 - N^{-D}$ on which the following holds.

$$\left|\mathcal{U}_{ia}(0,t_1) - \frac{1}{N}p_{t_1}(\gamma_i^{\mathrm{f}}, \gamma_j^{\mathrm{f}})\right| \le \frac{N^{\varepsilon+\varepsilon_2}}{Nt_1}\left(\frac{(Nt_1)^2}{\ell^2} + \frac{1}{(Nt_1)^{1/10}}\right) + \frac{N^{\varepsilon-\varepsilon_2/2}}{Nt_1} \tag{3.78}$$

3.6 Conclusion

Here we will frequently need to use the above bounds, which hold for fixed α on a set of large probability, for all $\alpha \in [0,1]$ simultaneously when bounding an integrand. This is accomplished using Lemma E.1 and the Remark that follows it in Appendix E of [27].

Proof of Theorem 3.2. By Lemma 3.6,

$$z_i(t_1,1) - z_i(t_1,0) = \hat{z}_i(t_1,1) - \hat{z}_i(t_1,0) + (\gamma_1(t_1,1) - \gamma_1(t_1,0)) \\ + O\left(N^{\varepsilon} t_1 \left(\frac{N^{\omega_A}}{N^{\omega_0}} + \frac{1}{N^{\omega_\ell}} + \frac{1}{\sqrt{NG}}\right)\right). \tag{3.79}$$

The last term is $o(1)$, and by Lemma 3.4 the difference $(\gamma_1(t_1,1) - \gamma_1(t_1,0))$ is also $o(1)$ for $j,k \le N^{\omega_0/2}$. Hence it suffices to bound the first term.

Recalling $u_i = \partial_\alpha \hat{z}_i$, we have

$$\hat{z}_i(t_1,1) - \hat{z}_i(t_1,0) = \int_0^1 u_i(t_1,\alpha)\, d\alpha. \tag{3.80}$$

We now use the finite speed of propagation estimates to show that the initial data far from zero makes a negligible contribution to $u_i(t_1,\alpha)$. We perform this estimate in two steps. First, we remark that u that is a solution of

$$\partial_t u = \mathcal{B} u + \xi, \tag{3.81}$$

where ξ satisfies the bound

$$|\xi_i(t)| \le \mathbb{1}_{\{|i| > N^{\omega_A}\}} N^C \tag{3.82}$$

with overwhelming probability for $0 \le t \le 1$. We define v_i as the solution

$$\partial_t v = \mathcal{B} v, \quad v_i(0) = u_i(0). \tag{3.83}$$

Then by the Duhamel formula,

$$u_i(t_1) - v_i(t_1) = \int_0^1 \sum_{|p| \le N} \mathcal{U}_{ip}(s,t_1)\xi_p(s)\, ds = \int_0^1 \sum_{N^{\omega_A} < |p| \le N} \mathcal{U}_{ip}(s,t_1)\xi_p(s)\, ds. \tag{3.84}$$

We fix $\delta_B > 0$ and consider i such that $|i| \le N^{\omega_A - \delta_B}$. By Lemma 3.7 and (3.82), we obtain

$$|u_i(t_1) - v_i(t_1)| \le N^{-10}. \tag{3.85}$$

For the second step, we fix $\varepsilon_a > 0$ and consider w defined as

$$\partial_t w = \mathcal{B} w, \quad w_i(0) = v_i(0)\mathbb{1}_{\{|i| \le N^{1+\varepsilon_a} t_1\}}. \tag{3.86}$$

We have

$$v_i(t_1) - w_i(t_1) = \sum_{N^{\omega_1 + \varepsilon_a}} \mathcal{U}_{ij}(0,t_1) u_j(0). \tag{3.87}$$

As in the first step, Lemma 3.7 shows the terms with $|j| > N^{\omega_A}$ are negligible. Fix $\varepsilon_b > 0$ such that $\varepsilon_b < \varepsilon_a$. Then by Lemma 3.8,

$$|v_i(t_1) - w_i(t_1)| \le \left| \sum_{N^{\omega_1+\varepsilon_a} < |j| \le N^{\omega_A}} \mathcal{U}_{ij}(0,t_1) u_j(0) \right| \le N^{\varepsilon + \omega_1} \sum_{|j| > N^{\omega_1 + \varepsilon_a}} \frac{1}{(i-j)^2} \le N^{-1-\varepsilon_a+\varepsilon}. \tag{3.88}$$

We then have

$$\left|\int_0^1 u_i(t_1,\alpha)\,d\alpha\right| \leq \left|\int_0^1 \sum_{|j|\leq Nt_1N^{\varepsilon_a}} \mathcal{U}(0,t_1,\alpha)_{ij}u_j(0)\,d\alpha\right| + N^{-1-\varepsilon_a+\varepsilon}. \tag{3.89}$$

Therefore, it suffices to estimate

$$\int_0^1 \sum_{|j|\leq Nt_1N^{\varepsilon_a}} \mathcal{U}(0,t_1,\alpha)_{ij}(z_j(0,1)-z_j(0,0))\,d\alpha. \tag{3.90}$$

Recall that by hypothesis the initial data are symmetric, $z_j = -z_{-j}$. Hence we just need to estimate

$$\int_0^1 \sum_{0<j\leq Nt_1N^{\varepsilon_a}} (\mathcal{U}_{ij}(0,t_1) - \mathcal{U}_{i,-j}(0,t_1)(z_j(0,1)-z_j(0,0))\,d\alpha. \tag{3.91}$$

By Theorem 3.17 we can replace $\mathcal{U}_{ij}(0,t_1)$ with $\frac{1}{N}p_{t_1}(\gamma_i^{\mathrm{f}},\gamma_j^{\mathrm{f}})$ and accrue an error of

$$E_0 = \frac{N^{\varepsilon+\varepsilon_2}}{Nt_1}\left(\frac{(Nt_1)^2}{\ell^2} + \frac{1}{(Nt_1)^{1/10}}\right) + \frac{N^{\varepsilon-\varepsilon_2/2}}{Nt_1}. \tag{3.92}$$

Then

$$(\mathcal{U}_{ij}(0,t_1) - \mathcal{U}_{i,-j}(0,t_1) \leq \frac{1}{N}(p_{t_1}(\gamma_i^{\mathrm{f}},\gamma_j^{\mathrm{f}}) - p_{t_1}(\gamma_i^{\mathrm{f}},-\gamma_j^{\mathrm{f}})) + E_0 \tag{3.93}$$

$$= \frac{1}{N}(p_{t_1}(\gamma_i^{\mathrm{f}},\gamma_j^{\mathrm{f}}) - p_{t_1}(-\gamma_i^{\mathrm{f}},\gamma_j^{\mathrm{f}})) + E_0 = \frac{1}{N}\int_{-\gamma_i^{\mathrm{f}}}^{\gamma_i^{\mathrm{f}}} p'_{t_1}(x,\gamma_j^{\mathrm{f}})\,dx + E_0. \tag{3.94}$$

Summing over $j \leq Nt_1N^{\varepsilon_a}$, then using Lemma 3.9 and the normalization of p_t gives

$$\begin{aligned}
\sum_{0<j\leq Nt_1N^{\varepsilon_a}} (\mathcal{U}_{ij}(0,t_1) - \mathcal{U}_{i,-j}(0,t_1) &\leq \sum_{0<j\leq Nt_1N^{\varepsilon_a}} E_0 + \frac{1}{N}\int_{-\gamma_1^{\mathrm{f}}}^{\gamma_1^{\mathrm{f}}} p'_{t_1}(x,\gamma_j^{\mathrm{f}}) \\
&\leq Nt_1N^{\varepsilon_a}E_0 + \frac{1}{t_1}\int_{-\gamma_1^{\mathrm{f}}}^{\gamma_1^{\mathrm{f}}} \sum_{0<j\leq Nt_1N^{\varepsilon_a}} \frac{1}{N}p_{t_1}(x,\gamma_j^{\mathrm{f}})\,dx \\
&\leq Nt_1N^{\varepsilon_a}E_0 + \frac{C}{t_1}\int_{-\gamma_1^{\mathrm{f}}}^{\gamma_1^{\mathrm{f}}} 1\,dx \\
&\leq Nt_1N^{\varepsilon_a}E_0 + \frac{N^{\varepsilon}}{Nt_1}.
\end{aligned} \tag{3.95}$$

Choosing $\varepsilon, \varepsilon_2, \varepsilon_a > 0$ small enough shows that for large enough N,

$$\sum_{0<j\leq Nt_1N^{\varepsilon_a}} (\mathcal{U}_{ij}(0,t_1) - \mathcal{U}_{i,-j}(0,t_1) \leq N^{-\delta} \tag{3.96}$$

for some small $\delta > 0$. Finally, using Lemma 3.5 we have with overwhelming probability

$$\sup_{j\leq Nt_1N^{\varepsilon_a}} |z_j(0,1) - z_j(0,0)| \leq N^{-1+\varepsilon} \tag{3.97}$$

for arbitrarily small $\varepsilon > 0$. Hence

$$\begin{aligned}
&\int_0^1 \sum_{|j|\leq Nt_1N^{\varepsilon_a}} \mathcal{U}(0,t_1,\alpha)_{ij}(z_j(0,1)-z_j(0,0))\,d\alpha \\
&\qquad \leq \int_0^1 \sum_{|j|\leq Nt_1N^{\varepsilon_a}} |\mathcal{U}(0,t_1,\alpha)_{ij}||(z_j(0,1)-z_j(0,0))|\,d\alpha \leq N^{-1-\delta'}
\end{aligned} \tag{3.98}$$

for some $\delta' > 0$. Undoing the time shift in the definition of $z_i(t,\alpha)$ completes the proof. □

4 Deformed local law

In this section we prove the deformed local law, Theorem 4.5, necessary for the proof of Lemma 3.5. A deformed local law for eigenvalues was previously shown in [28].

The structure of this section is as follows. In Subsection 4.1 we define some notions necessary for the rest of the section, state the main result, and compute Green function elements. In Subsection 4.2 we establish some preliminary estimates, then prove Lemma 4.9, a weak version of the deformed local law. Finally, in Subsection 4.3, we use a fluctuation averaging argument to upgrade this weak law and prove Theorem 4.5.

4.1 Preliminaries

4.1.1 Deformed Semicircle Law

It is well known that in the large N limit, many symmetric matrix ensembles have a macroscopic eigenvalue density that obeys the semicircle law, whose density is

$$\rho_{\mathrm{sc}}(x) = \mathbb{1}_{\{|x|<2\}} \frac{1}{2\pi}\sqrt{4-x^2}.$$

The *free convolution* of the semicircle law and a deterministic diagonal matrix $\overline{V} = (v_i)_{i=-N}^{N}$, with the $i=0$ index omitted, is defined by its Stieltjes transform,

$$m^{(N)}_{\mathrm{fc},t}(z) = \frac{1}{2N}\sum_{i=-N}^{N} \frac{1}{v_i - z - t m^{(N)}_{\mathrm{fc},t}(z)},$$

where again the $i=0$ term is omitted in the summation. There is a unique solution to this equation, and it is the Stieltjes transform of a measure absolutely continuous with respect to Lebesgue measure. We call the associated density $\rho^{(N)}_{\mathrm{fc},t}$. These facts and basic properties of the free convolution are proved in [8]. We often write $m_{\mathrm{fc},t}$ and $\rho_{\mathrm{fc},t}$ for these quantities, omitting dependence on N.

We now state a result on the stability of $m_{\mathrm{fc},t}$, proved in Lemma 7.2 of [28]. Recall the notion of (g,G)-regularity introduced in Definition 3.1.

Lemma 4.1. Assume that V is (g,G)-regular. For $q\in(0,1)$, $\sigma>0$, and N large enough, the following statements hold for $E\in(-qG,qG)$, $\eta\in[N^{-5},10]$, and t such that $gN^{\sigma}\le t\le N^{-\sigma}G^2$. The constants do not depend on σ or q.

(i) We have

$$c \le \operatorname{Im} m_{\mathrm{fc},t}(z) \le C \tag{4.1}$$

and hence

$$ct \le |v_i - z - t m_{\mathrm{fc},t}(z)| \tag{4.2}$$

for all v_i. Both statements hold uniformly in N.

(ii) We have

$$c \le \left| 1 - \frac{t}{2N}\sum_{-N}^{N} \frac{1}{(v_i - z - t m_{\mathrm{fc},t}(z))^2} \right| \le C$$

uniformly in N.

4.1.2 Stochastic Domination

We recall the notion of stochastic domination introduced in [17].

Definition 4.2. Let

$$X = (X^{(N)}(u)\colon N \in \mathbb{N}, u \in U^{(N)}), \quad Y = (Y^{(N)}(u)\colon N \in \mathbb{N}, u \in U^{(N)}) \tag{4.3}$$

be two sets of nonnegative random variables, where $U^{(N)}$ is a possibly N-dependent parameter set. We say that X is *stochastically dominated by Y, uniformly in u*, if for all $\varepsilon > 0$ and $D > 0$ we have

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[X^{(N)}(u) > N^{\varepsilon} Y^{(N)}(u)\right] \le N^{-D} \tag{4.4}$$

for large enough $N \ge N_0(\varepsilon, D)$. The stochastic domination is always uniform in all parameters that are not explicitly stated.

If X is stochastically dominated by Y, uniformly in u, we write $X \prec Y$. If for some complex family X we have $|X| \prec Y$, we also write $X = O_{\prec}(Y)$.

Observe that a sequence of events $E = (E^{(N)})$ holds with overwhelming probability if $1 - \mathbb{1}(E) \prec 0$. We also recall the basic properties of $\prec$.

Lemma 4.3. (i) Suppose that $X(u,v) \prec Y(u,v)$ uniformly in $u \in U$ and $v \in V$. If $|V| \le N^C$ for some constant C, then

$$\sum_{v \in V} X(u,v) \prec \sum_{v \in V} Y(u,v)$$

uniformly in u.

(ii) Suppose that $X_1(u) \prec Y_1(u)$ uniformly in u and $X_2(u) \prec Y_2(u)$ uniformly in u. Then $X_1(u)X_2(u) \prec Y_1(u)Y_2(u)$ uniformly in u.

(iii) if $X \prec Y + N^{-\varepsilon}X$ for some $\varepsilon > 0$, then $X \prec Y$.

The following large deviations estimates will be important for our work. Proofs may be found in, for example, [7].

Lemma 4.4. Let $\left(X_i^{(N)}\right)$, $\left(Y_i^{(N)}\right)$, $\left(a_{ij}^{(N)}\right)$, and $\left(b_{ij}^{(N)}\right)$ be independent families of random variables, where $N \in \mathbb{N}$ and $i, j \in \{1, \dots, N\}$. Suppose that all entries $X_i^{(N)}$ and $Y_i^{(N)}$ are independent and satisfy

$$\mathbb{E}X = 0, \quad \|X\|_p \le \mu_p$$

for all $p \in \mathbb{N}$ with some constants μ_p. In the following statements, Ψ can be any random variable.

(i) Suppose that $\left(\sum_i |b_i|^2\right)^{1/2} \prec \Psi$. Then $\sum_i b_i X_i \prec \Psi$.

(ii) Suppose that $\left(\sum_{i \ne j} |a_{ij}|^2\right)^{1/2} \prec \Psi$. Then $\sum_{i \ne j} a_{ij} X_i X_j \prec \Psi$.

(iii) Suppose that $\left(\sum_{i,j} |a_{ij}|^{1/2}\right)^{1/2} \prec \Psi$. Then $\sum_{i,j} a_{ij} X_i Y_j \prec \Psi$.

If all of the above random variables depend on an index u and the hypotheses of (i) – (iii) are uniform in u, then so are the conclusions.

4.1.3 Model and main result

We consider as in Section 3 a deterministic initial data matrix V. We let W be a matrix of i.i.d. random variables with distribution $\mathcal{N}(0, N^{-1})$ and define

$$H_t = V + \sqrt{t} W, \quad M_t = \begin{bmatrix} 0 & H_t \\ H_t^{\dagger} & 0 \end{bmatrix}. \tag{4.5}$$

Recall the eigenvalues of M_t, which we will call $\lambda_i(t)$, are the singular values of of H_t and their negatives. We also define

$$m_N(z) = \frac{1}{N} \sum_{i=-N}^{N} \frac{1}{\lambda_i(t) - z}. \tag{4.6}$$

suppressing the dependence of m_N on t in our notation.

In what follows we fix $\delta_1 > 0$ and always suppose that V is (g, G)-regular, as defined in Definition 3.1. We also suppose $z \in \mathcal{D}$ and $t \in T_\sigma$, which are defined by fixing $\sigma > 0$, $q \in (0,1)$, and setting

$$T_\sigma = \{t \colon gN^\sigma \le t \le N^{-\sigma} G^2\}, \quad \mathcal{D} = \{z = E + i\eta \colon E \in (-qG, qG), N^{\delta_1} \le N\eta \le 10N\} \tag{4.7}$$

The following theorem is the main result of this section.

Theorem 4.5. Let H_t be defined as above for some (g, G)-regular V. Fix $q \in (0,1)$ and $\sigma > 0$. Uniformly for $t \in T_\sigma$ and $z \in \mathcal{D}$, we have

$$|m_N(z) - m_{\mathrm{fc},t}(z)| \prec \frac{1}{N\eta}. \tag{4.8}$$

4.1.4 Reduction to diagonal V

In the proof below, we will assume that V is diagonal with entries $\{v_i\}_{i=1}^N$. If V is a general deterministic matrix, it has a singular value decomposition $V = ADB^\dagger$ with D diagonal and A, B orthogonal. Since the Gaussian ensemble W defined above is invariant under multiplication on the left and right by unitary matrices, we have

$$V + \sqrt{t}W = A(D + \sqrt{t}W')B^\dagger \tag{4.9}$$

for a Gaussian ensemble W' with the same entry distribution as W. Hence $V + \sqrt{t}W$ and $D + \sqrt{t}W'$ have the same distribution of eigenvalues, and we have shown that without loss of generality we may consider diagonal initial data.

4.1.5 Green Functions

We define the Green function matrix

$$G(z) = (M_t - zI)^{-1}. \tag{4.10}$$

For concreteness, we compute $G_{1,1}$. The other $G_{i,i}$ are analogous. We often write G_{ij} for $G_{i,j}$.

In the study of Hermitian random matrices, it is common to compute the diagonal Green function elements G_{ii} using the Schur complement formula. The off-diagonal entries G_{ij} where $i \neq j$ are asymptotically small and can be neglected. The Schur formula then produces an approximate fixed point equation for the diagonal elements G_{ii}. By analyzing the fixed point equations, one may bound the G_{ii} and obtain useful information about the eigenvalues and eigenvectors.

In the non-Hermitian setting, we would like to take the same approach. However, now the asymptotically non-trivial entries are not only the diagonal entries like G_{ii}, but also off-diagonal entries like $G_{i,i+N}$ and $G_{i+N,i}$, and the equations for the elements $(G_{ii}, G_{i,N+i}, G_{N+i,i}, G_{N+i,N+i})$ are closely coupled together. It turns out that considering these 2×2 blocks instead of individual entries yields a similar formulation to the Hermitian case, though one that requires more careful bookkeeping. This motivates the decision to work with 2×2 blocks in the following computation.

Applying the Schur complement formula with index set $\mathbb{T} = (1, N+1)$ yields

$$\begin{bmatrix} G_{1,1} & G_{1,N+1} \\ G_{N+1,1} & G_{N+1,N+1} \end{bmatrix} = (B_t - tv^\dagger G^{(\mathbb{T})} v)^{-1}, \tag{4.11}$$

where

$$B_t = \begin{bmatrix} -z & v_1 + \sqrt{t} w_{11} \\ v_1 + \sqrt{t} w_{11} & -z \end{bmatrix}, \quad v = \begin{bmatrix} 0 & w_{2,1} \\ 0 & w_{3,1} \\ \vdots & \vdots \\ w_{1,2} & 0 \\ w_{1,3} & 0 \\ \vdots & \vdots \end{bmatrix}. \tag{4.12}$$

We compute

$$v^\dagger G^{(\mathbb{T})} v = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$A_{11} = \sum_{i,j=1}^{N-1} w_{1,j+1} G^{(\mathbb{T})}_{N-1+j,n-1+i} w_{1,i+1}, \quad A_{12} = \sum_{i,j=1}^{N-1} w_{1,j+1} G^{(\mathbb{T})}_{N-1+j,i} w_{i+1,1}, \tag{4.13}$$

$$A_{21} = \sum_{i,j=1}^{N-1} w_{j+1,1} G^{(\mathbb{T})}_{j,n-1+i} w_{1,i+1}, \quad A_{22} = \sum_{i,j=1}^{N-1} w_{j+1,1} G^{(\mathbb{T})}_{j,i} w_{i+1,1}. \tag{4.14}$$

Notice that $A_{12} = A_{21}$, as $G^{(\mathbb{T})}$ is symmetric. Further, we may apply the Schur complement formula to the $(N-1) \times (N-1)$ blocks of $G^{(\mathbb{T})}$ and take the trace to see

$$\sum_{i=1}^{N-1} G^{(\mathbb{T})}_{N-1+i,N-1+i} = \sum_{i=1}^{N-1} G^{(\mathbb{T})}_{i,i}, \tag{4.15}$$

so $\mathbb{E}A_{11} = \mathbb{E}A_{22} = m^{(\mathbb{T})}(z)$.

From the formula for the inverse of a 2×2 matrix, we find

$$G_{11} = \frac{-z - tA_{22}}{(-!z - tA_{11})(-z - tA_{22}) - (v_1 + \sqrt{t} w_{11} - tA_{12})(v_1 + \sqrt{t} w_{11} - tA_{21})}, \tag{4.16}$$

$$= \frac{-z - tA_{22}}{(-z - (t/2)(A_{11} + A_{22}))^2 - ((t/2)(A_{11} - A_{22}))^2 - (v_1 + \sqrt{t} w_{11} - tA_{12})(v_1 + \sqrt{t} w_{11} - tA_{12})}. \tag{4.17}$$

Set

$$E_1 = t^2((A_{11} - A_{22})/2)^2, \quad E_2 = (A_{11} + A_{22})/2 - m^{(\mathbb{T})}. \quad r = m - m^{(\mathbb{T})}. \tag{4.18}$$

Writing m for $m_N(z)$, we have

$$= \frac{-z - tA_{22}}{(-z - tm^{(\mathbb{T})} + v_1 + \sqrt{t} w_{11} - tA_{12} - tE_2)(-z - tm^{(\mathbb{T})} - v_1 - \sqrt{t} w_{11} + tA_{12} - tE_2) - E_1}, \tag{4.19}$$

$$= \frac{-z - tA_{22}}{(-z - tm + v_1 + \sqrt{t} w_{11} - tA_{12} - tE_2 + tr)(-z - tm - v_1 - \sqrt{t} w_{11} + tA_{12} - tE_2 + tr) - E_1}. \tag{4.20}$$

Set, for $i > 0$,

$$g_i = \frac{1}{-z - tm + v_i}, \quad g_{-i} = \frac{1}{-z - tm - v_i}. \tag{4.21}$$

We define

$$E_3 = g_1(\sqrt{t}w_{11} - tA_{12} - tE_2 + tr) \tag{4.22}$$

and E_4 similarly. Then

$$G_{11} = \frac{-z - tA_{22}}{(-z - tm + v_1)(1 + E_3)(-z - tm - v_1)(1 + E_4) - E_1}, \tag{4.23}$$

$$= g_1 g_{-1} \frac{-z - tA_{22}}{[(1 + E_3)(1 + E_4) - g_1 g_{-1} E_1]}. \tag{4.24}$$

Define

$$K_1 = E_3 + E_4 + E_3E_4 - g_1g_{-1}E_1, \quad B_1 = \frac{-t([A_{22} - m^{(\mathbb{T})}] - r)}{-z - tm}. \tag{4.25}$$

Our expression for G_{ii} is, with K_i and B_i defined similarly,

$$G_{ii} = \frac{1}{2}(g_i + g_{-i}) \frac{1 + B_i}{1 + K_i}. \tag{4.26}$$

4.2 Weak law

We prove some bounds necessary for our bootstrap argument.

Lemma 4.6. For any $z \in \mathcal{D}$, $|m_{\mathrm{fc},t}(z) - m(z)| \le (\eta N)^{-1/2}$ implies

$$\frac{1}{N} \sum_i |g_i| + |g_{-i}| \le C \log N. \tag{4.27}$$

Proof. This follows as in the proof of Lemma 7.5 in [28], with the minor change that we are using m instead of $m_{\mathrm{fc},t}$ in the definition of g_i. However, if $|m_{\mathrm{fc},t} - m| \le (\eta N)^{-1/2}$, then by Lemma 4.1 we also have a lower bound $\operatorname{Im} m \ge c$, and this suffices to complete the proof. □

Lemma 4.7. Fix $z \in \mathcal{D}$. Let ϕ be the indicator function of some event, which may depend on z. If $\phi|m(z) - m_{\mathrm{fc},t}(z)| \prec N^{-c}$ for some $c > 0$, then

$$\phi \max_i |K_i| \prec \frac{1}{\sqrt{N\eta}}, \quad \phi \max_i |B_i| \prec \frac{1}{\sqrt{N\eta}}, \tag{4.28}$$

$$\frac{\phi t^2 (m_{\mathrm{fc},t} - m)}{2N} \sum_{-N}^{N} \frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)^2(-z - tm + v_i)} = O_{\prec}(N^{-c}). \tag{4.29}$$

Proof. We first consider B_1. Note that by the assumption we have a stability bound for m similar to Lemma 4.1 and this shows

$$\frac{t}{-z - tm} = O(1), \tag{4.30}$$

so it suffices to bound r and $A_{22} - m^{(\mathbb{T})}$. We have

$$|r| \le \frac{C}{N\eta}, \quad \left| A_{22} - m^{(\mathbb{T})} \right| \prec \frac{1}{\sqrt{N\eta}} \tag{4.31}$$

The first inequality follows from the eigenvalue interlacing lemma. (See, for example, Lemma 7.5 in [20].) For the second, we apply Lemma 4.4 and then Ward's identity to obtain

$$\left| A_{22} - m^{(\mathbb{T})} \right| \prec \frac{1}{N} \sqrt{\sum_{i,j} |G_{ij}^{(\mathbb{T})}|^2} = \sqrt{\frac{\operatorname{Im} m^{(\mathbb{T})}}{N\eta}} = \sqrt{\frac{\operatorname{Im} m_{\mathrm{fc},t} + r + (m - m_{\mathrm{fc},t})}{N\eta}} \prec \frac{1}{\sqrt{N\eta}}. \tag{4.32}$$

The final bound follows from the assumption on $m - m_{\mathrm{fc},t}$ and the upper bound on $\operatorname{Im} m_{\mathrm{fc},t}$ in Lemma 4.1. Finally, note these bounds are independent of i.

We now consider K_1. By the stability bound (4.2),

$$|g_1 g_{-1} t^2| \le c, \tag{4.33}$$

and by the large deviations bound Lemma 4.4,

$$\frac{A_{11} - A_{22}}{2} = O_\prec\left(\frac{1}{\sqrt{N\eta}}\right). \tag{4.34}$$

It remains to bound E_3 and E_4. We just do E_3, as E_4 is similar. Using (4.2), and noting that $\min(ct, \eta) \le |v_1 - z - tm_{\mathrm{fc},t}|$, which implies $\sqrt{c\eta t} \le |v_1 - z - tm_{\mathrm{fc},t}|$,

$$|E_3| \le \frac{|\sqrt{t} w_{11} - tA_{12} - tE_2 + tr|}{|v_1 - z - tm_{\mathrm{fc},t}|} \le \frac{|w_{11}|}{\sqrt{c\eta}} + |A_{12}| + |E_2| + |r|. \tag{4.35}$$

The r term was already bounded, and similar large deviations arguments apply to A_{12} and E_2. Because w_{11} has subexponential decay,

$$w_{11} \prec \frac{1}{\sqrt{N}}. \tag{4.36}$$

Again, these bounds are independent of i.

For the final claim, we apply the Cauchy-Schwarz inequality and note

$$\left|\frac{1}{2N}\sum_{-N}^{N} \frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)^2(-z - tm + v_i)}\right| \le \left(\frac{1}{2N}\sum_{-N}^{N}\frac{1}{|-z - tm_{\mathrm{fc},t} + v_i|^4}\right)^{1/2}\left(\frac{1}{2N}\sum_{-N}^{N}\frac{1}{|-z - tm + v_i|^2}\right)^{1/2} \tag{4.37}$$

We first bound the first factor in the right side of (4.37). By the stability bound Lemma 4.1, we get the bound

$$\frac{1}{2N}\sum_{-N}^{N}\frac{1}{|-z - tm_{\mathrm{fc},t} + v_i|^4} \le \frac{C}{2Nt^2}\sum_{-N}^{N}\frac{1}{|-z - tm_{\mathrm{fc},t} + v_i|^2} \tag{4.38}$$

Taking imaginary parts in the equation that defines $m_{\mathrm{fc},t}$ gives

$$\operatorname{Im} m_{\mathrm{fc},t} = \frac{1}{2N}\sum_{1}^{2N}\frac{t \operatorname{Im} m_{\mathrm{fc},t} + \eta}{|-z - tm_{\mathrm{fc},t} + v_i|^2}, \tag{4.39}$$

so that

$$\frac{1}{2N}\sum_{1}^{2N}\frac{1}{|-z - tm_{\mathrm{fc},t} + v_i|^2} = \frac{\operatorname{Im} m_{\mathrm{fc},t}}{t \operatorname{Im} m_{\mathrm{fc},t} + \eta} \le \frac{1}{t}. \tag{4.40}$$

We obtain

$$\left(\frac{1}{2N}\sum_{-N}^{N}\frac{1}{|-z - tm_{\mathrm{fc},t} + v_i|^4}\right)^{1/2} \le \frac{C}{t^{3/2}}. \tag{4.41}$$

We now consider the second factor in (4.37). We may sum the representation (4.26) for G_{ii} to obtain

$$m = \frac{1}{2N}\sum_{-N}^{N}\frac{1}{-z - tm + v_i} + R, \tag{4.42}$$

with $R = O_{\prec}\left(\frac{1}{\sqrt{\eta N}}\right)$, where we have used the bounds for K_i and B_i to Taylor expand

$$\frac{1+B_i}{1+K_i} = 1 + R' \tag{4.43}$$

for large N, with $R' = O_{\prec}\left(\frac{1}{\sqrt{\eta N}}\right)$, and then Lemma 4.6 to control the sum of the errors. There is a factor of $\log(N)$ that is absorbed by the stochastic domination.

Taking imaginary parts in (4.42) gives

$$\frac{\operatorname{Im} m}{\eta + t \operatorname{Im} m} = \frac{1}{2N}\sum_{-N}^{N} \frac{1}{|-z - tm + v_i|^2} + \frac{\operatorname{Im} R}{\eta + t \operatorname{Im} m}. \tag{4.44}$$

Then, using Lemma 4.1 to write $2\operatorname{Im} m \geq \operatorname{Im} m - R$ for $z \in \mathcal{D}$ and absorb the error term, we have

$$\left(\frac{1}{2N}\sum_{1}^{2N} \frac{1}{|-z - tm + v_i|^2}\right)^{1/2} \leq \frac{C}{t^{1/2}}. \tag{4.45}$$

Putting these bounds on each factor together, the expression we want to control is bounded above by

$$C\phi|m - m_{\text{fc},t}|t^2 \frac{1}{t^{3/2}} \frac{1}{t^{1/2}} \prec N^{-c}. \tag{4.46}$$

□

We also prove *a priori* bounds.

Lemma 4.8. If $\eta \geq 1$, then,

$$\max_i K_i = O_{\prec}\left(\frac{1}{\sqrt{N\eta}}\right), \quad \max_i B_i = O_{\prec}\left(\frac{1}{\sqrt{N\eta}}\right), \tag{4.47}$$

$$\frac{t^2}{2N}\sum_{-N}^{N} \frac{1}{(-z - tm_{\text{fc},t} + v_i)(-z - tm + v_i)} \leq 2t^2. \tag{4.48}$$

Proof. The first two bounds are proved as before, except we use the trivial estimate

$$\operatorname{Im} m \leq \frac{1}{\eta} \tag{4.49}$$

instead of $|m - m_{\text{fc},t}| \leq N^{-c}$ to estimate $\operatorname{Im} m^{(\mathbb{T})}$.

For the last bound, we use

$$|-z - tm + v_i| \geq \eta, \quad |-z - tm_{\text{fc},t} + v_i| \geq \eta, \quad |m| \leq \frac{1}{\eta}, \quad |m_{\text{fc},t}| \leq \frac{1}{\eta}. \tag{4.50}$$

Then

$$\frac{t^2(m - m_{\text{fc},t})}{2N}\left|\sum_{-N}^{N} \frac{1}{(z - tm + v_i)(-z - tm_{\text{fc},t} v_i)}\right| \leq \frac{2}{\eta}\frac{t^2}{2N}\frac{2N}{\eta^2} \leq 2t^2. \tag{4.51}$$

□

We now prove the weak local deformed law at the optimal scale using a bootstrapping argument. Our presentation follows [7].

Lemma 4.9. Suppose the initial values V are (g, G)-regular as in Definition 3.1. Then for $z \in \mathcal{D}$, we have $|m - m_{\text{fc},t}| \prec (N\eta)^{-1/2}$.

Proof. First, note that both m and $m_{\mathrm{fc},t}$ are N^2-Lipschitz continuous on $\mathcal{D}$. For m this is well known, and for $m_{\mathrm{fc},t}$ this is Lemma A.1 of [27]. It then suffices to prove the statement for the lattice $\hat{\mathcal{D}} = \mathcal{D} \cap (N^{-3}\mathbb{Z}^2)$. We will verify at the end of the proof that for $z \in \hat{\mathcal{D}}$ with $\eta \geq 1$ the claim follows from Lemma 4.8. We proceed assuming that the claim is true for such z.

For E such that $z_0 = E + i \in \hat{\mathcal{D}}$, define $\eta_k = 1 - kN^{-3}$ and $z_k = E + i\eta_k$. Fix $\sigma_1 < \delta_1/100$ and $D > 0$. Define

$$\Omega_k = \left\{ |m(z_k) - m_{\mathrm{fc},t}(z_k)| \leq \frac{N^{\sigma_1}}{\sqrt{N\eta}} \right\}. \tag{4.52}$$

Now recall the definition of $m_{\mathrm{fc},t}$ and the self-consistent equation (4.42) for m derived in the proof of Lemma 4.7. Subtracting these yields

$$m - m_{\mathrm{fc},t} = \frac{1}{2N} \sum_{i=-N}^{N} \frac{1}{-z - tm + v_i} - \frac{1}{-z - tm_{\mathrm{fc},t} + v_i} + R \tag{4.53}$$

$$= \frac{1}{2N} \sum_{i=-N}^{N} \frac{t(m - m_{\mathrm{fc},t})}{(-z - tm + v_i)(-z - tm_{\mathrm{fc},t} + v_i)} + R. \tag{4.54}$$

We obtain

$$(m - m_{\mathrm{fc},t}) \left(1 - \frac{t}{2N} \sum_{i=-N}^{N} \frac{1}{(-z - tm + v_i)(-z - tm_{\mathrm{fc},t} + v_i)} \right) = R. \tag{4.55}$$

By Lemma 4.8, $\mathbb{P}(\Omega_0^c) \leq N^{-D}$, as the second factor in (4.55) is bounded below and $|R| \prec (N\eta)^{-1/2}$. We now consider Ω_1. Because $m - m_{\mathrm{fc},t}$ is $2N^2$-Lipschitz on $\mathcal{D}$, we have

$$\mathbb{1}(\Omega_0)|m(z_1) - m_{\mathrm{fc},t}(z_1)| \leq \frac{N^{\sigma_1}}{\sqrt{N\eta}} + \frac{2}{N}. \tag{4.56}$$

Hence the hypothesis of Lemma 4.7 is verified for some $c > 0$ when $\phi = \mathbb{1}(\Omega_0)$ and $z = z_1$. Hence for $z = z_1$, in (4.55) we have $|R| \prec (N\eta)^{-1/2}$ and that the second factor on the left side is bounded below for large enough N. To see this, write

$$\begin{aligned} 1 - \frac{t}{2N} \sum_{i=-N}^{N} \frac{1}{(-z - tm + v_i)(-z - tm_{\mathrm{fc},t} + v_i)} &= \left(1 - \frac{t}{2N} \sum_{i=-N}^{N} \frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)^2} \right) \\ &+ \left(\frac{t}{2N} \sum_{i=-N}^{N} \frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)^2} - \frac{1}{(-z - tm + v_i)(-z - tm_{\mathrm{fc},t} + v_i)} \right). \end{aligned} \tag{4.57}$$

The first term is bounded below by Lemma 4.1, and the error term is $o(N^{-c})$ by Lemma 4.7, because it equals

$$\frac{t}{2N} \sum \frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)} \left(\frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)} - \frac{1}{(-z - tm + v_i)} \right) \tag{4.58}$$

$$= \frac{t}{2N} \sum \frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)} \left(\frac{-t(m - m_{\mathrm{fc},t})}{(-z - tm_{\mathrm{fc},t} + v_i)(-z - tm + v_i)} \right) \tag{4.59}$$

$$= \frac{(m_{\mathrm{fc},t} - m)t^2}{2N} \sum \frac{1}{(-z - tm_{\mathrm{fc},t} + v_i)^2(-z - tm + v_i)}. \tag{4.60}$$

We conclude

$$\mathbb{P}(\Omega_0 \cap \Omega_1^c) \leq N^{-D}. \tag{4.61}$$

Now we may apply this reasoning sequentially for all k such that $z_k \in \mathcal{D}$. Note that the $c > 0$ used to verify the hypothesis of Lemma 4.7 can be chosen to be the same for each step, so this lemma needs to be invoked only once. The conclusion follows by noting that $\mathbb{P}(\cap_k \Omega_k)$ can be made larger than $1 - N^{D_1}$ for any D_1 by taking D large enough.

For $z \in \hat{\mathcal{D}}$ with $\eta \geq 1$, we can use the same argument with the bounds in Lemma 4.8, which hold unconditionally, so there is no need for a bootstrapping argument. In particular, we do not need to use Lemma 4.6, since we have the trivial bound $|g_i| \leq \eta^{-1} \leq 1$. □

4.3 Strong law

We now improve the bound Lemma 4.9 using fluctuation averaging. For any random variable X, let $Q_i X$ denote the conditional expectation of X with respect to the ith column and ith row of M_t. For $I = (i, i+N)$, $J = (j, j+N)$, and any index set $\mathbb{T}$ containing pairs $(k, k+N)$, set

$$G_{IJ} = \begin{bmatrix} G_{i,j} & G_{i,N+j} \\ G_{i+n,j} & G_{N+i,N+j} \end{bmatrix}, \quad M_l^{(\mathbb{T})} = \begin{bmatrix} -z - tm^{(\mathbb{T})}(z) & v_l \\ v_l & -z - tm^{(\mathbb{T})}(z) \end{bmatrix}. \tag{4.62}$$

Lemma 4.10. For a matrix A with $A = B + R$,

$$A^{-1} = B^{-1} - B^{-1}RB^{-1} + B^{-1}RB^{-1}RA^{-1}. \tag{4.63}$$

Proof. Iterate $A^{-1} = B^{-1} - B^{-1}RA^{-1}$. □

With $i \leq N$ and $\mathbb{T} = (i, N+i)$, Schur's complement formula gives

$$G_{II}^{-1} = M_i^{(I)} + Q_i\left(G_{II}^{-1}\right), \tag{4.64}$$

and then Lemma 4.10 gives

$$G_{II} = \left[M_i^{(I)}\right]^{-1} - \left[M_i^{(I)}\right]^{-1} Q_1\left(G_{II}^{-1}\right)\left[M_i^{(I)}\right]^{-1} + \left[M_i^{(I)}\right]^{-1} Q_1\left(G_{II}^{-1}\right)\left[M_i^{(I)}\right]^{-1} Q_i\left(G_{II}^{-1}\right) G_{II}. \tag{4.65}$$

Define the deterministic quantities

$$S_i^{-1} = \begin{bmatrix} -z - tm_{\mathrm{fc},t}(z) & v_i \\ v_i & -z - tm_{\mathrm{fc},t}(z) \end{bmatrix}, \quad f_i = \frac{1}{-z - tm_{\mathrm{fc},t} + v_i}, \quad f_{-i} = \frac{1}{-z - tm_{\mathrm{fc},t} - v_i}. \tag{4.66}$$

Lemma 4.11. We have

$$\left\| S_i - \left[M_i^{(I)}\right]^{-1} \right\| \prec \frac{1}{\sqrt{N\eta}}, \tag{4.67}$$

and

$$\|S_i\| \leq \frac{|f_i| + |f_{-i}|}{2}. \tag{4.68}$$

Proof. Note that

$$S_i = \frac{1}{2}\begin{bmatrix} f_i + f_{-i} & f_i - f_{-i} \\ f_i - f_{-i} & f_i + f_{-i} \end{bmatrix}, \quad S_i^{-1} = M_i^{(I)} + \begin{bmatrix} t\varepsilon_1 & 0 \\ 0 & t\varepsilon_2 \end{bmatrix}, \tag{4.69}$$

with $|\varepsilon_k| \prec (N\eta)^{-1/2}$ by the weak law Lemma 4.9 and eigenvalue interlacing. We conclude using the same algebraic manipulations as in Section 4.1.5. □

Lemma 4.12. For $z \in \mathcal{D}$ we have

$$\frac{1}{N}\sum_i |f_i| + |f_{-i}| \leq C \log N. \tag{4.70}$$

Proof. This is Lemma 7.5 in [28]. □

Lemma 4.13. For $I = (i, i+N)$, $J = (j, j+N)$,

$$(G_{II})^{-1} = \left(G_{II}^{(J)}\right)^{-1} - (G_{II})^{-1}G_{IJ}(G_{JJ})^{-1}G_{JI}\left(G_{II}^{(J)}\right)^{-1}. \tag{4.71}$$

Proof. This is a consequence of Schur's formula. □

Lemma 4.14. The following claims hold for $z \in \mathcal{D}$ and $\mathbb{T}, \mathbb{S}$ with $|\mathbb{T}|, |\mathbb{S}| \leq \log N$, and such that $\mathbb{T}$ and $\mathbb{S}$ are composed of pairs $(k, k+N)$.

(i) We have

$$\left[M_i^{(\mathbb{T})}\right]^{-1} = \frac{1}{2}\begin{bmatrix} g_i + g_{-i} & g_i - g_{-i} \\ g_i - g_{-i} & g_i + g_{-i}\end{bmatrix} + R, \tag{4.72}$$

where E is a matrix such that $\|R\| \prec (|g_i| + |g_{-i}|)(N\eta)^{-1}$ and this bound is uniform in the index i. Hence

$$\left\|\left[M_i^{\mathbb{T}}\right]^{-1}\right\| \prec (|g_i| + |g_{-i}|) \leq \frac{C}{t}. \tag{4.73}$$

(ii) We have

$$\left\|\left[G_{II}^{(\mathbb{T})}\right]^{-1} - S_i^{-1}\right\| \prec \frac{t}{\sqrt{N\eta}} \tag{4.74}$$

and therefore

$$\left\|\left[G_{II}^{(\mathbb{T})}\right]^{-1} S_i\right\| \prec C. \tag{4.75}$$

(iii) We have

$$\left\|Q_i\left(\left[G_{II}^{(\mathbb{T})}\right]^{-1}\right)\right\| \prec \frac{t}{\sqrt{N\eta}}, \quad \left\|G_{II}^{(\mathbb{T})}\right\| \prec \frac{C}{t}. \tag{4.76}$$

(iv) For $I \neq J$, and $\mathbb{T}$ not containing I or J,

$$\left\|\left[G_{II}^{(\mathbb{T})}\right]^{-1} G_{IJ}^{(\mathbb{T})}\right\| \prec \sqrt{\frac{t}{N\eta}}, \tag{4.77}$$

$$\left\|G_{IJ}^{(\mathbb{T})}\right\| \prec \frac{\min(|g_i| + |g_{-i}|, |g_j| + |\overline{g}_j|)}{\sqrt{N\eta}}, \tag{4.78}$$

and therefore

$$\left\|G_{IJ}^{(\mathbb{T})}\left[G_{JJ}^{(\mathbb{T})}\right]^{-1} G_{JI}^{(\mathbb{T})}\right\| \prec \frac{\sqrt{t}}{N\eta}(|g_i| + |g_{-i}|). \tag{4.79}$$

(v) We have

$$\left\|[G_{II}^{(\mathbb{T})}]^{-1}[G_{II}^{(\mathbb{S})}]\right\| \prec C, \tag{4.80}$$

and hence (4.77) holds for any superscripts $\mathbb{T}$ and $\mathbb{S}$, and hence (4.79) holds for any combination of superscripts on the Green function elements.

Proof.

(i) We have

$$M_l^{(\mathbb{T})} = \begin{bmatrix} -z - tm(z) + tr & v_l \\ v_l & -z - tm(z) + tr\end{bmatrix}. \tag{4.81}$$

Here we have used the Cauchy interlacing theorem at most $\log N$ times to split off the error $|r| \prec (N\eta)^{-1}$, and the $\log N$ factor can be absorbed by the stochastic domination. The first claim now follows from the same algebraic manipulations as in our discussion of the Green functions. The second claim follows from the stability bound Lemma 4.1 and the weak law Lemma 4.9.

(ii) The first claim follows from using the representation of G_{II}^{-1} developed in our discussion of the Green function, only now applied in the same way to $\left[G_{II}^{(\mathbb{T})}\right]^{-1}$. The errors are bounded by Cauchy interlacing to control the removed rows and columns and the large deviations inequalities Lemma 4.4 are used to control the fluctuations. The second claim follows from the first and the analogue of the previous part for S_i.

(iii) The first claim is just a special case of (ii). The second follows from the explicit computation of the entries of G_{II} in Section 4.1.5 and the stability bound $|g_i| \leq Ct^{-1}$.

(iv) We first establish a Green function identity. We the Schur complement formula on the index set $I = (i, i+N)$ as in Section 4.1.5, except now we concentrate on the upper off-diagonal block. Write G_{II^c} for the sub-matrix of G whose rows are taken from the indices in I and columns from the indices in I^c. We obtain

$$G_{II^c} = -G_{II} v^\dagger G^{(I)}, \tag{4.82}$$

which implies

$$G_{II}^{-1} G_{IJ} = \sum_K^{(I)} H_{IK} G_{KJ}, \quad H_{IK} = \begin{bmatrix} 0 & \sqrt{t} w_{ik} \\ \sqrt{t} w_{ki} & 0 \end{bmatrix}. \tag{4.83}$$

We bound just the first entry of $G_{II} G_{IJ}$, as the rest are similar. This entry is

$$(G_{II}^{-1} G_{IJ})_{11} = \sum_k \sqrt{t} w_{ik} G_{k+n,j}^{(I)}. \tag{4.84}$$

By the large deviations estimates Lemma 4.4 and Ward's identity, we have

$$\left| \sum_k \sqrt{t} w_{ik} G_{k+n,j}^{(I)} \right| \prec \sqrt{\frac{t}{N}} \sqrt{\sum_k |G_{k+n,j}^{(I)}|^2} \leq \sqrt{\frac{t}{N}} \sqrt{\frac{\operatorname{Im} G_{kk}^{(I)}}{\eta}} \prec \sqrt{\frac{t}{N\eta}}. \tag{4.85}$$

In the last step, we used the bound $\operatorname{Im} G_{kk}^{(I)} \leq \operatorname{Im} m^{(I)} \leq \log(N)$ and absorbed the $\log N$ factor into the stochastic domination. For the second claim, we have a similar identity, obtained in the same way using the other off-diagonal block in the Schur complement formula.

$$G_{IJ} = \left(\sum_K^{(J)} G_{IK}^{(J)} H_{KJ} \right) G_{JJ} \tag{4.86}$$

We can expand this using the first identity to obtain

$$G_{IJ} = G_{II} \left(\sum_{M,K}^{(I,J)} H_{IM} G_{MK}^{(IJ)} H_{KJ} - H_{IJ} \right) G_{JJ}^{(I)} \tag{4.87}$$

As shown in the work in Section 4.1.5, up to an $O(1)$ factor, we have

$$\|G_{II}\| \leq |f_i| + |f_{-i}|, \quad \|G_{JJ}\| \leq |f_j| + |f_{-j}|. \tag{4.88}$$

The middle factor has a norm that is stochastically dominated by

$$t\sqrt{\frac{\log N}{N\eta}} + \frac{\sqrt{t}}{\sqrt{N}}, \tag{4.89}$$

where we used the large deviations bounds Lemma 4.4 and the Ward identity as before. The lemma follows from using $|g_i|, |g_{-i}| \le \min(t^{-1}, \eta^{-1})$. The third claim follows from the first two. Clearly we could repeat this argument for any $\mathbb{T}$ satisfying the given hypotheses.

(v) First, we show this implies the modification of (4.77). We want to bound $[G_{II}^{(\mathbb{T})}]^{-1} G_{IJ}^{(\mathbb{S})}$. Write this as

$$[G_{II}^{(\mathbb{T})}]^{-1}[G_{II}^{(\mathbb{S})}][G_{II}^{(\mathbb{S})}]^{-1}G_{IJ}^{(\mathbb{S})}. \tag{4.90}$$

The final two terms are bounded in norm by the previous part, so we need to bound

$$\left\| [G_{II}^{(\mathbb{T})}]^{-1}[G_{II}^{(\mathbb{S})}] \right\| \le C \tag{4.91}$$

as claimed. This follows from the work in Section 4.1.5 by estimating

$$\left\| [G_{II}^{(\mathbb{S})}]^{-1} - [G_{II}^{(\mathbb{T})}]^{-1} \right\| \prec \sqrt{\frac{1}{N\eta}} \tag{4.92}$$

using the explicit representation there.

□

We record an elementary fact for later use.

Lemma 4.15. For any symmetric matrix M and integer $r > 0$,

$$\|M\|^{2r} \le \operatorname{Tr} M^{2r}. \tag{4.93}$$

The following arguments are based on the proof of Lemma 7.15 in [28].

Lemma 4.16. For $z \in \mathcal{D}$ and p even,

$$\mathbb{E} \left\| \frac{1}{N} \sum S_i Q_i \left(G_{II}^{-1} \right) S_i \right\|^p \prec \frac{1}{(N\eta)^p}. \tag{4.94}$$

Proof. We proceed as in the proof of Lemma 4.7 in Appendix B of [18], by computing moments. We first consider the case $p = 2$. Recall that if Y is a random variable independent of the ith row and ith column, then $\mathbb{E}Q_i(X)Y = \mathbb{E}Q_i(XY) = 0$. We have

$$\mathbb{E} \left\| \frac{1}{N} \sum_{i=1}^N S_i Q_i \left(G_{II}^{-1} \right) S_i \right\|^2 \le \operatorname{Tr} \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left[S_i Q_i \left(G_{II}^{-1} \right) S_i \right]^2 \tag{4.95}$$

$$+ \operatorname{Tr} \frac{1}{N^2} \sum_{i \neq j}^N \mathbb{E} \left[S_i Q_i \left(G_{II}^{-1} \right) S_i S_j Q_j \left(G_{JJ}^{-1} \right) S_j \right] = A_1 + A_2. \tag{4.96}$$

By Lemma 4.14, and the fact that, up to a constant, the norm of a matrix bounds its trace, we have

$$|A_1| \prec C \frac{1}{N\eta} \frac{1}{N^2} \sum_{i=1}^N (|f_i| + |f_{-i}|)^2 \le \left(\frac{1}{N\eta} \right)^2. \tag{4.97}$$

Here we used

$$\frac{1}{N} \sum_{i=1}^N |f_i|^2 + |f_{-i}|^2 \le \frac{1}{\eta}, \tag{4.98}$$

which comes from $|f_i| \le \eta^{-1}$.

For A_2, we note that S_i is deterministic, so we may move it inside the Q_i.

$$A_2 = \frac{1}{N^2} \sum_{i \neq j}^N \mathbb{E} Q_i \left(S_i G_{II}^{-1} S_i \right) Q_j \left(S_j G_{JJ}^{-1} S_j \right). \tag{4.99}$$

By Lemma 4.13, we can write

$$Q_i\left(S_i\left[\left(G_{II}^{(J)}\right)^{-1}-(G_{II})^{-1}G_{IJ}(G_{JJ})^{-1}G_{JI}\left(G_{II}^{(J)}\right)^{-1}\right]S_i\right) \tag{4.100}$$

Recall that if Y is a random variable independent of the jth row and jth column, then $\mathbb{E}Q_j(X)Y=\mathbb{E}Q_j(XY)=0$. So the first term will cancel when multiplied against the $Q_j(\cdot)$ term. We are left with multiplying two terms of the following form.

$$Q_i\left(S_i(G_{II})^{-1}G_{IJ}(G_{JJ})^{-1}G_{JI}\left(G_{II}^{(J)}\right)^{-1}S_i\right) \tag{4.101}$$

Using Lemma 4.14 again, we see that this is stochastically dominated by $(N\eta)^{-1}(|g_i|+|g_{-i}|)$. Specifically, we use (4.79) on the middle three terms and (4.75) on the outside pairs. Then

$$A_2\le\frac{1}{(N\eta)^2}\frac{1}{N^2}\sum_{i\neq j}(|f_i|+|f_{-i}|)(|f_j|+|f_{-j}|) \tag{4.102}$$

Now we sum over i and j separately, picking up $\log$ terms by Lemma 4.12, and obtain the bound $A_2\prec(N\eta)^{-2}$.

We now discuss even $p>1$. Our strategy, as in [28], is to adapt the proof of Theorem 4.7 in [18] to the deformed case. Applying Lemma 4.15 to $M=\left\|\sum S_iQ_i\left(G_{II}^{-1}\right)S_i\right\|^p$ yields, using the notation in [28],

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}S_iQ_i\left(G_{II}^{-1}\right)S_i\right\|^p\le\frac{1}{N^p}\mathbb{E}\left[\operatorname{Tr}\sum_{k_1,\cdots,k_{2p}}\prod S_{k_i}Q_{k_i}\left(G_{K_iK_i}^{-1}\right)S_{k_i}\right] \tag{4.103}$$

$$=\operatorname{Tr}\frac{1}{N^p}\sum_{\Gamma\in\mathcal{P}_p}\sum_{i_1,\dots,i_r}\mathbb{1}_{\{\Gamma=\Gamma(i)\}}\mathbb{E}\left[S_{i_1}Q_{i_1}\left(G_{I_1I_1}^{-1}\right)S_{i_1}\dots S_{i_p}Q_{i_p}\left(G_{I_pI_p}^{-1}\right)S_{i_p}\right]. \tag{4.104}$$

We now apply the algorithm in [18] to decompose each term, using the analogous substitutions (which also hold with any superscript $\mathbb{T}$ added to the Green function matrices)

$$G_{IJ}\to G_{IJ}^{(K)}+G_{IK}G_{KK}^{-1}G_{KJ}, \tag{4.105}$$

$$(G_{II})^{-1}\to\left(G_{II}^{(J)}\right)^{-1}-(G_{II})^{-1}G_{IJ}(G_{JJ})^{-1}G_{JI}\left(G_{II}^{(J)}\right)^{-1}. \tag{4.106}$$

We see that these substitutions create non-commutative polynomials in the Green functions with the properties that inverted Green function matrices alternate with non-inverted ones, and that the lower indices pair across adjacent terms. The algorithm yields binary strings σ_k and an expansion into monomials.

$$\mathbb{E}\left[S_{i_1}Q_{i_1}\left(G_{I_1I_1}^{-1}\right)S_{i_1}\dots S_{i_p}Q_{i_p}\left(G_{I_pI_p}^{-1}\right)S_{i_p}\right]$$
$$=\sum_{\sigma_1,\dots,\sigma_p}\mathbb{E}[S_{i_1}Q_{i_1}\left(F_{i_1}\right)_{\sigma_1}S_{i_1}\dots S_{i_p}Q_{i_p}\left(F_{i_p}\right)_{\sigma_p}S_{i_p}] \tag{4.107}$$

Set $\Phi=(N\eta)^{-1/2}$. We now claim it suffices to establish the key bound

$$\left\|[\mathbb{E}S_{i_1}Q_{i_1}\left(F_{i_1}\right)_{\sigma_1}S_{i_1}\dots S_{i_p}Q_{i_p}\left(F_{i_p}\right)_{\sigma_p}S_{i_p}]\right\|\prec C\Psi^{p+s}\prod_{i_k}(|f_{i_k}|+|f_{-i_k}|) \tag{4.108}$$

where s is the number of lone labels. This is proved in the Lemma 4.17, following this proof. Assuming this lemma, we will conclude the proof.

It follows from Lemma 4.17 that for a partition Γ with $l = |\Gamma|$,

$$\left\| \frac{1}{N^p} \sum_{i_1,\dots,i_p} \mathbb{1}_{\Gamma=\Gamma(i)} \mathbb{E}[\dots] \right\| \prec \frac{\Psi^{p+s}}{N^p} \sum_{k_1=1}^{N} \cdots \sum_{k_l=1}^{N} (|f_{k_1}|+|f_{-k_1}|)^{d_1} \cdots (|f_{k_l}|+|\overline{f}_{k_l}|)^{d_l}, \quad (4.109)$$

where the d_l are the sizes of the blocks in Γ. Then $|f_i| \le \eta^{-1}$ implies the bound

$$\le \Psi^{p+s} \frac{1}{N^p} \frac{1}{\eta^{p-l}} \sum_{k_1=1}^{N} \cdots \sum_{k_l=1}^{N} (|f_{k_1}| + |f_{-k_1}|) \cdots (|f_{k_l}| + |f_{-k_l}|), \quad (4.110)$$

$$\le \Psi^{p+s} \frac{1}{N^{p-l}} \frac{1}{\eta^{p-l}} \log(N)^l \prec \Psi^{2p+s} \Psi^{2p-2l}. \quad (4.111)$$

Using $p + s + (2p - 2l) \ge 2p$ we have

$$\mathbb{E} \left\| \frac{1}{N} \sum S_i Q_i \left(G_{II}^{-1}\right) S_i \right\|^p \prec \sum_{\Gamma \in \mathcal{P}_p} \Psi^{2p} \le (Cp)^{Cp} \Psi^{2p} \prec \Psi^{2p}. \quad (4.112)$$

This concludes the proof. □

Lemma 4.17.

$$\left\| \mathbb{E}[S_{i_1} Q_{i_1} (F_{i_1})_{\sigma_1} S_{i_1} \dots S_{i_p} Q_{i_p} \left(F_{i_p}\right)_{\sigma_p} S_{i_p}] \right\| \prec C \Psi^{p+s} \prod_{i_k} (|f_{i_k}| + |f_{-i_k}|) \quad (4.113)$$

Proof. We claim that if $b(\sigma_k)$ is the number of ones appearing in σ_k, then

$$\|S_{i_k} Q_{i_k} (F_{i_k})_{\sigma_k} S_{i_k}\| \prec \Psi^{b(\sigma_k)+1} (|f_{i_k}| + |f_{-i_k}|). \quad (4.114)$$

If $\sigma = 0$, then we have a single term of the form $S_{i_k} Q_{i_k} \left((G_{II}^{(T)})^{-1}\right) S_{i_k}$, and using Lemma 4.14 we have $\|S_{i_k} Q_{i_k} \left((G_{II}^{(T)})^{-1}\right) S_{i_k}\| \le \Psi(|f_{i_k}| + |f_{-i_k}|)$. If $\sigma > 1$, we can proceed similarly, except each interior pair of diagonal and off-diagonal terms gains a factor of Ψ by (4.77), and there are σ of these.

Now we have various cases.

(i) One of the monomials $(F_{i_k})_{\sigma_k}$ is not maximally expanded. Then $(F_{i_k})_{\sigma_k}$ contains $\ge 2p$ off-diagonal Green function entries, and $b(\sigma_k) \ge 2p - 1 \ge 2r + s$. The result follows from the previous discussion.

(ii) Every monomial is maximally expanded, and for every lone label a there is a label $b \in \{1, \dots, p\} \setminus \{a\}$ such that the monomial (F_{i_b}) contains an off-diagonal resolvent entry with lower index I_a. Then $\sum b(\sigma_k) \ge s$ and again we are finished.

(iii) Every monomial is maximally expanded, and there is a lone label without a matching lower index in another monomial, say i_1. Then, since the S_i are deterministic and the other terms are maximally expanded, the other $S_{i_k} Q_{i_k} (F_{i_k})_{\sigma_k} S_{i_k}$ are independent of the i_1 row and column. Then we can write

$$\mathbb{E}\left[S_{i_1} Q_{i_1} (F_{i_1})_{\sigma_1} S_{i_1} \dots S_{i_p} Q_{i_p} \left(F_{i_p}\right)_{\sigma_p} S_{i_p}\right]$$
$$= \mathbb{E}\left[Q_{i_1} \left(S_{i_1} (F_{i_1})_{\sigma_1} S_{i_1} \dots S_{i_p} Q_{i_p} \left(F_{i_p}\right)_{\sigma_p} S_{i_p}\right)\right] = 0. \quad (4.115)$$

□

Lemma 4.18. For $z \in \mathcal{D}$, we have entrywise

$$\frac{1}{N}\sum S_i Q_i\left(G_{II}^{-1}\right) S_i \prec \frac{1}{N\eta}. \tag{4.116}$$

Proof. This follows from Lemma 4.16 and Markov's inequality. □

Proof of Theorem 4.5. We consider each term in (4.65), average over N, and take trace. For the first term we get

$$g_i + g_{-i} + \text{error}, \tag{4.117}$$

where

$$|\text{error}| \le \frac{C}{N\eta}\frac{1}{2N}\left(\sum_{i=1}^{N}|g_i| + |g_{-i}|\right) \le \frac{C\log N}{N\eta}. \tag{4.118}$$

It remains to show the contributions from averaging the last two terms are negligible. The second term is dealt with by using Lemma 4.11 to make the replacement

$$\left[M_i^{(I)}\right]^{-1} Q_1\left(G_{II}^{-1}\right)\left[M_i^{(I)}\right]^{-1} \to S_i Q_1\left(G_{II}^{-1}\right) S_i \tag{4.119}$$

and then Lemma 4.18 to bound the averaged error from this replacement. The largest terms in the error have the form

$$\left\|\left(S_i - \left[M_i^{(I)}\right]^{-1}\right) Q_1\left(G_{II}^{-1}\right)\left[M_i^{(I)}\right]^{-1}\right\| \prec \frac{1}{\sqrt{N\eta}}\frac{t}{\sqrt{N\eta}}\frac{1}{t} = \frac{1}{N\eta}. \tag{4.120}$$

By Lemma 4.14, the third term is

$$\left[\frac{1}{2}\begin{bmatrix} g_i + g_{-i} & g_i - g_{-i} \\ g_i - g_{-i} & g_i + g_{-i}\end{bmatrix} + O\left(\frac{|g_i| + |g_{-i}|}{N\eta}\right)\right] Q_1\left(G_{II}^{-1}\right)\left[M_i^{(I)}\right]^{-1} Q_i\left(G_{II}^{-1}\right) G_{II} \tag{4.121}$$

Again by Lemma 4.14

$$\left\|Q_1\left(G_{II}^{-1}\right)\left[M_i^{(I)}\right]^{-1} Q_i\left(G_{II}^{-1}\right) G_{II}\right\| \le C\frac{t}{\sqrt{N\eta}}\frac{1}{t}\frac{t}{\sqrt{N\eta}}\frac{1}{t} \le \frac{C}{N\eta}, \tag{4.122}$$

and

$$\left\|\frac{1}{2}\begin{bmatrix} g_i + g_{-i} & g_i - g_{-i} \\ g_i - g_{-i} & g_i + g_{-i}\end{bmatrix} + O\left(\frac{|g_i| + |g_{-i}|}{N\eta}\right)\right\| \le C(|g_i| + |g_{-i}|). \tag{4.123}$$

Hence the average over the third term is negligible. Note the average over the $|g_i| \pm |g_{-i}|$ picks up a negligible $\log N$ factor, as above.

Finally, we have

$$m = \frac{1}{2N}\sum_{-N}^{N}\frac{1}{-z - tm + v_i} + R, \tag{4.124}$$

where $|R| \prec (N\eta)^{-1}$, and we can repeat the proof of Lemma 4.9 with the improved error $(N\eta)^{-1}$ in place of $(N\eta)^{-1/2}$ to conclude. □

5 Removal of time evolution

In this section we show how to complete the proof of universality given the main homogenization result Theorem 3.2. In Subsection 5.1 we prove a local law for sparse matrices, which is necessary in the following subsections. In Subsection 5.2 we prove short time universality for sparse random matrices. Finally, in Subsection 5.3 we show how to remove the time evolution through a Green function comparison argument.

5.1 Sparse local law

For clarity, in this subsection we consider just the case of sparse ensembles where the variances are equal, $s_{ij} = N^{-1}$, but the case of a general doubly stochastic variance matrix satisfying the conditions in Definition 2.1 can be handled with minor modifications. The key point is that in each of these cases the limiting spectral distribution is a semicircle. More general variance matrices (which give rise to new limit distributions), along with correlated entries, are considered in Section 6.

We prove a weak local law for the singular values of sparse matrices. The symmetric case was considered in [19]. We recall that our model is $M = B + f\,|w\rangle\,\langle w|$, using the notation of Section 2. The next lemma implies it is enough to prove such a law for B. Let the singular values of M be $(\mu_i)_{i=1}^N$ and the singular values of B be $(\lambda_i)_{i=1}^N$.

Lemma 5.1. The singular values of M and B are interlaced,

$$\lambda_{j+1} \geq \mu_j \geq \lambda_{j-1}. \tag{5.1}$$

Proof. Note that M is a rank 1 perturbation of B. The result follows from Weyl's inequality and Majorization for singular values. □

Lemma 5.2. Letting m_{ij} denote the entries of M, we have

$$|m_{ij}| \prec \frac{1}{q}. \tag{5.2}$$

Proof. This follows from Markov's inequality. □

We also require a slight modification of Lemma 3.8 from [19].

Lemma 5.3. Let $a_1, \ldots, a_N$ be centered and independent random variables satisfying

$$\mathbb{E}|a_i|^p \leq \frac{C^p}{Nq^{p-2}} \tag{5.3}$$

for all p. Then for any $A_i \in \mathbb{C}$ and $B_{ij} \in \mathbb{C}$,

$$\left|\sum_{i=1}^N A_i a_i\right| \prec \frac{\max_i |A_i|}{q} + \left(\frac{1}{N}\sum_{i=1}^N |A_i|^2\right)^{1/2}, \tag{5.4}$$

$$\left|\sum_{i=1}^N \overline{a_i} B_{ii} a_i - \sum_{i=1}^N \sigma_i^2 B_{ii}\right| \prec \frac{B_d}{q}, \tag{5.5}$$

$$\left|\sum_{1\leq i\neq j\leq N} \overline{a_i} B_{ij} a_j\right| \prec \frac{B_o}{q} + \left(\frac{1}{N^2}\sum_{i\neq j} |B_{ij}|^2\right)^{1/2}, \tag{5.6}$$

where σ_i^2 is the variance of a_i and

$$B_d = \max_i |B_{ii}|, \quad B_o = \max_{i\neq j} |B_{ij}|. \tag{5.7}$$

Further, if $a_1, \ldots, a_N$ and $b_1, \ldots, b_N$ are independent random variables satisfying the above moment condition, then for $B_{ij} \in \mathbb{C}$ we have

$$\left|\sum_{i,j=1}^N a_i B_{ij} b_j\right| \prec \left[\frac{B_d}{q^2} + \frac{B_o}{q}\left(\frac{1}{N^2}\sum_{i\neq j} |B_{ij}|^2\right)^{1/2}\right] \tag{5.8}$$

Let $K = \begin{bmatrix} 0 & B \\ B^\dagger & 0 \end{bmatrix}$ be the symmetric $2N \times 2N$ block matrix formed from B and $G_{ij}(z)$ be entries of the Green function of K. Let $m(z)$ be the Stieltjes transform of K. Define

$$\Lambda_o = \max_{i \neq j} |G_{ij}|, \quad \Lambda_d = \max_i |G_{ii} - m_{\text{sc}}|, \quad \Lambda = |m - m_{\text{sc}}|. \tag{5.9}$$

Repeating the Green function calculations for the deformed case, we have

$$G_{ii} = g\left(\frac{1+B_i}{1+K_i}\right), \tag{5.10}$$

with

$$g = \frac{1}{-z-m}, \quad K_i = E_3 + E_4 + E_3E_4 - g^2E_1, \quad B_i = \frac{-([A_{22} - m^{(\mathbb{T})}] - r)}{-z-m}, \tag{5.11}$$

$$E_1 = \left(\frac{A_{11} - A_{22}}{2}\right)^2, \quad E_2 = \frac{A_{11} + A_{22}}{2} - m^{(\mathbb{T})}, \quad E_1 = g(h_{ii} - A_{12} + E_2 + r), \quad r = m - m^{(\mathbb{T})}. \tag{5.12}$$

For any $\delta > 0$, set

$$\mathcal{D}_\delta = \{z = E + i\eta \colon E \in (-1,1), N^\delta \le N\eta \le 10N\}. \tag{5.13}$$

We will proceed largely as in the proof of the deformed law. The key difference is that we have better stability for m_{sc} (see Lemma 6.2 in [20]), so we can prove the local law for a larger spectral domain.

Lemma 5.4. Suppose $z \in \mathcal{D}_\delta$. Let ϕ be the indicator function of some event, which may depend on z. If $\phi(\Lambda_o + \Lambda_d) \prec N^{-c}$ for some $c > 0$, then

$$\phi \max_i |K_i| \prec \left(q^{-1} + (N\eta)^{-1/2}\right), \quad \phi \max_i |B_i| \prec \left(q^{-1} + (N\eta)^{-1/2}\right), \tag{5.14}$$

$$\left(1 - \frac{1}{(-z-m)(-z-m_{\text{sc}})}\right) = O_\prec(N^{-c}). \tag{5.15}$$

Proof. For concreteness, we consider B_1, but our bounds will be uniform in i. By the stability bound for m_{sc} and the hypothesis on m, we have $\frac{1}{-z-m} \le C$. By Cauchy's interlacing lemma, $r \le C(N\eta)^{-1}$. Using Lemma 5.3, reasoning as in the proof of the deformed law, we have

$$|A_{22} - m^{(\mathbb{T})}| \prec \frac{\Lambda_o}{q} + \frac{1}{\sqrt{N\eta}}. \tag{5.16}$$

Combining these completes the proof for the B_i.

For the K_i, it remains to bound E_3 and E_4. As these are similar to what was done before, we just sketch the proof for E_3.

$$|E_3| \le |h_{11} - A_{12} - E_2 + r| \tag{5.17}$$

The r term was already bounded. Large deviations arguments using Lemma 5.3 suffice to bound A_{12} and E_2, and h_{11} is bounded using Lemma 5.2. Combining these completes the proof.

For the final bound, we write

$$1 - \frac{1}{(-z-m)(-z-m_{\text{sc}})} = \left(1 - \frac{1}{(-z-m_{\text{sc}})^2}\right) + \left(\frac{1}{(-z-m_{\text{sc}})^2} - \frac{1}{(-z-m)(-z-m_{\text{sc}})}\right). \tag{5.18}$$

The first term equals $1 - m_{\text{sc}}^2$ and by Lemma 6.2 of [20] it is bounded above by a constant in $\mathcal{D}_\delta$. The second term is $O(N^{-c})$, which follows from the hypotheses and the bounds in the aforementioned lemma. □

Lemma 5.5. For $z \in \mathcal{D}_\delta$, Let ϕ be the indicator function of some event, which may depend on z. If $\phi(\Lambda_o + \Lambda_d) \prec N^{-c}$ for some $c > 0$, then

$$\Lambda_d \prec \Lambda + \frac{1}{q} + \frac{1}{\sqrt{N\eta}}, \quad \Lambda_o \prec \frac{1}{q} + \frac{1}{\sqrt{N\eta}}. \tag{5.19}$$

Proof. The first claim is proved the same way as in display (3.39) in [19]. We use the explicit expression (5.10) for G_{ii} above to compute

$$G_{ii} - G_{jj} \leq C\left(\frac{1}{\sqrt{N\eta}} + \frac{1}{q}\right). \tag{5.20}$$

The claim follows by fixing i and averaging over j.

The second claim is proved as in Lemma 3.13 of [19]. We use

$$G_{IJ} = G_{II}\left(\sum_{M,K}^{(I,J)} H_{IM} G_{MK}^{(J)} H_{KJ} - H_{IJ}\right) G_{JJ}^{(I)}. \tag{5.21}$$

By hypothesis we find

$$\|G_{II}\| \leq C, \quad \|G_{JJ}\| \leq C, \tag{5.22}$$

so using Lemma 5.3 on the individual entries of the matrix expression,

$$\left\| G_{II}\left(\sum_{M,K}^{(I,J)} H_{IM} G_{MK}^{(J)} H_{KJ} - H_{IJ}\right) G_{JJ} \right\| \prec C\left(\frac{1}{q} + \frac{\Lambda_o}{q} + \left(\frac{1}{N^2}\sum_{k,l}^{(IJ)} \left|G_{kl}^{(IJ)}\right|^2\right)^{1/2}\right). \tag{5.23}$$

By Ward's identity,

$$\frac{1}{N^2}\sum_{k,l}^{(IJ)} \left|G_{kl}^{(IJ)}\right|^2 = \frac{1}{N^2\eta}\sum_{k}^{(IJ)} \operatorname{Im} G_{kk}^{(ij)} \leq \frac{\operatorname{Im} m}{N\eta} + \frac{C\Lambda_o^2}{N\eta}, \tag{5.24}$$

where the last inequality follows from using

$$|G_{ij}| \leq C, \quad c \leq |G_{ii}| \leq C, \quad G_{ij} = G_{ij}^{(k)} + \frac{G_{ik}G_{kj}}{G_{kk}}. \tag{5.25}$$

repeatedly. Taking the maximum over $i \neq j$ gives

$$\Lambda_o \prec \frac{C}{q} + o(1)\Lambda_o + \sqrt{\frac{\operatorname{Im} m}{N\eta}}, \tag{5.26}$$

which implies the claim. □

Lemma 5.6. If $\eta \geq 2$, then,

$$\max_i |K_i| \prec \left(q^{-1} + (N\eta)^{-1/2}\right), \quad \max_i |B_i| \prec \left(q^{-1} + (N\eta)^{-1/2}\right), \tag{5.27}$$

$$\left(1 - \frac{1}{(-z-m)(-z-m_{\mathrm{sc}})}\right) \geq c. \tag{5.28}$$

Proof. The proof is similar to the above using the trivial estimates for $\eta \geq 2$ as in Lemma 4.8. □

Lemma 5.7. If $\eta \geq 2$,

$$\Lambda_d(z) + \Lambda_o(z) \prec \frac{1}{\sqrt{N}} + \frac{1}{q}. \tag{5.29}$$

Proof. The bound on Λ_o follows from the same calculation as in Lemma 5.5, where we now use Lemma 5.6 to bound the error terms and the trivial estimates $|G_{ij}| \leq \eta^{-1}$ to bound the Green function entries. For Λ_d we estimate using (5.10)

$$\Lambda_d = |m_{\rm sc} - G_{ii}| = \left| m_{\rm sc} - \frac{1}{-z-m}(1+R) \right| \leq \left| \frac{1}{z+m} - \frac{1}{z+m_{\rm sc}} \right| + C|R| \tag{5.30}$$

$$\leq \left| \frac{m - m_{\rm sc}}{(m+z)(m_{\rm sc}+z)} \right| + C|R| \leq \frac{\Lambda_d}{3/2} + C|R|, \tag{5.31}$$

where we used

$$|z + m_{\rm sc}| = |m_{\rm sc}|^{-1} \geq 2, \quad |m - m_{\rm sc}| \leq 1. \tag{5.32}$$

The conclusion follows by combining the Λ_d terms on the left and using Lemma 5.6 to bound R. □

Let

$$H = \begin{bmatrix} 0 & M \\ M^\dagger & 0 \end{bmatrix} \tag{5.33}$$

be the symmetric $2N \times 2N$ block matrix formed from M, and define $\tilde{m}$ to be the Stieltjes transform of H.

Lemma 5.8. Uniformly for $z \in \mathcal{D}_\delta$, we have $|\tilde{m} - m_{\rm sc}| \prec q^{-1/2} + (N\eta)^{-1/2}$.

Proof. As noted above, it is enough to establish the theorem for m, the Stieltjes transform of K. Define, following the proof of Lemma 4.9, the lattice $\hat{\mathcal{D}}_\delta = \mathcal{D}_\delta \cap (N^{-3}\mathbb{Z}^2)$. We have already shown in Lemma 5.7 that the claim holds for $z \in \hat{\mathcal{D}}_\delta$ with $\eta \geq 2$. As in the proof of the deformed weak law, it suffices to prove the result holds uniformly for elements of the lattice $\hat{\mathcal{D}}_\delta$ with $\eta < 2$. Define $n_k = 2 - kN^{-3}$ and $z_k = E + i\eta_k$. Fix $\sigma > 0$ and $D > 0$, and define

$$\Omega_k = \left\{ \Lambda_o(z_k) + \Lambda_d(z_k) \leq \frac{N^\sigma}{\sqrt{N\eta}} + \frac{1}{q} \right\}. \tag{5.34}$$

Note that $\Lambda_d \geq \Lambda$, so $\Lambda(z_k) \leq \frac{N^\sigma}{\sqrt{N\eta}} + \frac{1}{q}$ on Ω_k.

It is well known that $m_{\rm sc}$ satisfies a self consistent equation

$$m_{\rm sc} = \frac{1}{-z - m_{\rm sc}}. \tag{5.35}$$

Using Lemma 5.6, we may Taylor expand (5.10) to find

$$G_{ii} = \frac{1}{-z-m}(1+R'), \tag{5.36}$$

where $|R'| \prec q^{-1} + (N\eta)^{-1/2}$. By the stability estimate $\operatorname{Im} m_{\rm sc} \geq c$ and working on the set Ω_0 to control Λ, we have

$$m = \frac{1}{-z-m} + R \tag{5.37}$$

where $|R| \prec q^{-1} + (N\eta)^{-1/2}$. Subtracting the two self consistent equations yields

$$(m - m_{\rm sc}) \left(1 - \frac{1}{(-z-m)(-z-m_{\rm sc})} \right) = R. \tag{5.38}$$

By Lemma 5.7, $\mathbb{P}(\Omega_0^c) \le N^{-D}$. We now consider Ω_1. Because $m - G_{ii}$ is $2N^2$-Lipschitz on $\mathcal{D}$, we have

$$\mathbb{1}(\Omega_0)|m(z_1) - G_{ii}(z_1)| \le \frac{N^\sigma}{\sqrt{N\eta}} + \frac{2}{N} + \frac{1}{q}. \tag{5.39}$$

This shows that $\Lambda_d(z_1) \le N^{-c}$ for some $c > 0$, and similar reasoning applies to Λ_o. Then, by Lemma 5.4, the coefficient of $(m - m_{\text{sc}})$ in the self-consistent equation is bounded below and $|R| \prec q^{-1} + (N\eta)^{-1/2}$. Hence $\Lambda(z_1) \prec (N\eta)^{-1/2} + q^{-1}$. We conclude by Lemma 5.5 that

$$\mathbb{P}(\Omega_0 \cap \Omega_1^c) \le N^{-D}. \tag{5.40}$$

We may apply this reasoning sequentially for all k such that $z_k \in \mathcal{D}$. The conclusion follows by noting that $\mathbb{P}(\cap_k \Omega_k)$ can be made larger than $1 - N^{D_1}$ for any D_1 by taking D large enough. □

5.2 Short time universality for sparse matrices

Let M_N be a sparse matrix ensemble and define H by forming a $2N \times 2N$ symmetric block matrix as in (2.10):

$$H = \begin{bmatrix} 0 & M_N \\ M_N^\dagger & 0 \end{bmatrix}. \tag{5.41}$$

We now define the perturbed matrix H_t that we show universality for. Because we must accommodate the possibly unequal variance structure, we cannot simply add a Gaussian matrix. Instead, we evolve the nonzero entries of H according to the following Ornstein-Uhlenbeck dynamics from [10].

$$d(h_{ij}(t) - f) = \frac{dB_{ij}(t)}{\sqrt{N}} - \frac{1}{2Ns_{ij}}(h_{ij}(t) - f)\,dt \tag{5.42}$$

Here the B_{ij} are independent Brownian motions. The entries of the evolved matrix H_t satisfy

$$h_{ij}(t) = f + \exp\left(-\frac{t}{2Ns_{ij}}\right)(h_{ij}(0) - f) + \frac{1}{\sqrt{N}}\int_0^t \exp\left(-\frac{s-t}{2Ns_{ij}}\right) dB_{ij}(s). \tag{5.43}$$

We choose this dynamics because it preserves the mean and variance of the initial entries, and because the resulting entries are Gaussian divisible. With $r = \min\{Ns_{ij}\}$ and G a $2N \times 2N$ symmetrized version of a $N \times N$ ensemble of independent standard Gaussian variables,

$$H_t \overset{d}{=} H_t^{(1)} + \sqrt{\frac{r(1 - \exp(-t/r))}{N}}G, \tag{5.44}$$

where

$$\left(H_t^{(1)}\right)_{ij} \overset{d}{=} f + e^{-\frac{t}{2Ns_{ij}}}(h_{ij}(0) - f) + \sqrt{Ns_{ij}\left(1 - e^{-\frac{t}{Ns_{ij}}}\right) - r(1 - e^{-t/r})}\,\frac{\tilde{B}_{ij}(t)}{\sqrt{N}}, \tag{5.45}$$

and the $\tilde{B}_{ij}$ are a family of symmetric, independent Brownian motions. Note that

$$\sqrt{t} \asymp \sqrt{r(1 - \exp(-t/r))} \tag{5.46}$$

because r is bounded below.

Before invoking Theorem 3.2, we prove a lemma that assists in showing that sparse matrices are (g, G)-regular.

Lemma 5.9. The largest singular value of M satisfies $\mu_n \le N^C$ for some C with overwhelming probability.

Proof. The largest singular value of M is equal to the largest eigenvalue of the symmetrized sparse matrix $\begin{bmatrix} 0 & M \\ M^\dagger & 0 \end{bmatrix}$. The lemma then follows from the proof of Lemma 4.3 in [19]. □

We now obtain short time universality for the new dynamics. Given a matrix M, let $\lambda_1(t, M)$ denote the least singular value of M evolved according to the dynamics (5.42).

Lemma 5.10. Let (M_N) be a sparse matrix ensemble and let W be a Gaussian ensemble of i.i.d. $\mathcal{N}(0, N^{-1})$ variables. Given $\varepsilon > 0$, there exists $\delta > 0$ and a coupling of the processes $\lambda_1(t, M_n)$ and $\lambda_1(t, W)$ such that

$$|\lambda_1(t_a, M_N) - \lambda_1(t_a, W)| \le N^{-1-\delta} \tag{5.47}$$

for some $t_a \le N^\varepsilon/N$ with overwhelming probability.

Proof. Recall that for any t we have

$$H_t \overset{d}{=} H_t^{(1)} + \sqrt{\frac{r(1-\exp(-t/r))}{N}}G \overset{d}{=} H_t^{(1)} + \frac{1}{\sqrt{N}} B_{r(1-\exp(-t/r))}, \tag{5.48}$$

where G is a matrix of i.i.d. standard Gaussians and $B_{r(1-\exp(-t/r))}$ is a matrix Brownian motion considered at the fixed time $r(1-\exp(-t/r))$.

Note that up to a factor of $1 + O(t)$, $H_t^{(1)}$ is a sparse matrix and obeys the weak local semicircle law, Lemma 5.8. Then Lemma 5.8 holds for $(1+O(t))^{-1}H_t^{(1)}$ on the optimal scale $g = N^{-1+\nu}$ for any $\nu > 0$. Further, by invoking Lemma 5.9, this matrix is (g, G)-regular for any such g. The additional factor of $1 + O(t)$ does not affect the (g, G)-regularity of the singular values if $t \le N^{-1-\varepsilon}$ by the argument at the end of the proof of Lemma 6.3 in [26]. In summary, we find that $H_t^{(1)}$ is (g, G)-regular for any $g = N^{-1+\nu}$ with overwhelming probability.

By making ν small enough and choosing constants appropriately in the statement of Theorem 3.2, we may take $t_a \le N^{-1+\varepsilon}$ in the statement of that theorem and apply it to complete the proof. More precisely, we condition on the entries of $H_{t_a}^{(1)}$ and apply Theorem 3.2 to obtain a conditional coupling of the singular value processes. Then, because the hypotheses of Theorem 3.2 hold for the singular values of $H_{t_a}^{(1)}$ with overwhelming probability, by the weak law Lemma 5.8 and Lemma 5.9, we obtain the desired coupling with overwhelming probability after removing the conditioning. □

5.3 Green function comparison

We now control the distribution of the least singular value of a stable matrix ensemble H_N in terms of Green functions. Fix a matrix H from this ensemble. We retain the notation H_t for the dynamics in the previous subsection.

For any $r > 0$, define $\chi_r = \mathbb{1}_{(-r,r)}$. For $\eta > 0$, we set

$$\theta_\eta = \frac{\eta}{\pi(x^2+\eta^2)} = \frac{1}{\pi}\operatorname{Im}\frac{1}{x - i\eta}. \tag{5.49}$$

In particular, for any $r > 0$ we have

$$\operatorname{Tr}\chi_r \star \theta_\eta(H_s) = \frac{N}{\pi}\int_{-r}^{r} \operatorname{Im} m_s(y + i\eta)\, dy. \tag{5.50}$$

We fix $\varepsilon, r > 0$ and set

$$\eta_1 = N^{-1-99\varepsilon}, \quad l = N^{-1-3\varepsilon}, \quad l_1 = lN^{2\varepsilon}, \quad E = \frac{r}{N}. \tag{5.51}$$

We also consider a time parameter t. We are interested in the case $0 \le t \le N^{\varepsilon_0}/N$ for some small $\varepsilon_0 > 0$. For $t \le N^{\varepsilon_0}/N$, note that H_t still satisfies the weak local law at the optimal scale. As described in the proof of Theorem 6.3 of [26], it is a consequence of the weak local law (in particular the fact that $\operatorname{Im} m(z)$ is bounded down to the optimal scale) that there exists C such that, for any interval I with length $|I| \ge N^{-1+\delta}$,

$$|\{\lambda_i(t) \in I\}| \le C|I|N \tag{5.52}$$

holds with overwhelming probability.

Lemma 5.11. Fix t such that $0 \le t \le N^{\varepsilon_0}/N$ and $\varepsilon > 0$. With overwhelming probability,

$$|\operatorname{Tr}\chi_E(H_t) - \operatorname{Tr}\chi_E \star \theta_{\eta_1}(H_t)| \le C\Big(N^{-2\varepsilon} + n(-E-l, -E+l) + n(E-l, E+l)\Big). \tag{5.53}$$

Proof. By the argument in the proof of Lemma 6.1 in [39],

$$|\chi_E(x) - \chi_E \star \theta_{\eta_1}(x)| \le C\eta_1 \left(\frac{2E}{d_1(x)d_2(x)} + \frac{\chi_E(x)}{d_1(x) + d_2(x)}\right) \tag{5.54}$$

where $d_1 = |E - x| + \eta_1$ and $d_2 = |-E - x| + \eta_1$, and the right side is bounded by a constant if $\min\{d_i\} \le l$ and is $O(\eta_1/l)$ if $\min\{d_i\} \ge l$. We obtain by (5.52), on a set of probability greater than $1 - N^{-D}$,

$$\begin{aligned} |\operatorname{Tr}\chi_E(H_t) - \operatorname{Tr}\chi_E \star \theta_{\eta_1}(H_t)| &\le C\Big(\operatorname{Tr} f_1(H_t) + \operatorname{Tr} f_2(H_t) + \frac{\eta_1}{l} n(-E+l, E-l)\Big) &(5.55)\\ &+ C\Big(n(-E-l, -E+l) + n(E-l, E+l)\Big) &(5.56)\\ &\le C\Big(\operatorname{Tr} f_1(H_t) + \operatorname{Tr} f_2(H_t) + \frac{\eta_1 N^{\varepsilon}}{l}\Big) &(5.57)\\ &+ C\Big(n(-E-l, -E+l) + n(E-l, E+l)\Big), &(5.58)\end{aligned}$$

where

$$f_1(x) = \frac{2\eta_1 E}{d_1(x)d_2(x)}\mathbb{1}(x \le -E-l), \quad f_2(x) = \frac{2\eta_1 E}{d_1(x)d_2(x)}\mathbb{1}(x \ge E+l). \tag{5.59}$$

We now describe how to bound $\operatorname{Tr} f_2(H_t)$. The term $\operatorname{Tr} f_1(H_t)$ is similar. For $x \ge E + l$, we have

$$f_2(x) = \frac{2E\eta_1}{(|E-x| + \eta_1)(2E + \eta_1 + |E-x|)} \le \frac{N^{-1-99\varepsilon}}{|E-x|}. \tag{5.60}$$

Set $\alpha = 3 - E$. We consider the N intervals

$$\left[E + l, E + \frac{\alpha}{N}\right], \left[E + \frac{\alpha}{N}, E + \frac{2\alpha}{N}\right], \left[E + \frac{2\alpha}{N}, E + \frac{3\alpha}{N}\right], \dots, \left[3 - \frac{\alpha}{N}, 3\right], \tag{5.61}$$

where the first is of a different size than the rest. By (5.52), each interval contains at most N^{ϵ} eigenvalues. We also consider the interval $[3, \infty]$. Using this decomposition, we obtain

$$\begin{aligned} \operatorname{Tr} f_1(H_t) &\le N^{-1-99\varepsilon}N^{\varepsilon}\left(\frac{1}{l} + \frac{N}{\alpha} + \frac{N}{2\alpha} + \frac{N}{3\alpha} + \dots + \frac{N}{N\alpha}\right) + NN^{-1-99\varepsilon}\\ &\le \frac{N^{-98\varepsilon}}{\alpha}(N^{3\varepsilon} + \log(N)). \end{aligned} \tag{5.62}$$

This completes the proof. □

Lemma 5.12. Fix t such that $0 \le t \le N^{\varepsilon_0}/N$ and $\varepsilon > 0$. There is a constant C such that, with overwhelming probability,

$$\operatorname{Tr} \chi_{E-l_1} \star \theta_{\eta_1}(H_t) - CN^{-\varepsilon} \le \operatorname{Tr} \chi_E(H_t) \le \operatorname{Tr} \chi_{E+l_1} \star \theta_{\eta_1}(H_t) + CN^{-\varepsilon}. \quad (5.63)$$

Proof. We see Lemma 5.11 holds with E replaced by $y \in [E-l, E+l]$. Recall $l_1 = lN^{2\varepsilon}$. Hence, with overwhelming probability,

$$\operatorname{Tr} \chi_E(H_t) \le \frac{1}{l_1} \int_E^{E+l_1} \operatorname{Tr} \chi_y(H_t)\, dy \quad (5.64)$$

$$\le \frac{1}{l_1} \left(\int_E^{E+l_1} \operatorname{Tr} \chi_y(H_t) \star \theta_{\eta_1}(H_t)\, dy + CN^{-2\varepsilon} + Cn(y-l, y+l) + Cn(-y-l, -y+l) \right) dy \quad (5.65)$$

$$\le \operatorname{Tr} \chi_{E+l_1} \star \theta_{\eta_1}(H_t) + CN^{-2\varepsilon} + \frac{Cl}{l_1} \left(n(E-2l_1, E+2l_1) + n(-E-2l_1, -E+2l_1) \right). \quad (5.66)$$

By (5.52), the two counting functions in the above expression are at most CN^{ε}, so we obtain an error of $N^{-\varepsilon}$ and

$$\operatorname{Tr} \chi_E(H_t) \le \operatorname{Tr} \chi_{E+l_1} \star \theta_{\eta_1}(H_t) + N^{-\varepsilon}. \quad (5.67)$$

A matching lower bound is proved similarly. □

As in [33], we fix a smooth function $q : \mathbb{R} \to \mathbb{R}_+$ such that $q(x)$ is decreasing for $x \ge 0$, $q(x) = 1$ for $|x| \le 1/9$, and $q(x) = 0$ for $|x| \ge 2/9$.

Lemma 5.13. Fix t such that $0 \le t \le N^{\varepsilon_0}/N$ and $\varepsilon > 0$. For any $D > 0$, we have

$$\mathbb{E} q(\operatorname{Tr} \chi_{E+l_1} \star \theta_{\eta_1}(H_t)) - N^{-D} \le \mathbb{P}(n(-E, E) = 0) \le \mathbb{E} q(\operatorname{Tr} \chi_{E-l_1} \star \theta_{\eta_1}(H_t)) + N^{-D}. \quad (5.68)$$

Proof. When Lemma 5.12 holds, $n(-E, E) = 0$ implies $\operatorname{Tr} \chi_{E-l_1} \star \theta_{\eta_1}(H_t) \le 1/9$ with overwhelming probability. Hence

$$\mathbb{P}(n(-E, E) = 0) \le \mathbb{P}(\operatorname{Tr} \chi_{E-l_1} \star \theta_{\eta_1}(H_t) \le 1/9) + N^{-D}, \quad (5.69)$$

and Markov's inequality applied to $q(\operatorname{Tr} \chi_{E-l_1} \star \theta_{\eta_1}(H_t))$ yields

$$\le \mathbb{P}(q(\operatorname{Tr} \chi_{E-l_1} \star \theta_{\eta_1}(H_t)) \ge 1) + N^{-D} \le \mathbb{E} q(\operatorname{Tr} \chi_{E-l_1} \star \theta_{\eta_1}(H_t)) + N^{-D}. \quad (5.70)$$

Also, again using Lemma 5.12,

$$\mathbb{E} q(\operatorname{Tr} \chi_{E+l_1} \star \theta_{\eta_1}(H_t)) \le \mathbb{P}(\operatorname{Tr} \chi_{E+l_1} \star \theta_{\eta_1}(H_t) \le 2/9) \quad (5.71)$$

$$\le \mathbb{P}(n(-E, E) \le 2/9 + CN^{-\varepsilon}) + N^{-D} = \mathbb{P}(n(-E, E) = 0) + N^{-D}. \quad (5.72)$$

□

In the work [10], which analyzed the eigenvector moment flow for generalized Wigner matrices and covariance matrices, the authors developed a purely dynamical approach to Green function comparison. We implement it here in the current context. We require the following modification of Lemma A.1 in [10], which asserts the continuity of the above dynamics. The proof is essentially the same.

The deformed matrix $\theta^{ab} H_t$ is defined as

$$(\theta^{ab} H_t)_{kl} = f + \theta^{ab}_{kl}(h_{kl}(t) - f),$$

where $\theta^{ab}_{kl} = 1$ if $k, l \neq a, b$ and some number $0 \leq \theta^{ab}_{kl} \leq 1$ otherwise, where we impose the symmetry condition $\theta^{ab}_{ab} = \theta^{ab}_{ba}$. We define the index set $\mathcal{I}$ to be the entries in the off-diagonal blocks of the $2N \times 2N$ symmetrized matrix:

$$\mathcal{I} = \{(i,j) \colon 1 \leq i, j, \leq 2N, i \leq N < j \text{ or } j \leq N < i\}. \tag{5.73}$$

Lemma 5.14. Let H be a $2N \times 2N$ symmetric matrix, with entries independent up to the symmetry constraint $h_{ij} = h_{ji}$ and the $N \times N$ blocks on the main diagonal all zero. Suppose the other entries satisfy $\mathbb{E}[h_{ij}] = f$ and $\mathbb{E}[(h_{ij} - f)^2] = s_{ij}$ with $cN^{-1} \leq s_{ij} \leq CN^{-1}$. Denote $\partial_{ij} = \partial_{h_{ij}}$. Suppose F is a smooth function of the matrix elements (h_{ij}) in the upper off-diagonal block of H satisfying

$$\sup_{0 \leq s \leq t} \sup_{\theta^{ab}} \mathbb{E}[(N^2|h_{ab}(s) - f|^3 + N|h_{ab} - f|)\partial^{(3)}_{ab} F(\theta^{ab} H_s)] \leq B \tag{5.74}$$

where the supremum is taken over deformations in the off-diagonal block indices $(i,j) \in \mathcal{I}$. Then

$$|\mathbb{E}F(H_t) - \mathbb{E}F(H_0)| \leq CtB. \tag{5.75}$$

We will use Lemma 5.14 to study the expressions appearing above:

$$\mathbb{E}q(\operatorname{Tr} \chi_{E+l} \star \theta_{\eta_1}(H_s)) = \mathbb{E}q\left(\frac{N}{\pi}\int_{-E-l}^{E+l} \operatorname{Im} m(y + i\eta_1)\, dy\right). \tag{5.76}$$

Lemma 5.15. There exists $\varepsilon > 0$ and $c > 0$ such that, for all $t \leq N^{\varepsilon}/N$,

$$|\mathbb{E}q(\operatorname{Tr} \chi_{E+l} \star \theta_{\eta_1}(H_0)) - \mathbb{E}q(\operatorname{Tr} \chi_{E+l} \star \theta_{\eta_1}(H_t))| = O(N^{-c}), \tag{5.77}$$

and a similar statement holds with χ_{E+l} replaced by χ_{E-l}.

Proof. It suffices to control the derivatives of

$$\frac{N}{\pi}\int_{-E-l}^{E+l} \operatorname{Im} m(y + i\eta_1)\, dy \tag{5.78}$$

in order to apply the Lemma 5.14, since the derivatives of q are bounded independent of N. Recall here that $\eta_1 = N^{-1-99\varepsilon}$ is below the natural scale. By Lemma 5.16, which is proved below, we have

$$\mathbb{P}(|\partial^k_{ab} m_s(y + i\eta_1)| \leq CN^{3(k+1)100\varepsilon}) \geq 1 - N^{-D} \tag{5.79}$$

for any $D > 0$ and $s \in [0,t]$, with a deterministic upper bound of $CN^{3(1+100\varepsilon)}$. Then, moving the derivatives inside the integral, bounding the resulting integrand, and using the fact that the factor of N is canceled by the fact the integral is over an interval of length $O(N^{-1})$, we obtain

$$B \leq C(NN^{-\alpha}N^{\sigma_1} + N^{-D}N^{\sigma_2}), \tag{5.80}$$

where σ_1 can be made as small as desired by adjusting ε. Hence $B = O(N^{1-\sigma})$ for some $\sigma > 0$. We conclude by choosing ε small enough in order to make $Bt = o(1)$. ☐

Lemma 5.16. With the notation above,

$$\mathbb{P}(|\partial^k_{ab} m_s(y + i\eta_1)| \leq CN^{3(k+1)100\varepsilon}) \geq 1 - N^{-D}, \tag{5.81}$$

$$|\partial^k_{ab} m_s(y + i\eta_1)| \leq CN^{3(1+100\varepsilon)}. \tag{5.82}$$

Proof. We first suppose that each deformed matrix $\theta^{ab}H_s$ has Green functions elements $G_{ij}(z)$ bounded by a constant C independent of all other parameters for energies $E \in [-1,1]$ and $\eta \geq N^{-1+\varepsilon}$, with overwhelming probability. We fix $s \in [0,t]$ and indices a, b. Let $G(z)$ be the Green function for this matrix. Defining

$$\Gamma(z) = \max_{i,j} |G_{ij}(z)| \vee 1, \tag{5.83}$$

we have by Lemma 2.1 of [6] that

$$\Gamma(E + i\eta_1) \leq N^{100\varepsilon}\Gamma(E + i\eta), \tag{5.84}$$

where $\eta = N^{-1+\varepsilon}$. We have $\Gamma(E + i\eta) \leq C$ for some constant C with overwhelming probability. Given this control over the individual Green functions elements, the argument used in the proof of Lemma 5.2 in [26] proves the first claim. The second claim follows from the trivial deterministic bound

$$G(E + i\eta) \leq \frac{1}{\eta}. \tag{5.85}$$

We now show that we have a uniform constant bound on the deformed G_{ij}. This follows from the proof of Theorem 6.3 in [26]. Note that in that reference the bound is obtained with overwhelming probability for fixed s, and implicitly a standard stochastic continuity argument gives the bound uniformly in s. □

6 Random matrices with correlated entries

In this section, we prove the universality of the least singular value for a class of non-symmetric square random matrices with correlated entries. Our main result is Theorem 6.2. We first sketch the proof of a local law for the smallest singular values. Our method extends the one in [12], and here we consider the more general case of exponentially decaying correlations. We then show how to adapt the method in the previous section on the removal of the time evolution. As a consequence, universality holds for the smallest singular value. The same result can be proved for non-Hermitian random matrices with complex entries in the same way.

6.1 Model

Consider a family of centered real random variables $(x_{ij})_{1\leq i,j\leq N}$ that satisfy

$$\mathbb{E}[x_{ij}x_{kl}] = \frac{1}{N}\xi_{ijkl}, \quad 1 \leq i,j,k,l \leq N.$$

We introduce a sparsity parameter $q = N^\tau$ for some $\tau \in (0,1]$. Assume that for any $p \geq 2$, there is a constant $\mu_p < +\infty$ such that

$$\sup_{i,j} \mathbb{E}[|x_{ij}|^p] \leq \frac{\mu_p^p}{Nq^{p/2-1}}, \quad \forall N \in \mathbb{N}.$$

Furthermore, we assume that ξ has a profile, in the sense that there is a function $\phi : [0,1]^2 \times \mathbb{Z}^2 \to \mathbb{R}$ such that

$$\xi_{ijkl} = \phi(i/N, j/N, k-i, l-j),$$

and that ϕ is piecewise Hölder-continuous with respect to the first two variables. We impose exponential decay on the correlation. For any index set $\mathcal{A} \in \{(i,j) : 1 \leq i,j \leq N\}$

we define $\mathcal{F}_\mathcal{A}$ to be the σ-algebra generated by $(X_{ij})_{(i,j)\in\mathcal{A}}$. For any two index sets $\mathcal{A},\mathcal{B}$ we define their distance by

$$\mathrm{d}(\mathcal{A},\mathcal{B}) = \max_{(i,j)\in\mathcal{A},(i',j')\in\mathcal{B}} |i-i'| \vee |j-j'|.$$

We assume that there are universal constants $c_1, c_2 > 0$ such that for any random variables $Z_1 \in \mathcal{F}_\mathcal{A}, Z_2 \in \mathcal{F}_\mathcal{B}$ with $\mathrm{Var}[Z_1] = \mathrm{Var}[Z_2] = 1$, the following inequality holds,

$$\mathrm{Cov}[Z_1, Z_2] \leq c_1 \exp(-c_2 \mathrm{d}(\mathcal{A},\mathcal{B})). \tag{6.1}$$

As usual we write the symmetrized version of X, namely a $2N$ by $2N$ matrix H defined by

$$H = \begin{bmatrix} 0 & X^* \\ X & 0 \end{bmatrix}.$$

It is easy to see that H is a special case of the model in [12] without the positive definite condition (see Definition 2.2 in [12]), since H has many 0 entries. However, we can still consider an alternative positive definiteness condition in the current case.

Definition 6.1. Let $\Sigma^{(N)} \in \mathbb{R}^{N^2\times N^2}$ be the covariance matrix of the family of real random variables $(x_{ij})_{1\leq i,j\leq N}$ with $\mathbb{E}[X_{ij}X_{kl}] = \xi^{(N)}_{ijkl}$. We say that ξ is positive definite with lower bound $c_0 > 0$ if $\Sigma^{(N)} \geq c_0$ for all N.

We now state the main result of this section.

Theorem 6.2. For the class of correlated sparse matrices whose correlation comes from a positive definite profile function, as defined above, the conclusion of Theorem 2.2 holds.

6.2 Local law

6.2.1 Concentration

Condition 6.1 is weaker than the finite-ranged correlation enforced in [12], because every pair of entries could be correlated, although exponentially weakly. Nevertheless the same concentration estimates hold for linear combinations and quadratic forms (see Lemma 3.6 in [12]).

Lemma 6.3. Let $k \in \mathbb{N}$, $\mathcal{A} \subset \mathbb{N}$ be such that $\mathrm{d}(\{k\},\mathcal{A}) \geq \log^2 N$. Let $(A_i), (B_{ij})$ be families of random variables that are $\mathcal{F}_\mathcal{A}$-measurable with upper bounds $\max_i |A_i| \vee \max_{i\neq j}|B_{ij}| \leq N^{100}$. Then,

$$\left|\sum_i A_i x_{ki} - \mathbb{E}\left[\sum_i A_i x_{ki}\right]\right| \prec \frac{\max_i |A_i|}{\sqrt{q}} + \sqrt{\frac{1}{N}\sum_i |A_i|^2} + \exp(-c\log^2 N).$$

$$\left|\sum_{i,j} B_{ij} x_{ki}x_{kj} - \mathbb{E}\left[\sum_{i,j} B_{ij} x_{ki}x_{kj}\right]\right| \prec \frac{\max_{i\neq j}|B_{ij}|}{\sqrt{q}} + \sqrt{\frac{1}{N}\sum_{i,j} |B_{ij}|^2} + \exp(-c\log^2 N).$$

Proof. Note that x_{ki} is centered, therefore

$$\sum_i \mathbb{E}[A_i x_{ki}] = \sum_i \mathrm{Cov}[A_i, x_{ki}] \leq c_1 \exp(-c_2\log^2 N) \sum_i \left(\mathrm{Var}[A_i]\mathrm{Var}[x_{ki}]\right),$$

which is bounded by $\exp(-c\log^2 N)$ for some constant $c > 0$. Therefore the first inequality is reduced to proving

$$\left|\sum_i A_i x_{ki}\right| \prec \frac{\max_i |A_i|}{\sqrt{q}} + \sqrt{\frac{1}{N}\sum_i |A_i|^2}.$$

We split the sum into $\lfloor \log^2 N \rfloor$ parts, each part being the sum of weakly correlated random variables. Specifically, write $S_l = \sum_i A_{i\lfloor \log^2 N\rfloor + l}$ so that

$$\sum_i A_i x_{ki} = \sum_{1\le l\le \lfloor \log^2 N\rfloor} S_l.$$

Heuristically, each S_l can be viewed as the sum of independent random variables, because the summands have very weak correlation with each other. Let $\tilde{A}_i$, $\tilde{x}_{ki}$ be independent copies of A_i and x_{ki}, and let $\tilde{S}_l$ be defined likewise by replacing A_i, x_{ki} with their copies. It is easy to see that for any $p \ge 2$,

$$\mathbb{E}|S_l|^p = \mathbb{E}|\tilde{S}_l|^p + O(c_p \exp(-c\log^2 N))$$

by expanding out $|S_l|^p$ and collecting all the cross terms, which are exponentially small. This implies that

$$|S_l| \prec |\tilde{S}_l| + O(\exp(-c\log^2 N)).$$

For $\tilde{S}_l$ one can apply Lemma A.1 in [19] to get

$$|\tilde{S}_l| \prec \frac{\max_i |A_i|}{\sqrt{q}} + \sqrt{\frac{1}{N}\sum_i |A_i|^2}.$$

Therefore we summarize the estimates above and see

$$\sum_i A_i x_{ki} \prec \log^2 N \left(\frac{\max_i |A_i|}{\sqrt{q}} + \sqrt{\frac{1}{N}\sum_i |A_i|^2} + O(\exp(-c\log^2 N)) \right).$$

The factor $\log^2 N$ can be absorbed into $\prec$ by definition. This proves the first inequality.

The second inequality follows from a very similar argument. One only needs to split the sum into $O(\log^2 N)$ parts, each of which is the sum of weakly correlated random variables. The weak correlation will not worsen the estimate, as we saw in the proof of the first inequality. □

6.2.2 Self-consistent equation

As usual we define the Green function $G(z)$ by

$$G(z) = (H - z)^{-1}.$$

Define a map $\Xi : \mathbb{C}^{2N\times 2N} \to \mathbb{C}^{2N\times 2N}$ through

$$\Xi(M) = \mathbb{E}[HMH],$$

which can be explicitly defined entrywise:

$$\Xi(M)_{ik} = \sum_{1\le j,\, l\le 2N} \mathbb{E}[h_{ij}h_{kl}M_{jl}].$$

We introduce three control parameters:

$$\Gamma = 1 \vee \max_{i,j} |G_{ij}|, \quad \gamma = 1 \vee \max_i \max_{\mathbb{I},\mathbb{J}} \left\| \left(G^{(\mathbb{J})}_{\mathbb{I},\mathbb{I}} \right)^{-1} \right\|, \quad \Phi = \frac{1}{\sqrt{N\eta}} + \frac{1}{\sqrt{q}}. \tag{6.2}$$

With the help of Lemma 6.3 and repeating the argument of Lemma 3.9 in [12], one can prove an estimate that is the same as Lemma 3.9 in [12]:

$$G(-z - \Xi(G)) = I + O_\prec(\Gamma^5 \gamma^3 \Phi). \tag{6.3}$$

Here the $O_\prec$ notation is in the entrywise sense. We omit the details here, but point out that the only difference in the proof is that the error term here is bigger by a factor of $O(\log^2 N)$, which is negligible in the context of stochastic domination. It remains to show that G is close to the solution M to the following equation.

$$M(-z - \Xi(M)) = I \tag{6.4}$$

The whole argument in [12] goes through except for the part where the positive-definiteness condition on the tensor ξ is used to prove the stability of (6.4) (that is, the solution is stable under small perturbation of the equation). We now discuss the necessary changes.

In [12] the equation was transformed into a continuous version. It was shown by a discretization argument that (6.4) is stable in the bulk if and only if the following continuous equation is stable in the bulk.

$$u(\theta, s) = \frac{1}{-z - Su(\theta, s)}, \quad \theta, s \in [0,1] \tag{6.5}$$

Here the operator S is given by

$$Su(\theta, s) = \iint \hat{\phi}(\theta, \vartheta, s, t)\, d\vartheta\, dt,$$

where $\hat{\phi}$ is the Fourier transform of the symmetrized version of ϕ in the latter two variables (since H is the symmetrized version of X). In particular,

$$\hat{\phi}(\theta, \vartheta, s, t) = \sum_{1 \le k,l \le N} \phi(\theta, \vartheta, k, l) \exp(\mathrm{i}2\pi(sk - tl)), \quad \text{ for } (\theta, \vartheta) \in [0, 1/2] \times [1/2, 1],$$

and $\hat{\phi}(\theta, \vartheta, \cdot, \cdot) = \hat{\phi}(\vartheta, \theta, \cdot, \cdot)$ for $(\theta, \vartheta) \in [1/2, 1] \times [0, 1/2]$ and $\hat{\phi} = 0$ for other (θ, ϑ). In [12] it was shown that $0 < c_0 \le \hat{\phi}(\theta, \vartheta, s, t) \le C_0$ for some universal constants c_0 and C_0 (Lemma 4.15) under the positive definite condition in that paper. Here we use the new positive definite condition (Definition 6.1) to prove upper and lower bounds on $\hat{\phi}$.

Lemma 6.4. Suppose that ξ is positive definite with lower bound $c_0 > 0$ in the sense of Definition 6.1. For $\theta, \vartheta \in [0, 1/2] \times [1/2, 1]$ or $\theta, \vartheta \in [1/2, 1] \times [0, 1/2]$, we have

$$\hat{\phi}(\theta, \vartheta, s, t) \in [c_0, C_0], \quad \forall s, t \in [0,1].$$

Here C_0 is an universal constant depending on $c_1, c_2 > 0$ in (6.1).

Proof. For any $(s,t) \in [0,1]^2$, take an arbitrary real continuous function $g \in C([0,1]^2)$ with $\iint |g|^2 = 1$. For each $N \in \mathbb{N}$ define a random variable

$$Y_N = \frac{1}{N} \sum_{1 \le i,j \le N} h_{ij}\, g\left(\frac{i}{N}, \frac{j}{N}\right) \exp(\mathrm{i}2\pi(si - tj)).$$

By definition of positive definiteness and the decay of correlation, we have $c_0 \le \mathrm{Var} Y_N \le C_0$. One can explicitly compute the variance of Y_N:

$$\mathrm{Var} Y_N = \frac{1}{N^2} \sum_{i,j,k,l} \phi\left(\frac{i}{N}, \frac{j}{N}, k, l\right) g\left(\frac{i}{N}, \frac{j}{N}\right) g\left(\frac{i+k}{N}, \frac{j+l}{N}\right) \exp(\mathrm{i}2\pi(sk - tl)).$$

Let $N \to \infty$ and use the fact that $c_0 \le \mathrm{Var} Y_N \le C_0$. We have

$$\iint g(\theta, \vartheta) \hat{\phi}(\theta, \vartheta, s, t)\, d\theta\, d\vartheta \in [c_0, C_0].$$

Since g was arbitrary, we conclude that $\hat{\phi}(\theta, \vartheta, s, t) \in [c_0, C_0]$. □

The stability of equation (6.5) is very similar to the case in [12] and was analyzed in [2]. It follows from Proposition 3.10 (ii) in [2] that the following estimate holds.

Proposition 6.5. Let u solve (6.5) and u' solve a perturbed version of (6.5), namely

$$u' = \frac{1}{-z - Su'} + r.$$

There exist universal constants $\varepsilon > 0$, and $C > 0$ such that if $|\operatorname{Re} z| \leq \varepsilon, \operatorname{Im} z \in (0, 10], \|u - u'\|_\infty \leq \varepsilon$, then

$$\|u' - u\|_\infty \leq C \|r\|_\infty .$$

Via the discretization method in [12], we can prove the stability for equation (6.4). Below, $\|A\|_\infty$ means $\max_{i,j}|A_{ij}|$.

Proposition 6.6. Let M be the solution to (6.4) and let M' solve a perturbed version of (6.4), namely

$$M'(-z - \Xi(M')) = I + R.$$

There exist universal constants $\varepsilon > 0$, and $C > 0$ such that if $|\operatorname{Re} z| \leq \varepsilon, \operatorname{Im} z \in (0, 10], \|M - M'\|_\infty < \varepsilon$, then,

$$\|M - M'\|_\infty < C \|R\|_\infty .$$

Proposition 6.2.2 and the estimate (6.3) allow us to prove the following theorem.

Theorem 6.7. Let M the the solution of (6.4). There exists a universal constant $\varepsilon > 0$ such that for any $\kappa > 0$,

$$\max_{ij}|G_{ij} - M_{ij}| \prec \Phi,$$

uniformly for all $z \in \{E + \mathrm{i}\eta : \eta \in [N^{-1+\kappa}, 10), E \in [-\varepsilon, \varepsilon]\}$.

Corollary 6.8. There exist universal constants $\varepsilon > 0$ and $c > 0$ such that with overwhelming probability,

$$c \leq \operatorname{Im}\left(\frac{1}{2N} \operatorname{Tr} G\right) \leq c^{-1} \tag{6.6}$$

uniformly for all $z \in \{E + \mathrm{i}\eta : \eta \in [N^{-1+\kappa}, 10), E \in [-\varepsilon, \varepsilon]\}$.

6.3 Universality

Let $(B_{ij}(t))_{1\leq i,j\leq N}$ be a family of Brownian motions that has the same correlation structure as (x_{ij}):

$$\mathbb{E}[B_{ij}(t)B_{kl}(t)] = t\mathbb{E}[x_{ij}x_{kl}] = t\xi_{ijkl}/N. \tag{6.7}$$

Define $x_{ij}(t)$ by the SDE

$$\mathrm{d}x_{ij} = dB_{ij} - \frac{x_{ij}}{2}\, dt.$$

We show that when $t \ll N^{-1}\sqrt{q}$, the evolution does not affect the local statistics of the smallest singular values. Following the argument in [12], we prove the same result as in Lemma 6.1 in [12].

Lemma 6.9. Let $x = (x_k)_{1\leq k\leq m}$ be an array of real centered random variables such that $\sup_k \mathbb{E}[|x_k|^3] \leq \kappa_3^3$ and $\operatorname{Corr}[Z_1, Z_2] \leq c_1 \exp(-c_2 d)$ for all nontrivial random variables $Z_1 \in \sigma(x_1, \cdots, x_k)$, $Z_2 \in \sigma(x_{k+d}, \cdots, x_m)$ and any $1 \leq k \leq k + d \leq m$. Let f be a C^2 function on $\mathbb{R}^m$ with $\left\|D^2 f\right\| \vee \kappa_3 \vee m \leq N^{100}$. Then,

$$\mathbb{E}[f(x)x_i] = \sum_k \mathbb{E}[\partial_k f(x)]\mathbb{E}[x_i x_k] + O\left(\log^2 N\|D^2 f\|_\infty \kappa_3^3 + \exp(-c\log^2 N)\right).$$

Proof. If f is a linear function in x, then the equality is exact without error terms. In general, define

$$\mathbb{T} = \{j : |i-j| \le \log^2 N\}, \quad \mathbb{U} = \{j : |i-j| \le 2\log^2 N\}.$$

Denote $x^{(\mathbb{T})} = (x_k \mathbb{1}_{k \notin \mathbb{T}})$, $x^{(\mathbb{U})} = (x_k \mathbb{1}_{k \notin \mathbb{U}})$. By Taylor's expansion,

$$f(x) = f(x^{(\mathbb{T})}) + \sum_{k \in \mathbb{T}} \partial_k f(x^{(\mathbb{T})}) x_k + \frac{1}{2} \sum_{k,l \in \mathbb{T}} \int_0^1 (1-t) \partial_{kl} f(x^{(\mathbb{T})} + t(x - x^{(\mathbb{T})})) x_k x_l \, dt.$$

We expand the second term further,

$$\sum_{k \in \mathbb{T}} \partial_k f(x^{(\mathbb{T})}) x_k = \sum_{k \in \mathbb{U}} \partial_k f(x^{(\mathbb{T})}) x_k + \sum_{k \in \mathbb{T}, l \in \mathbb{U}} \int_0^1 (1-t) \partial_{kl} f(x^{(\mathbb{U})} + t(x^{(\mathbb{T})} - x^{(\mathbb{U})})) x_k x_l \, dt.$$

Therefore, $f(x)$ can be written as

$$f(x) = f(x^{(\mathbb{T})}) + \sum_{k \in \mathbb{U}} \partial_k f(x^{(\mathbb{U})}) x_k + O\left(\sum_{k,l \in \mathbb{U}} \sup_{\theta \in [0,1]^{\mathbb{U}}} |f(\theta x)| |x_k x_l| \right). \tag{6.8}$$

Here $\theta x = (x_k \mathbb{1}_{k \in \mathbb{U}} \theta_k + x_k \mathbb{1}_{k \notin \mathbb{U}})$. Therefore, we can compute

$$\mathbb{E}[f(x) x_i] = \mathbb{E}[f(x^{(\mathbb{T})}) x_i] + \sum_{k \in \mathbb{T}} \mathbb{E}[\partial_k f(x^{(\mathbb{U})}) x_k x_i] + O\left(\left\| D^2 f \right\|_\infty \kappa_3^3 \right).$$

The first and second term are expectations of products of weakly correlated random variables. So,

$$\mathbb{E}[f(x) x_i] = \sum_{k,l} \mathbb{E}[\partial_{kl} f(x^{(\mathbb{U})})] \mathbb{E}[x_k x_l] + O\left(\left\| D^2 f \right\|_\infty \kappa_3^3 + \exp(-c \log^2 N) \right).$$

Using Taylor expansion again, we can replace $\partial_{kl} f(x^{(\mathbb{U})})$ by $\partial_{kl} f(x)$ with the cost of a small error term. Therefore,

$$\mathbb{E}[f(x) x_i] = \sum_{k,l} \mathbb{E}[\partial_{kl} f(x)] \mathbb{E}[x_k x_l] + O\left(\left\| D^2 f \right\|_\infty \kappa_3^3 + \exp(-c \log^2 N) \right).$$

□

We can similarly imitate the proof of Lemma 6.2 of [12] to prove the following.

Lemma 6.10. Suppose f is a C^3 function on $\mathbb{R}^{N \times N}$. Then

$$\mathbb{E}[f(H_t) - f(H_0)] = O\left(\exp(-c \log^2 N) + t N q^{-1/2} \mathbb{E} \sup_\theta \partial^{(k)} f(\theta H) \right). \tag{6.9}$$

The rest of the argument is essentially the same as the one in Section 5. Similarly to Section 5.2, we may decompose the correlated OU dynamics (6.7) as

$$x_{ij}(t) = \tilde{x}_{ij}(t) + w_{ij}, \tag{6.10}$$

where the (w_{ij}) are i.i.d. Gaussian and the $\tilde{x}_{ij}(t)$ have a positive definite correlation structure. Then we see the local law holds for $\tilde{X}(t) = (\tilde{x}_{ij}(t))$, and the rest of the arguments in Section 5 go through, since they do not rely on the structure of the matrix. The only necessary change is that in the proof of Lemma 5.16 we need to use a different method to show the (non-deformed) Green function entries are uniformly bounded by a constant down to the scale $N^{-1+\delta}$. Using the local law, this reduces to showing the entries of M are bounded, and this is a consequence of the definition of M and the analogue of Lemma 4.20 in [12]. The regularity necessary for Theorem 3.2 is provided by Corollary 6.8 and the analogue of Lemma 5.9 for the correlated model in this Section. The analogue is proved by splitting into $\log^4(N)$ weakly correlated matrices and applying the argument in the proof of Lemma 5.9. This completes the proof of Theorem 6.2.

A Singular value dynamics

This appendix collects information on the SDE

$$d\lambda_k = \frac{1}{\sqrt{N}} dB_k + \frac{1}{2N} \sum_{j \neq k} \left(\frac{1}{\lambda_k - \lambda_j} + \frac{1}{\lambda_k + \lambda_j} \right) dt. \tag{A.1}$$

A.1 Existence and uniqueness of solutions

Let Δ be the region where $\lambda_1 < \lambda_2 < \cdots < \lambda_N$ and $|\lambda_1| < |\lambda_2| < \cdots < |\lambda_N|$. We show that given initial data in Δ, there is a unique strong (continuous) solution that stays in Δ for all time. We follow the arguments in Section 4.3 of [3], explaining the necessary changes. We also show the solutions of this equation are the singular values of a matrix Brownian motion process.

Throughout, N will be fixed, and $\lambda(t) = (\lambda_1(t), \ldots, \lambda_N(t))$.

Lemma A.1. Fix an initial condition $\lambda(0) \in \Delta$. There exists a unique strong solution $(\lambda(t)_{t \geq 0}) \in C(\mathbb{R}^+, \Delta)$.

Proof. In the proof of Lemma 4.3.3 in [3], replace the given definition of f with

$$f(x) = \frac{1}{N} \sum_i x_i^2 - \frac{1}{2N^2} \sum_{i \neq j} \log |x_i - x_j| + \log |x_i + x_j|. \tag{A.2}$$

Then the estimates (4.3.6) given still hold, and

$$df(\lambda^R(t)) = \sum_{i=1}^N \partial_i f(\lambda^R(t))\, d\lambda_i^R + \frac{1}{2} \sum_{i,j} \partial_i \partial_j f(\lambda^R(t))\, d\langle \lambda_i^R, \lambda_j^R \rangle. \tag{A.3}$$

As in [3], we define

$$u_{i,1} = \sum_{k \neq i} \frac{1}{x_i - x_k}, \quad u_{i,2} = \sum_{k \neq i} \frac{1}{(x_i - x_k)^2}, \tag{A.4}$$

$$\overline{u}_{i,1} = \sum_{k \neq i} \frac{1}{x_i + x_k}, \quad \overline{u}_{i,2} = \sum_{k \neq i} \frac{1}{(x_i + x_k)^2}. \tag{A.5}$$

We have the identities

$$\sum x_i (u_{i,1}(x) + \overline{u}_{i,1}(x)) = N(N-1) \tag{A.6}$$

$$\sum (u_{i,1} + \overline{u}_{i,1})^2 - u_{i,2} - \overline{u}_{i,2} = 0. \tag{A.7}$$

The equation becomes (suppressing the λ_i^R in the arguments of the u functions):

$$df(\lambda^R(t)) = 1\, dt + \frac{1}{N^2} \sum_i \left(\lambda_i^R - \frac{1}{2N} (u_{i,1} + \overline{u}_{i,1}) \right) u_{i,1} dt + \frac{1}{N^3} \sum u_{i,2}\, dt + dM(t) \tag{A.8}$$

$$= 1 + \frac{1}{2} - \frac{1}{N} + dM(t). \tag{A.9}$$

Since M is a martingale with zero expectation,

$$E[f(\lambda^R(t \wedge T_M)] \leq 1.5 E[t \wedge T_M] + f(\lambda^R(0)). \tag{A.10}$$

Now we are finished, as in [3], by a Borel-Cantelli argument. □

We now explain why solutions to (A.1) have the same distribution as the singular values of M_t from Section 3. We recall the SDE for the eigenvalues of $M_t^\dagger M_t$ given in Appendix C of [10]:

$$d\lambda_k = 2\sqrt{\lambda_k}\frac{dB_k}{\sqrt{N}} + \left(1 + \sum_{j \neq k} \frac{\lambda_k + \lambda_l}{\lambda_k - \lambda_l}\right) dt. \tag{A.11}$$

Existence and uniqueness of solutions to (A.11) was shown in [11].

An application of Itô's lemma to (A.1) yields

$$d(\lambda_k^2) = \frac{2\lambda_k}{\sqrt{N}}\, dB_k + 1\, dt + \sum_{j \neq k} \frac{\lambda_k^2 + \lambda_j^2}{\lambda_k^2 - \lambda_j^2}\, dt. \tag{A.12}$$

Then λ_k^2 almost solves the SDE given in [10], except here we are not always choosing the positive square root of λ_k^2. However, we obtain a weak solution by noting that λ^2 solves (A.11) with the Brownian motions chosen as $\overline{B}_k = \mathrm{sgn}(\lambda_k) B_k$. (Note that by the Lévy criterion, the $\overline{B}_k$'s are indeed independent Brownian motions.) Hence the solutions of (A.1) have the desired distribution.

A.2 Interpolation

Here we provide the details of the construction of the interpolated solutions (3.13) and show they are differentiable with respect to α.

We first construct solutions $z_i(t, \alpha)$ for $\alpha \in \mathbb{Q} \cap [0, 1]$ using the argument in the previous subsection. Because there are a countable number of solutions, each of which exists individually except possibly on some set of measure zero in the probability space Ω, they all exist and satisfy the SDE simultaneously on a set of full measure $E_1 \subset \Omega$. For $\alpha_1, \alpha_2 \in \mathbb{Q} \cap [0, 1]$, define

$$\tilde{u}_i(t, \alpha_1, \alpha_2) = z_i(t, \alpha_1) - z_i(t, \alpha_2). \tag{A.13}$$

Then

$$\partial_t \tilde{u}_i(t) = \sum_j B_{ij}(\tilde{u}_j - \tilde{u}_i), \quad B_{ij} = \frac{\mathbb{1}_{i \neq \pm j}}{2N(z_i(\alpha_1) - z_j(\alpha_1))(z_i(\alpha_2) - z_j(\alpha_2))}, \tag{A.14}$$

$$\tilde{u}_i(0) = (\alpha_1 - \alpha_2)(z_i(0, 1) - z_i(0, 0)). \tag{A.15}$$

Suppose $\tilde{u}_k(t)$ is a particle where $\tilde{u}_k(t) = \max_i \tilde{u}_i(t)$. Because the particles are ordered, the coefficients B_{ij} are positive, so $\partial_t \tilde{u}_k(t) \leq 0$. We conclude that $\|\tilde{u}(t)\|_\infty$ is non-increasing, and

$$\|\tilde{u}(t)\|_\infty \leq (\alpha_1 - \alpha_2)(\|z(0, 1)\|_\infty + \|z(0, 0)\|_\infty) \tag{A.16}$$

holds uniformly for all t on E_1.

Since $z_i(t, \alpha)$ is Lipschitz in α (with Lipschitz constant depending on the random initial data), it extends uniquely to a (random) function $z(t, \alpha)$ continuous in $\alpha \in [0, 1]$. Further, since the uniform limit of continuous functions is continuous, the Lipschitz estimate shows the paths in the variable t are continuous for all α.

Fix $\alpha_0 \in [0, 1]$. If $\tilde{z}_i(t, \alpha_0)$ is a solution a.s., then the same reasoning that led to (A.16) shows that $\tilde{z}_i(t, \alpha_0, \omega) = z_i(t, \alpha_0, \omega)$ for a set of full measure in Ω. (Note, however, that this set of full measure may vary with the choice of α_0). By Fubini's theorem, $z_i(t, \alpha, \omega)$ is solution for a set of (ω, α) of full measure in the product space $\Omega \times [0, 1]$. This completes the construction of the interpolated solutions.

Fix $\omega \in E_1$. Since Lipschitz functions are differentiable almost everywhere and their derivatives satisfy the fundamental theorem of calculus, we see that $\partial_\alpha z_i(t, \alpha, \omega)$ exists for almost every $\alpha \in [0, 1]$ (with the exceptional set depending on ω) and

$$z_i(t, 1, \omega) - z_i(t, 0, \omega) = \int_0^1 \partial_\alpha z(t, \alpha, \omega)\, d\alpha.$$

Hence this relation holds for every $\omega \in E_1$ and therefore almost surely.

References

[1] B. Adlam. The local Marchenko–Pastur law for sparse covariance matrices. http://www.people.fas.harvard.edu/~adlam/thesis.pdf, 2013.

[2] J. Alt, L. Erdős, and T. Krüger. Local law for random Gram matrices. *Electronic Journal of Probability*, 22(25), 2017. MR-3622895

[3] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2010. MR-2760897

[4] Z. Bai and J. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Science Press, 2006. MR-2567175

[5] A. Basak and M. Rudelson. Invertibility of sparse non-Hermitian matrices. *Advances in Mathematics*, 310:426–483, 2017. MR-3620692

[6] R. Bauerschmidt, A. Knowles, and H.-T. Yau. Local semicircle law for random regular graphs. *Comm. Pure Appl. Math.*, 70:1898–1960, Oct. 2017. MR-3688032

[7] F. Benaych-Georges and A. Knowles. Lectures on the local semicircle law for Wigner matrices. *Preprint arXiv:1601.04055*, 2016. MR-3792624

[8] P. Biane. On the free convolution with a semi-circular distribution. *Indiana Univ. Math. J.*, 46(3):705–718, 1997. MR-1488333

[9] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin. Fixed energy universality for generalized Wigner matrices. *Comm. Pure Appl. Math.*, Dec. 2015. MR-3541852

[10] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicity. *Comm. Math. Phys.*, 2013. MR-3606475

[11] M.-F. Bru. Diffusions of perturbed principal component analysis. *Journal of Multivariate Analysis*, 29(1):127–136, 1989. MR-0991060

[12] Z. Che. Universality of random matrices with correlated entries. *Electronic Journal of Probability*, 22(30):1–38, 2017. MR-3629874

[13] T. Claeys, T. Neuschel, and M. Venker. Boundaries of sine kernel universality for Gaussian perturbations of Hermitian matrices. *arXiv preprint arXiv:1712.08432*, 2017.

[14] N. Cook. Lower bounds for the smallest singular value of structured random matrices. *The Annals of Probability*, 46(6):3442–3500, 2018. MR-3857860

[15] I. Dumitriu and Y. Zhu. Sparse general Wigner-type matrices: Local law and eigenvector delocalization. *arXiv preprint arXiv:1808.07611*, 2018.

[16] Erdős, S. Péché, J. A. Ramirez, and B. Schlein. Bulk universality for Wigner matrices. *Comm. Pure Appl. Math.*, 63(7):895–925, 2010. MR-2662426

[17] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Inst. Henri Poincaré*, 14(8):1837–1926, 2013. MR-3119922

[18] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electronic Journal of Probability*, 18(59), 2013. MR-3068390

[19] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős–Rényi graphs I: local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013. MR-3098073

[20] L. Erdős and H.-T. Yau. Dynamical approach to random matrix theory. *Courant Lecture Notes in Mathematics*, 28, 2017. MR-3699468

[21] L. Erdős, B. Schlein, and H.-T. Yau. Universality of random matrices and local relaxation flow. *Invent. Math.*, 185(1):75–119, 2011. MR-2810797

[22] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Annales de l'I.H.P. Probabilités et statistiques*, 48(1):1–46, 2012. MR-2919197

[23] L. Erdős and H.-T. Yau. Universality of local spectral statistics of random matrices. *Bull. Amer. Math. Soc.*, 49(3):377–414, 2012. MR-2917064

[24] D. Grabiner. Brownian motion in a Weyl chamber, non-colliding particles, and random matrices. *Ann. Inst. Henri Pointcarè Probab. Stat.*, 35(2):177–204, 1999. MR-1678525

[25] Y. He, A. Knowles, and M. Marcozzi. Local law and complete eigenvector delocalization for supercritical Erdős–Rényi graphs. *arXiv preprint arXiv:1808.09437*, 2018.

[26] J. Huang, B. Landon, and H.-T. Yau. Bulk universality of sparse random matrices. *J. Math. Phys.*, 56(12):123301, 2015. MR-3429490

[27] B. Landon, P. Sosoe, and H.-T. Yau. Fixed energy universality for Dyson Brownian motion. *Preprint arXiv:1609.09011*, 2016.

[28] B. Landon and H.-T. Yau. Convergence of local statistics of Dyson Brownian motion. *Comm. Math. Phys.*, 355:949?1000, Nov. 2017. MR-3687212

[29] J. O. Lee, K. Schnelli, B. Stetler, and H.-T. Yau. Bulk universality for deformed Wigner matrices. *Ann. Probab.*, 44(3):2349–2425, 2016. MR-3502606

[30] E. H. Lieb and M. Loss. Analysis, volume 14 of Graduate Studies in Mathematics. *American Mathematical Society, Providence, RI,*, 4, 2001. MR-1817225

[31] A. Litvak and O. Rivasplata. Smallest singular value of sparse random matrices. *Studia Mathematica*, 212(3), Jun. 2011. MR-3009072

[32] V. A. Marchenko and L. A. Pastur. The distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.*, 72:507–536, 1967. MR-0208649

[33] N. S. Pillai and J. Yin. Universality of covariance matrices. *The Annals of Applied Probability*, 24(3):935–1001, 2014. MR-3199978

[34] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. *arXiv preprint arXiv:1003.2990*, 2010. MR-2827856

[35] D. Spielman and S. Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians*, volume I, pages 597–606, Beijing, 2002. Higher Ed. Press. MR-1989210

[36] T. Tao and V. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, 169:595–632, 2009. MR-2480613

[37] T. Tao and V. Vu. Random matrices: the distribution of the smallest singular value. *Geometric and Functional Analysis*, 20(1):260–297, 2010. MR-2647142

[38] T. Tao and V. Vu. Smooth analysis of the condition number and the least singular value. *Mathematics of computation*, 79(272):2333–2352, 2010. MR-2684367

[39] J. Yin, L. Erdős, and H.-T. Yau. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math*, 229(3):1435–1515, 2012. MR-2871147

Acknowledgments. The authors thank H.-T. Yau for suggesting the problem, useful discussions, and helpful comments on our preliminary draft. P.L. thanks A. Aggarwal, B. Landon, and P. Sosoe for useful discussions. Finally, the authors thank the referee for their comments and suggestions, which significantly improved the paper.