

Edge universality of correlated Gaussians

Arka Adhikari*

Ziliang Che[†]

Abstract

We consider a Gaussian random matrix with correlated entries that have a power law decay of order $d > 2$ and prove universality for the extreme eigenvalues. A local law is proved using the self-consistent equation combined with a decomposition of the matrix. This local law along with concentration of eigenvalues around the edge allows us to get a bound for extreme eigenvalues. Using a recent result of the Dyson-Brownian motion, we prove universality of extreme eigenvalues.

Keywords: universality; random matrix; correlation.

AMS MSC 2010: 15B52; 82B44.

Submitted to EJP on March 27, 2018, final version accepted on January 31, 2019.

1 Introduction

The Wigner-Dyson-Mehta conjecture asserts that the local eigenvalue statistics of large random matrices are universal in the sense that they depend only on the symmetry class of the model - real symmetric or complex Hermitian - but are otherwise independent of the underlying details of the model. There are two types of universality results. Bulk universality involves the spacing distribution eigenvalues that lie well within the support of the limiting spectral distribution, while edge universality involves the extreme eigenvalues.

There has recently been a lot of progress made in proving the Wigner-Dyson-Mehta conjecture in a increasingly large class of models. In [7, 8, 10, 11, 14, 13], universality was proved for Wigner matrices whose entries are independent and have identical variance; parallel results are obtained independently in various cases in [18, 17]. In [3, 1], this type of result was extended to more general variance patterns, while still maintaining the independence of matrix entries.

Most of the previous works rely heavily on the independence between matrix entries, and deal with bulk universality. Only recently have people proved results on models with

*Harvard University, United States of America. E-mail: adhikari@math.harvard.edu

[†]Harvard University, United States of America.
E-mail: che@math.harvard.edu

general correlation structure. In [6, 4, 5], bulk universality is proved for matrices where the correlation decays fast enough, e.g., faster than polynomial decay. In a recent paper [9], Erdős et al. consider a model where the correlation between matrix entries has a power law decay of order $d \geq 12$ in the long range and $d \geq 2$ in the short range. They use a combinatorial expansion to get optimal local law, then prove bulk universality. They remark in Example 2.12 that in the Gaussian case, $d \geq 2$ for both long range and short range correlation is sufficient to satisfy the assumptions of their main theorem.

In this paper, we prove edge universality for Gaussian matrices with a correlation structure that decays as a power law of order $d > 2$, namely $|\mathbb{E}[h_{ij}h_{kl}]| \leq \frac{1}{|i-l|^d + |j-k|^d}$ where h_{ij} are the entries of the random matrix H .

By a standard rule of thumb, it is much easier to prove universality when the matrix being considered is closer to one such that all elements are independent. When we have stronger correlations it becomes significantly more difficult to apply the few known algebraic techniques to derive a local law. When one applies row removal to get a self-consistent equation, the correlations make it significantly more difficult to bound the error term. When applying the loop equation, one would also derive poor concentration bounds through naive choices of the integrating region. It is believed that $d > 2$ is the optimal region where one might still be able to get universality estimates. It becomes most difficult to prove the local law in this regime and one has to be very careful with controlling errors.

In order to deal with the difficulty involved in deriving some of these error estimates, it becomes necessary to expand the matrix to large orders and perform a sophisticated combinatorial expansion of the matrix entries of the Green's function. We give a proof of this result that does not require this expansion or a corresponding fluctuation averaging type ¹ result, instead we rely on a decomposition of Gaussian random variable.

Correlated matrices regularly occur in various statistical applications. A population researcher may seek to determine the existence of various subpopulations, where the correlations within one subpopulation are greater than those between different subpopulations. Biological researchers studying genetic history of species or protein structure and gene expression use these sorts of correlations to determine the genealogical relationship of species or to construct a map between parts of DNA and the protein it encodes. Since the decay $d > 2$ is optimal, proving universality in this regime would get a statement that would be robust for all possible applications.

We follow a three step strategy to prove universality:

1. Prove a local law for the empirical eigenvalue distribution at small scales.
2. Study the convergence of the DBM (Dyson-Brownian motion) in short time scales to local equilibrium.
3. Prove that the eigenvalue spacing distribution does not change too much during the short time evolution of DBM.

Step 1, finding the local law, is generally the most difficult and model dependent. The strategy in proving this local law is deriving a self-consistent equation for the Green's function $G = (H - z)^{-1}$.

One can heuristically derive a self consistent equation by taking expectation and performing integration by parts on $G(H - z) = I$. One notices that there is a linear operator

¹Fluctuation Averaging is a central limit theorem type result on the diagonal elements of the Green's function. Since these elements are correlated, a proof of this averaging requires analyzing higher order terms in a combinatorial expansion of the Green's function. See [12] Ch 10.

S such that $\mathbb{E}[G(-S(G) - z)] = 1$. Removing the expectation creates some error term. The goal is to show that a small error exists with high probability on our matrix ensemble, as is done in [4, 6].

From [6], it is known that the self-consistent equation for correlated matrix entries is of the form $G(-S(G) - z) = I$. This can be transformed into the following vector equation via local Fourier transform.

$$g(x)(-\Psi(g)(x) - z) = 1, \quad x \in [0, 1]^2. \quad (1.1)$$

where $\Psi : L^\infty([0, 1]^2) \rightarrow L^\infty([0, 1]^2)$ is an integral operator, which is the continuous version of S . There are two difficulties in our case: getting a small error for our self-consistent equation and proving the stability of the equation near the edge.

In order to get a small error for the self-consistent equation, we avoid the procedure of removing blocks of elements, which requires combinatorial expansion, but instead applied integration by parts and concentration results along a careful decomposition of the probability space. This gives us a weak local law which can be bootstrapped to give an even better bound for the expected value of the Green's function. Once we have bounds on the expected value, we use the concentration of eigenvalues about its mean value in order to show a version of upper bound for the top eigenvalue along the edge.

In order to prove the stability, we first embed the matrix space into the continuous space $C^\infty([0, 1]^2)$, up to small errors. However, entry-wise error is not small enough to allow this embedding. We noticed the fact that the operator S has a smoothing effect and will reduce the error. The smoothing effect of S is due to the fact that it acts as a convolution with a decaying function; this convolution effectively regularizes the error. A double iteration of the operator $F(G) = (-S(G) - z)^{-1}$ created a matrix $F(F(G))$ that satisfies

$$F(F(G)) = F(F(F(G))) + R, \quad (1.2)$$

where R has sufficiently fast decay on off-diagonal entries. A similar strategy based on the smoothing effect of F is also used in [5]. Then we can embed and apply stability of the continuous solution. In order to prove the decay properties of the double iteration, we applied a perturbation around a fixed matrix that is known to have decay of matrix entries.

With sufficiently strong upper bounds on the top eigenvalue and lower bounds on the bottom eigenvalue, we are able to use the result of [16] to get universality for the extreme eigenvalues conditioned on a sub- σ -field, where the scaling factors and edge locations vary. Then we prove the existence of a scaling factor uniform for all matrices in the model, which gives us the final universality statement.

The main novelties in this paper are: the usage of a Gaussian random variable decomposition, which allows us to get better concentration estimates for the Green's function about its expected value; the extension of the result of [16] via uniformization of scaling factors; improving the analysis of stability of the self-consistent equations in the slow-decay regime where $d > 2$.

The structure of this paper is as follows. The second section is devoted to proving a self-consistent equation with sufficiently small error. The third section of this paper involves proving stability of the self-consistent equation to get a local law to prove an upper bound on eigenvalues. The final section uses this upper bound in order to prove universality.

2 Derivation of self-consistent equation

2.1 The model and assumptions

Before we start defining our model, we will start with some preliminary notation. As is standard, the GOE is a matrix $(Z_{ij} + Z_{ji})$ where the Z_{ij} are i.i.d. copies of an i.i.d. random variable $N(0, \frac{1}{2N})$.

We need a notion of toroidal distance in order to later define correlation decay. In the sequel, we will use letters i, j, k, l as indices for the matrix entries; for an N by N matrix, we view these indices as elements in $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$. On $\mathbb{Z}/N\mathbb{Z}$ we define the natural distance $\text{dist}_{\mathbb{Z}/N\mathbb{Z}}(i, j) := \min\{|i - j + kN| \mid k \in \mathbb{Z}\}$, which for simplicity of notation we still denote by $|i - j|$ unless there is danger of confusion.

Definition 2.1 (Correlated Gaussian Ensembles). *For $N \in \mathbb{N}$, we consider a symmetric matrix $H = (h_{ij}^{(N)})_{1 \leq i, j \leq N}$ whose entries are centered Gaussian random variables with correlation $\xi_{ijkl}^{(N)} := N\mathbb{E} [h_{ij}^{(N)} h_{kl}^{(N)}]$. For simplicity of notation we omit superscript (N) notation.*

We need the following assumptions on the behavior of the ξ_{ijkl}

1. We have a limiting profile for the covariances. Assume there is a Lipschitz function $\phi : \mathbb{T}^2 \times \mathbb{Z}^2 \rightarrow \mathbb{R}$ such that

$$\xi_{ijkl} = \phi(i/N, j/N, k - i, l - j) + O(N^{-1}), \forall i \leq k, j \leq l. \quad (2.1)$$

2. There is correlation decay

$$|\xi_{ijkl}| \leq c_1^2 \max \left\{ \frac{1}{(|i - k| + |j - l| + 1)^d}, \frac{1}{(|i - l| + |j - k| + 1)^d} \right\}. \quad (2.2)$$

for some $d > 2$.

3. We have a nonsingularity condition. We assume that the covariance matrix ξ_{ijkl} is strictly positive semidefinite.

$$[\xi]_{ijkl} > 0 \quad (2.3)$$

Since we are dealing with Gaussian matrices, this is equivalent to $c_2 > 0$, such that H allows a decomposition

$$H = c_2 X + Y, \quad (2.4)$$

where X is a GOE matrix independent from Y .

Remark 2.2. The third nonsingularity condition was used to gain control of an inverse operation during our study of the loop equation.

There are examples of Gaussian ensembles with correlation decay that do not satisfy universality. As a simple counterexample, consider the following ensemble matrix

$$H = G \otimes X \quad (2.5)$$

where G is a GOE matrix and X can be the 2 by 2 identity matrix or the 2 x 2 matrix with all 1s. If X were the identity, all eigenvalues appear with double multiplicity and universality cannot hold. If X was the all 1 matrix, then it would have half rank. The small GOE component causes sufficient spreading of the eigenvalues to get universality.

Remark 2.3. Though writing our ensemble in the form $H = c_2 X + Y$ where X and Y are independent would suggest that we use results from free probability to prove a local law, we would need to have local law results on the matrix Y . However, here, Y is a Gaussian matrix of a similar type as H and we do not know a priori good local laws for Y . Our method here avoids this issue of infinite regress.

We fix as a control parameter an arbitrary $\alpha \in (2, d)$. We say that a constant is universal if it only depends on c_1, c_2, d, α and ϕ . In this paper, we denote $a \lesssim b$ if there is a universal constant $c > 0$ such that $a \leq cb$. We also denote $a \sim 1$ if $a \lesssim 1$ and $1 \lesssim a$.

For $\beta > 0$ and any matrix A (finite square or infinite) we define the following norms,

$$\|A\|_\beta := \sup_{i,j} (|A_{ij}|(1 + |i - j|)^\beta), \quad |A|_\infty := \max_{i,j} |A_{ij}|. \tag{2.6}$$

Our main result is the following.

Theorem 2.4. *Let H be a Correlated Gaussian Ensemble as in Definition 2.1.*

Let $\lambda_1 \leq \dots \leq \lambda_N$ be the eigenvalues of H . Let $\hat{\lambda}_1 \leq \dots \leq \hat{\lambda}_N$ be the eigenvalues of an N by N GOE matrix.

There exists a universal constant γ such that for any $f \in C^1(\mathbb{R}^{k-1})$, the following inequality holds for N large enough for some small $c > 0$.

$$\begin{aligned} &|\mathbb{E}[f(\gamma N^{2/3}(\lambda_2 - \lambda_1), \dots, \gamma N^{2/3}(\lambda_k - \lambda_1))] - \\ &\mathbb{E}[f(N^{2/3}(\hat{\lambda}_2 - \hat{\lambda}_1), \dots, N^{2/3}(\hat{\lambda}_k - \hat{\lambda}_1))]| \leq N^{-c} \end{aligned} \tag{2.7}$$

2.2 The loop equation

The following lemma is one of the building blocks of the loop equation.

Lemma 2.5. *Let $Z = (Z_k)_{k=1}^p$ be a centered Gaussian random vector in \mathbb{R}^p with covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$. Let $f \in C^1(\mathbb{R}^p)$. Then,*

$$\mathbb{E}[f(Z)Z_l] = \sum_{k=1}^p \mathbb{E}[\partial_k f(Z)] \Sigma_{kl}, \forall 1 \leq l \leq p.$$

Proof. This directly follows from an identity known as Stein’s lemma, which says that if $X \sim N(0, 1)$ and $h \in C^1(\mathbb{R})$, then $\mathbb{E}[h(X)X] = \mathbb{E}[h'(X)]$. □

We also use the following decomposition lemma, which allows us to construct a special sigma algebra. We then derive the loop equation by taking the conditional expectation with respect to this sigma algebra. As we will see later, this sigma algebra has the benefit that we can get strong concentration bounds upon removing conditional expectation.

Lemma 2.6. *Let $Z = (Z_k)_{k=1}^p$ be a centered Gaussian random vector. Let $1 \leq q < p$. Then, there is a constant matrix $(a_{kl})_{1 \leq l \leq q, q+1 \leq k \leq p}$ such that*

$$Z_k = \sum_{l=1}^q a_{kl} Z_l + \tilde{Z}_k,$$

where the collection $(Z_l)_{l=1}^q \cup (\tilde{Z}_k)_{k=q+1}^p$ forms a Gaussian random vector such that the latter $(\tilde{Z}_k)_{k=q+1}^p$ are independent from the former $(Z_l)_{l=1}^q$

Proof. This can be done by a carefully chosen linear transform. □

We start with the trivial matrix identity $G(H - z) = I$, which can be written as follows

$$\sum_k G_{ik} h_{kj} - z G_{ij} = \delta_{ij}, \quad i, j \in \mathbb{Z}/N\mathbb{Z}. \tag{2.8}$$

Without loss of generality, fix $j = 1$. According to Lemma 2.6 we may write,

$$h_{ab} = \sum_{k=1}^N \gamma_{abk1} h_{k1} + \tilde{h}_{ab}, \tag{2.9}$$

where \tilde{h}_{ab} is a Gaussian random variable that is independent from $(h_{k1})_{1 \leq k \leq N}$. In particular, $\gamma_{a1k1} = \delta_{ak}$, $\tilde{h}_{a1} = 0$, $\forall a \in \mathbb{Z}/N\mathbb{Z}$. In order to apply Lemma 2.5 on (2.8), let \mathcal{F}_1 be the σ -algebra generated by $(\tilde{h}_{ab})_{a \neq 1, b \neq 1}$. Define the conditional expectation operator

$$\mathbb{E}_1[\cdot] := \mathbb{E}[\cdot | \mathcal{F}_1].$$

We will then be able to apply Lemma 2.5 to get the following

$$\delta_{i1} = \sum_k \mathbb{E}_1[G_{ik} h_{k1}] - z \mathbb{E}_1[G_{i1}] = -\frac{1}{N} \sum_{k,a,b} \mathbb{E}_1[G_{ia} G_{bk}] \xi_{abk1} - z \mathbb{E}_1[G_{i1}]. \tag{2.10}$$

For technical reasons, we define the cut-off version of ξ as follows,

$$\tilde{\xi}_{iklj} = \min\{\max\{\xi_{iklj}, -c_1^2|i-j|^{-d}\}, c_1^2|i-j|^{-d}\} \tag{2.11}$$

, so that $\tilde{\xi}_{iklj}$ has a power-law decay as i and j gets farther. Define a linear map $S : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{N \times N}$ by

$$(S(M))_{pq} := \frac{1}{N} \sum_{\alpha,\beta} \tilde{\xi}_{p\alpha\beta q} M_{\alpha\beta}. \tag{2.12}$$

Therefore, (2.10) is equivalent to

$$- \mathbb{E}_1 [[GS(G)]_{i1}] - z \mathbb{E}_1 [G_{i1}] = \delta_{i1} + O(N^{-1} \max_{k,l} |G_{kl}|^2). \tag{2.13}$$

Notice that the expectation operator \mathbb{E}_1 is equivalent to integrating over N weakly dependent Gaussian random variables, we may remove the expectation up to the cost of some small error terms, after which, we would get a self-consistent equation in the following form.

$$G(-S(G) - z) = I + \text{error}. \tag{2.14}$$

Define a map $F : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{N \times N}$ via

$$F(M) = (-z - S(M))^{-1}. \tag{2.15}$$

Then the above equation can be written as the perturbation of a fixed point equation

$$G = F(G) + \text{error}. \tag{2.16}$$

Here the error is entry-wise bounded by roughly $O((N\eta)^{-\frac{1}{2}})$. However, this entry-wise bound is not strong enough to use the stability of the equation $G = F(G)$. Therefore, we iterate the map F on G to get

$$F(F(G)) = F(F(F(G))) + \text{new error}. \tag{2.17}$$

The new error term has a power-law decay on the off-diagonal entries; this comes from two facts that will be established in detail later. The first is the fact that the $S(G)$ operator acts like a convolution operator with ξ_{ijkl} which allows us to get decay on the entries of $S(G)$. Secondly, we use the fact that the inverse of a matrix whose entries have decay will also have decay of such order. Hence, two applications of F will give us an error that is much smaller than the original error and we can get a good estimate on $F(F(G))$. Using $F(F(G))$ we can recover G and get a bound on $|G - G_0|$ where G_0 is some deterministic matrix.

2.3 Limiting version of self-consistent equation

In order to define a local law, we need a space in which we can define our limiting self-consistent equation. This limiting equation is best described as the space of convolution operators on the torus.

Consider $\mathcal{K} := C(\mathbb{T}^2)$ and $\mathcal{K}_+ := \{g \in \mathcal{K} \mid \text{Im } g(s, u) > 0, \forall s, u \in \mathbb{T}\}$. Recall the function ϕ in (2.1). Define

$$\varphi(s, t, u, v) := \sum_{k,l} \phi(s, t, k, l) e^{-2\pi i(uk-vl)}. \tag{2.18}$$

The argument in Lemma 4.15 of [6] can be modified to show that $\varphi \sim 1$. We give a brief summary of that argument as follows. One can write, by the definition of profile,

$$\int \int \varphi(\theta, \phi, s, t) g(\theta, \phi)^2 d\theta d\phi = \lim_{N \rightarrow \infty} \text{Var} Y_N \tag{2.19}$$

where $Y_N := \frac{1}{N} \sum_{i,j} h_{ij} g(i/N, j/N) e^{2\pi i(si-tj)}$. $\text{Var} Y_N$ can be controlled via the decay assumptions on ξ_{ijkl} . Choosing g carefully gives pointwise bounds on φ . This is the only place where the argument differs from Lemma 4.15 in [6] which used the finite range of correlations.

The decay condition (2.2) guarantees that φ is Lipschitz. Define $\Psi : \mathcal{K}_+ \rightarrow \mathcal{K}_+$ via $\Psi(h)(s, u) := \int \int_{\mathbb{T}^2} \varphi(s, t, u, v) h(t, v) dt dv$ and $\Phi : \mathcal{K}_+ \rightarrow \mathcal{K}_+$ via $\Phi(h) := (-\Psi(h) - z)^{-1}$. Consider the fixed point equation $g = \Phi(g)$, or equivalently,

$$g(-\Psi(g) - z) = 1. \tag{2.20}$$

This will be the infinite limit of our earlier matrix equation

$$G(-S(G) - z) = I. \tag{2.21}$$

With a proper embedding of the infinite solution g in a finite matrix space, the equation (2.20) can be understood as a small perturbation of (2.21)

Equations like (2.20) are studied in detail in [2]. Since the function φ is bounded above and below away from 0, the function Φ satisfies conditions A1-A3 and the block fully indecomposable² condition of Definition 2.9 of [2]. Also, since φ is Lipschitz, it satisfies (2.22) in that article. Therefore, their Theorem 2.6 says that the above equation has a unique solution $g \in \mathcal{K}_+$, and there is a universal constant $c_3 < +\infty$ such that

$$\sup_{z \in \mathbb{C}^+} \|g\|_\infty \leq c_3. \tag{2.22}$$

Let $m(z) := \int \int_{\mathbb{T}^2} g(s, u) ds du$. Then m is the Stieltjes transform of a compactly supported probability measure ν on \mathbb{R} , i.e.,

$$m(z) = \int_{\mathbb{R}} \frac{\nu(dx)}{x - z}, \quad \forall z \in \mathbb{C}^+. \tag{2.23}$$

Then Theorem 2.6 in [2] says that ν has a $\frac{1}{3}$ -Hölder continuous density $\rho \in C_c(\mathbb{R})$ such that it has square-root behavior at the left and right edges, i.e., let

$$E_L := \inf \text{supp } \nu, \quad E_R := \sup \text{supp } \nu. \tag{2.24}$$

²This is a condition relating to the positivity of the coefficients of ψ . This allows for some regularity of the solution and is easily checked for ψ

Then, there are $c_L, c_R > 0$ s.t.

$$\rho(E_L + t) = c_L \sqrt{t} + O(t), \quad \rho(E_R - t) = c_R \sqrt{t} + O(t), \quad \text{as } t \rightarrow 0_+. \quad (2.25)$$

For $h \in \mathcal{K}$, define the Fourier coefficients $\hat{h}(s, k) := \int_{\mathbb{T}} h(s, u) e^{-2\pi i k u} du$. On \mathcal{K} we may define a norm $\|\cdot\|_\beta$ for $\beta \geq 0$:

$$\|h\|_\beta := \sup_{s \in \mathbb{T}, k \in \mathbb{Z}} |\hat{h}(s, k)| (1 + |k|)^\beta. \quad (2.26)$$

In view of Theorem 3.2 of this work, it is easy to see that $\|g\|_\alpha \vee \|g^{-1}\|_\alpha \lesssim 1$ on any bounded subdomain of \mathbb{C}^+ . For any $N \in \mathbb{N}$, define a discretization operator $D^{(N)} : \mathcal{K} \rightarrow \mathbb{C}^{N \times N}$ by

$$D(h)_{ij} := \hat{h}(i/N, j - i). \quad (2.27)$$

We have the following lemma concerning the discretization $D(g)$:

Lemma 2.7. *Let $a, b \in \mathcal{K}$. Assume that b is Lipschitz in the first variable in the sense that $|b(s, u) - b(s', u)| \leq L|s - s'|, \forall s, s' \in \mathbb{T}, u \in \mathbb{T}$. Then, $\|D(a)D(b) - D(ab)\| \lesssim N^{-\frac{1}{2}}(L + \|b\|_\alpha)\|a\|_\alpha$, also, $\|D(a)D(b)^* - D(ab)^*\| \lesssim N^{-\frac{1}{2}}(L + \|b\|_\alpha)\|a\|_\alpha$ where we specifically need $\alpha > 2$*

Proof. By definition, $(D(a)D(b) - D(ab))_{ij} = \sum_k \hat{a}(i/N, k - i)(\hat{b}(k/N, j - k) - \hat{b}(i/N, j - k))$, therefore, using the decay of \hat{a} and the Lipschitz continuity of \hat{b} , we have

$$|(D(a)D(b) - D(ab))_{ij}| \leq \sum_k \frac{\|a\|_\alpha}{|k - i|^\alpha} \frac{L|k - i|}{N} \lesssim N^{-1}L\|a\|_\alpha. \quad (2.28)$$

On the other hand, $\|D(a)D(b) - D(ab)\|_\alpha \lesssim \|a\|_\alpha \|b\|_\alpha$, hence

$$|(D(a)D(b) - D(ab))_{ij}| \leq \|a\|_\alpha \|b\|_\alpha (1 + |i - j|)^{-\alpha}. \quad (2.29)$$

Therefore $\|D(a)D(b) - D(ab)\|_{l^\infty \rightarrow l^\infty} \lesssim (L + \|b\|_\alpha)\|a\|_\alpha \sum_k (N^{-1} \wedge |k|^{-2}) \lesssim N^{-\frac{1}{2}}(L + \|b\|_\alpha)\|a\|_\alpha$. Here we used the fact that $\alpha > 2$ to get decay. Similarly, the $l^1 \rightarrow l^1$ norm is bounded by the same quantity, hence the operator norm has the same bound by interpolation. The second estimate follows from a similar argument. \square

Fix z ; in the following, we will let $g(s, t)$ be the solution of (2.20) at this point z . We now define $Z(z) := \{g(s, t) | s, t \in \mathbb{T}\}$. Equation (2.20) gives us that Z is bounded away from 0 and $+\infty$. For $K > 0$ let $\mathcal{D}_K = \{z \in \mathbb{C}^+ | |z| \leq K\}$.

Corollary 2.8. *There is an $N(K) > 0$ such that for any $N > N(K)$ and $z \in \mathcal{D}_K$, the singular spectrum of $D(g)$ is in the $N^{-\frac{1}{6}}(\log N)^{-1}$ -neighborhood of $Z(z)$.*

Proof. Let $\theta \ll 1$ be some parameter to be chosen. Let $x \in \mathbb{R}_+$ s.t. $\text{dist}(x, Z(z)) \geq \theta$. Let $h := \frac{1}{|g|^2 - x^2}$. Then

$$\begin{aligned} \|(D(g)D(g)^* - x^2)D(h) - I\| &\leq \|D(g)D(g)^* - D(|g|^2)\| \|D(h)\| \\ &\quad + \|D(|g|^2 - x^2)D(h) - D(1)\|. \end{aligned}$$

According to Lemma 2.7, we have $\|(D(g)D(g)^* - x^2)D(h) - I\| \lesssim N^{-\frac{1}{2}}[\|h\|_2 + L]$, where L is the Lipschitz constant of h with respect to the first variable. By chain rule we know that $\|h''\|_\infty \lesssim \theta^{-3}$ and $L \lesssim \theta^{-2}$. Therefore, $\|h\|_2 \lesssim \theta^{-3}$ and hence $\|(D(g)D(g)^* - x^2)D(h) - I\| \lesssim N^{-\frac{1}{2}}\theta^{-3}$. Choose $\theta = N^{-\frac{1}{6}}(\log N)^{-1}$. Then $D(g)D(g)^* - x^2$ is invertible for N large enough. That means x is not in the singular spectrum of $D(g)$. \square

Corollary 2.9. *Let $R := D(g)(-S(D(g)) - z) - I$. Then, for any $z \in \mathcal{D}_K$,*

$$|R_{ij}| \leq C(K)N^{-1} \wedge |i - j|^{-2}. \tag{2.30}$$

In particular, $\|R\| \leq C(K)N^{-\frac{1}{2}}$.

Proof. According to Lemma 2.7 and equation (2.20), we know

$$|(D(g)(-D(\Psi(g)) - z) - I)_{ij}| \lesssim N^{-1} \wedge |i - j|^{-2}.$$

By definition, $(D(\Psi(g)))_{kl} = \sum_p \int_{\mathbb{T}} \phi(k/N, t, l - k, p) \hat{g}(t, p) dt$,

$(S(D(g)))_{kl} = \frac{1}{N} \sum_{p,q} \phi(k/N, q/N, l - k, p) \hat{g}(q/N, p)$. Using the Lipschitz-ness of ϕ and g , we have $|(D(\Psi(g)))_{kl} - (S(D(g)))_{kl}| \lesssim N^{-1} \wedge |k - l|^{-2}$. Therefore,

$$|(D(g)(-S(D(g)) - z) - I)_{ij}| \lesssim N^{-1} \wedge |i - j|^{-2},$$

as desired. □

Corollary 2.10. *Recall the definition (2.15) of F . For all sufficiently large N , there exists a constant $c > 0$ such that $\|F(D(g)) - D(g)\| \vee \|F(F(D(g))) - D(g)\| \leq cN^{-\frac{1}{2}}$. In particular, the singular spectrum of $F(D(g))$ and $F(F(D(g)))$ are contained in a compact subset of \mathbb{R}_+ .*

Proof. Using the notation from the previous corollary, if $D(g)(-S(D(g)) - z) - I = R$, then

$$F(D(g)) = (I + R)^{-1}D(g).$$

Since $\|R\| \lesssim N^{-\frac{1}{2}}$ and $\|D(g)\| \lesssim 1$, we know $\|(I + R)^{-1}D(g) - D(g)\| \lesssim N^{-\frac{1}{2}}$. From perturbation theory we know that the singular spectrum of $F(D(g))$ is within $N^{-\frac{1}{2}}$ of that of $D(g)$, therefore it is a compact subset of \mathbb{R}_+ . On the other hand, a simple algebraic calculation yields

$$F(F(D(g))) = (I + F(D(g))S(F(D(g))R))^{-1}F(D(g)).$$

Note that $\|F(D(g))S(F(D(g))R)\| \lesssim N^{-\frac{1}{2}}$, so the singular spectrum of $F(F(D(g)))$ is within the $O(N^{-\frac{1}{2}})$ neighborhood of that of $F(D(g))$, hence is a compact subset of \mathbb{R}_+ . □

For $z \in \mathbb{C}^+$, define

$$\kappa(z) := \text{dist}(z, \text{supp } \nu), \quad \rho(z) := \rho(\text{Re } z), \quad \omega(z) := \kappa(z)^{\frac{2}{3}} + \rho(z)^2. \tag{2.31}$$

Theorem 2.8 in [2] implies the following stability result:

Lemma 2.11. *There is a universal constant c_6 such that if $\tilde{g} \in \mathcal{K}$ satisfies*

$$\tilde{g}(-\Psi(\tilde{g}) - z) = 1 + r \tag{2.32}$$

and $\|\tilde{g} - g\|_\infty \leq c_6(\kappa^{\frac{2}{3}} + \rho)$, then $\|\tilde{g} - g\|_\infty \leq c_6^{-1}\omega^{-1}$.

2.4 Concentration lemmas

The following lemma says that a Lipschitz function of weakly dependent Gaussian random variables concentrates around its expectation.

Lemma 2.12. *Let $X = (X_1, \dots, X_n)$ be an array of centered Gaussian random variables with covariance matrix Σ . Let $f : \mathbb{R}^N \rightarrow \mathbb{R}$ be a Lipschitz function, such that $|f(x) - f(y)| \leq L|x - y|, \forall x, y \in \mathbb{R}^N$. Then*

$$\mathbb{P} [|f(X) - \mathbb{E}f(X)| \geq t] \leq 2e^{-\frac{t^2}{2L^2\|\Sigma\|}}, \quad \forall t > 0.$$

Proof. Let $Y = \Sigma^{-1/2}X$ so that Y is an n -dimensional random vector with independent $N(0, 1)$ components. In [19],

$$\mathbb{P} \left[\left| f(\Sigma^{\frac{1}{2}}Y) - \mathbb{E}f(\Sigma^{\frac{1}{2}}Y) \right| \geq t \right] \leq 2e^{-\frac{t^2}{2L_1^2}} \text{ for all } t > 0.$$

Here L_1 is the Lipschitz constant for the function $y \mapsto f(\Sigma^{\frac{1}{2}}y)$. It is easy to see that $L_1 \leq L\|\Sigma\|^{\frac{1}{2}}$, which concludes the proof. \square

In the future, we will frequently use the following lemma.

Lemma 2.13. *Let $A \in \mathbb{C}^{N \times N}$. Assume that there are $\beta > 0, \theta > 1$, s.t.*

$|A_{ij}| \leq \beta((|i - j| + 1)^{-\theta} + N^{-1}), \forall 1 \leq i, j \leq n$. Then $\|A\| \leq \frac{\beta\theta}{\theta - 1}$. More generally, for any $p \in [1, +\infty]$, we have $\|A\|_{l^p \rightarrow l^p} \leq \frac{\beta\theta}{\theta - 1}$.

Proof. Without loss of generality let $\beta = 1$. For any vector $v \in \mathbb{R}^n$,

$$\begin{aligned} \|Av\|_{\infty} &= \max_k \left| \sum_i A_{ki}v_i \right| \\ &\leq \|v\|_{\infty} \max_k \left(\sum_i (|i - k| + 1)^{-\theta} + N^{-1} \right) \\ &\leq \|v\|_{\infty} \left(\int_1^{+\infty} x^{-\theta} dx + 1 \right). \end{aligned}$$

Therefore, $\|A\|_{l^{\infty} \rightarrow l^{\infty}} \leq \frac{\beta\theta}{\theta - 1}$. Similarly, $\|A\|_{l^1 \rightarrow l^1} = \|A^*\|_{l^{\infty} \rightarrow l^{\infty}} \leq \frac{\beta\theta}{\theta - 1}$. By interpolation,

$$\|A\|_{l^p \rightarrow l^p} \leq \|A\|_{l^{\infty} \rightarrow l^{\infty}}^{\frac{1}{p}} \|A\|_{l^1 \rightarrow l^1}^{1 - \frac{1}{p}} \leq \frac{\beta\theta}{\theta - 1}, \quad \forall p \in [1, +\infty]. \quad \square$$

Recall that in Section 2.2 we defined a map S (see (2.12)). Thanks to the decay condition (2.2), the operator S is a bounded operator, as will be seen in the following lemma.

Lemma 2.14. *Let $A \in \mathbb{C}^{N \times N}$. Then there is a universal constant $c > 0$ such that the following inequalities hold.*

1. $\|S(A)\|_{d-1} \leq c\|A\|_{\infty}$.
2. $\|S(A)\|_{l^p \rightarrow l^p} \leq c\|A\|_{\infty}, \forall p \in [1, +\infty]$.
3. $\|S(A)\|_d \leq c\|A\|_{d-1}$.
4. $\|S(A)\|_{d-\frac{1}{2}} \leq c\|A\|$.

Proof. By definition $|S(A)_{ij}| = |\frac{1}{N} \sum_{k,l} \xi_{iklj} A_{kl}| \leq \frac{|A|_\infty}{N} \sum_{k,l} |\xi_{iklj}|$. According to (2.2), $\frac{1}{N} \sum_{k,l} |\xi_{iklj}| \lesssim \frac{1}{(1+|i-j|)^{d-1}}$. Hence $|S(A)_{ij}| \lesssim \frac{|A|_\infty}{(1+|i-j|)^{d-1}}$, which implies the first inequality. Setting $\theta = d - 1$ in Lemma 2.13, we see that $\|S(A)\|_{lp \rightarrow lp} \lesssim |A|_\infty, \forall p \in [1, +\infty]$, which implies the second inequality. If $\|A\|_{d-1} < +\infty$, then $|S(A)_{ij}| = |\frac{1}{N} \sum_{k,l} \xi_{iklj} A_{kl}| \leq (1 + |i - j|)^{-d} \frac{1}{N} \sum_{k,l} |A_{kl}| \lesssim (1 + |i - j|)^{-d}$. This proves the third inequality. As for the fourth inequality, we use Cauchy-Schwarz inequality to see that

$$|S(A)_{ij}| \leq \frac{1}{N} \left(\sum_{k,l} |\xi_{iklj}|^2 \right)^{\frac{1}{2}} \left(\sum_{k,l} |A_{kl}|^2 \right)^{\frac{1}{2}} \lesssim (1 + |i - j|)^{\frac{1}{2}-d} \|A\|. \quad \square$$

2.5 Error estimate

Recall the decomposition (2.9)

$$h_{ab} = \sum_{k=1}^N \gamma_{abk1} h_{k1} + \tilde{h}_{ab}, \quad \forall a, b \in \mathbb{Z}/N\mathbb{Z}. \quad (2.33)$$

Taking the co-variance with h_{l1} for any $l \in \mathbb{Z}/N\mathbb{Z}$, we see that

$$\xi_{abl1} = \sum_{k=1}^N \gamma_{abk1} \xi_{l1k1}, \quad \forall l \in \mathbb{Z}/N\mathbb{Z}.$$

Note that by assumption (2.2) the matrix $\Sigma_1 := (\xi_{l1k1})_{l,k \in \mathbb{Z}/N\mathbb{Z}}$ satisfies $|\xi_{l1k1}| \lesssim \frac{1}{(1+|l-k|)^\alpha}$ and by (2.4), $\|\Sigma_1^{-1}\| \leq c_2^{-1}$. Therefore Lemma 3.2 implies that $|(\Sigma_1^{-1})_{ij}| \lesssim (1 + |i - j|)^{-\alpha}$ and hence by Lemma 2.13 we have $\|\Sigma_1^{-1}\| \leq c$. Let ∇_1 denote the partial gradient with respect to the first column $(h_{k1})_{1 \leq k \leq N}$. Use the fact that $\frac{\partial G_{ij}}{\partial h_{ab}} = -G_{ia} G_{bj}$ and the chain rule, we have

$$\|\nabla_1 G_{ij}\|^2 \leq \sum_k \left| - \sum_{a,b} G_{ia} G_{bj} \gamma_{abk1} \right|^2 \lesssim \sum_k \left| - \sum_{a,b} G_{ia} G_{bj} \xi_{abk1} \right|^2.$$

In the second inequality above we have used the boundedness of $\|\Sigma_1^{-1}\|$. Let

$$\Gamma = \max_{i,j} |G_{ij}| \vee 1, \quad \gamma := \max_i \text{Im } G_{ii} \vee \eta. \quad (2.34)$$

Use the decay rate (2.2),

$$\|\nabla_1 G_{ij}\|^2 \leq C\Gamma^2 \sum_k \left(\sum_a \frac{|G_{ia}|^2}{(|a-k|+1)^{\alpha-1}} + \sum_b \frac{|G_{bj}|^2}{(|b-k|+1)^{\alpha-1}} \right)^2. \quad (2.35)$$

Since $\alpha - 1 > 1$, the operator norm of the matrix $\left(\frac{1}{(|a-k|+1)^{\alpha-1}} \right)_{1 \leq a,k \leq N}$ is bounded by $C(\alpha - 2)^{-1}$, according to Lemma 2.13. Therefore,

$$\|\nabla_1 G_{ij}\|^2 \leq C\Gamma^2 \left(\sum_a |G_{ia}|^2 + \sum_b |G_{bj}|^2 \right) \leq C\Gamma^2 \gamma \eta^{-1}. \quad (2.36)$$

In the second inequality we used Ward Identity. Similarly,

$$\|\nabla_1 (GS(G))_{ij}\| \leq \left\| \sum_p \nabla_1 G_{ip} (S(G))_{pj} \right\| + \left\| \sum_p G_{ip} \nabla_1 (S(G))_{pj} \right\|. \quad (2.37)$$

Define a short-hand notation $Q_{kl} := \|\nabla_1 G_{kl}\|$. By (2.36) we have $|Q|_\infty^2 \leq C\Gamma^2\gamma\eta^{-1}$. Then

$$\begin{aligned} \|\nabla_1(GS(G))_{ij}\| &\leq \sum_p Q_{ip}|(S(G))_{pj}| + \sum_p |G_{ip}| \frac{1}{N} \sum_{k,l} |\xi_{pklj}| Q_{kl} \\ &\leq |Q|_\infty \|S(G)\|_{l^\infty \rightarrow l^\infty} + |Q|_\infty \frac{\Gamma}{N} \sum_{k,l,p} |\xi_{pklj}|. \end{aligned} \tag{2.38}$$

Now we use the bound (2.36), and use the decay (2.2) as well as Lemma 2.14 to see,

$$\|\nabla_1(GS(G))_{ij}\|^2 \leq C\Gamma^4\gamma\eta^{-1}. \tag{2.39}$$

The observation above yields the following lemma.

Lemma 2.15. *Let $z = E + i\eta \in \mathbb{C}^+$ and $K \geq 1$, then there is a universal constant $c > 0$ such that*

$$-GS(G) - Gz = I + R,$$

where $\mathbb{P} \left[|R|_\infty \geq t\sqrt{\frac{K^4\gamma}{N\eta}}, \Gamma \leq K \right] \leq 2N^2e^{-ct^2}, \forall t \geq 1$.

Proof. For any $K > 0$ let $\chi : \mathbb{R} \rightarrow [0, 1]$ be a smooth function s.t. $|\chi'| \leq 1$ and $\chi = 1$ on $[-K, K]$ and $\chi = 0$ outside $[-3K, 3K]$. Define

$$\tilde{G} = \chi(\Gamma)G.$$

Then $\|\nabla \tilde{G}_{ij}\|^2 \lesssim K^2\gamma\eta^{-1}$. According to Lemma 2.12,

$$\mathbb{P} \left[|\tilde{G}_{ij} - \mathbb{E}_j \tilde{G}_{ij}| \geq t\sqrt{\frac{K^2\gamma}{N\eta}} \right] \leq 2e^{-ct^2}.$$

Note that $\tilde{G} = G$ on the event $\{\Gamma \leq K\}$. Therefore,

$$\mathbb{P} \left[\max_{i,j} |G_{ij} - \mathbb{E}_j G_{ij}| \geq t\sqrt{\frac{K^2\gamma}{N\eta}}, \Gamma \leq K \right] \leq 2N^2e^{-ct^2}.$$

On the other hand, in view of (2.39), a similar argument yields,

$$\mathbb{P} \left[\max_{i,j} |(GS(G))_{ij} - \mathbb{E}_j (GS(G))_{ij}| \geq t\sqrt{\frac{K^4\gamma}{N\eta}}, \Gamma \leq K \right] \leq 2N^2e^{-ct^2}.$$

Now we go back to the identity (2.13), removing \mathbb{E}_1 at the cost of some error term, and replacing 1 with a generic j , to see

$$-GS(G) - Gz = I + R,$$

where $\mathbb{P} \left[|R|_\infty \geq t\sqrt{\frac{K^4\gamma}{N\eta}}, \Gamma \leq K \right] \leq 2N^2e^{-ct^2}$. □

In particular, for a crude bound, we may take $t = \log N$ and take $K = 2/\eta$ so that $\mathbb{P}[\Gamma > K] = 0$. The lemma above yields,

Corollary 2.16. *Let R satisfy*

$$G(-S(G) - z) = I + R.$$

Then $|R|_\infty \leq \frac{8 \log N}{\sqrt{N\eta^6}}$ with probability $1 - N^{-c \log N}$.

3 The local law for correlated Gaussian ensembles

3.1 Power law decay of inverse matrices

Lemma 3.1. *Let $A, B \in \mathbb{C}^{N \times N}$, $\beta_{1,2} > 1$, then*

$$\|AB\|_{\min\{\beta_1, \beta_2\}} \leq C_{\min\beta_1, \beta_2} \|A\|_{\beta_1} \|B\|_{\beta_2}.$$

Proof. Note that by definition, $\|A\|_{\min\{\beta_1, \beta_2\}} \|B\|_{\min\{\beta_1, \beta_2\}} \leq \|A\|_{\beta_1} \|B\|_{\beta_2}$, so it is sufficient to prove the case where $\beta_1 = \beta_2 = \beta$. Without loss of generality assume $\|A\|_{\beta} = \|B\|_{\beta} = 1$, then,

$$|(AB)_{ik}| \leq \sum_j \frac{1}{(1 + |i - j|)^{\beta}} \frac{1}{(1 + |j - k|)^{\beta}}.$$

Since either $|i - j|$ or $|j - k|$ is $\geq |i - k|/2$, the above quantity is bounded by

$$|(AB)_{ik}| \leq 2 \sum_{l \in \mathbb{Z}} \frac{1}{(1 + \frac{|i-k|}{2})^{\beta}} \frac{1}{(1 + |l|)^{\beta}} \leq \frac{2}{(1 + \frac{|i-k|}{2})^{\beta}} \left(1 + 2 \int_1^{+\infty} \frac{dx}{x^{\beta}}\right),$$

which is bounded by $2^{\beta+1} \frac{\beta+1}{\beta-1} (1 + |i - k|)^{-\beta}$. \square

The following argument is based off a similar argument of Jaffard [15].

Theorem 3.2. *Let $d > \frac{3}{2}$ and assume that a matrix $A = I + B$ (finite or infinite) satisfies $\|B\| < 1$ and $\|A\|_d < +\infty$. Then, for any $\delta > 0$, there exists a polynomial dependent on d and $\delta \geq 0$ such that $\|A^{-1}\|_{d-1/2-\delta} \leq P_{d,\delta}(\|A\|_d, \frac{1}{1-\|B\|})$.*

If $d > 1$ and there exists an $\epsilon > 0$ such that $\|B\| \leq 1 - \epsilon$, then $\|A^{-1}\|_{d-\delta} \leq C(\delta, \epsilon, \|A\|_d)$.

We will show matrix element decay of the solution to the self-consistent equation. Though we will only really apply this to the solution of the limiting equation (2.21), the following theorem will phrase the result in terms of Matrices for convenience of notation.

Proposition 3.3. *Let M be the solution to the following equation*

$$M(-z - S(M)) = I.$$

where S is defined by (2.12) for a covariance structure ξ_{ijkl} with decay $d > 2$. If there exists a constant $c > 0$ such that $\|M\|, \|M^{-1}\| \leq c$, then we have that $\|M\|_{\alpha} \leq C(c, \alpha)$.

Proof. Notice that we are able to write

$$M = ((M^{-1})^* M^{-1})^{-1} (M^{-1})^* \quad (3.1)$$

By the equation of M , we have $M^{-1} = -z - S(M)$. Let us first estimate the decay of M^{-1} . By Lemma 2.14 we have $\|M^{-1}\|_{d-\frac{1}{2}} = \|-z - S(M)\|_{d-\frac{1}{2}} \lesssim \|M\|$. By Lemma 3.1 we have $\|M^{-1}(M^{-1})^*\|_{d-\frac{1}{2}} \lesssim \|M\|^2$. We would now like to apply theorem 3.2 to $(M^{-1}(M^{-1})^*)^{-1}$.

For any general positive semi-definite matrix, A , we will be able to write it as $A = \frac{\lambda_1 + \lambda_n}{2} [I + B]$ where λ_1 and λ_n are respectively the largest and smallest eigenvalues of A .

Theorem 3.2 is applied to the matrix $I + B$. The operator norm bound on B will be $\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$. The important factor $r = 1 - \|B\|$ will be $\frac{2\lambda_n}{\lambda_1 + \lambda_n}$. Theorem 3.2 now shows that the matrix decay of A^{-1} will be the same matrix decay of A .

Applying this logic to the positive semidefinite matrix $(-z - S(M))(-z - S(M))^*$, one will obtain that λ_1 and λ_n are both of some bounded constant order. Thus, we see we have matrix decay of order $d - \frac{1}{2}$.

Finally, we apply Lemma 3.1 to the equation (3.1). Notice that the two terms $(M^{-1})^*$ and $((M^{-1})M^{-1})^{-1}$ both have decay $d - \frac{1}{2}$. Thus, we see the product, and thus M , will have a decay of order $d - \frac{1}{2}$.

One we know M has decay $d - \frac{1}{2}$, we can show that $S(M)$ has even better decay; it will have decay of order α . We can repeat the same argument, but with this better decay estimate, to show that M has matrix decay of order α . \square

Now we define $J : \mathbb{C}^{N \times N} \rightarrow \mathcal{K}$, such that for any $A \in \mathbb{C}^{N \times N}$ and $i \in \mathbb{Z}/N\mathbb{Z}$, $u \in \mathbb{T}$,

$$J(A)(i/N, u) := \sum_{k=i-\lfloor N/2 \rfloor}^{i+\lfloor N/2 \rfloor} A_{i,i+k} e^{2\pi i k u}. \tag{3.2}$$

and $J(A)(s, u)$ is linear in s for $s \in [i/N, (i + 1)/N]$. It is easy to check that

$$D(J(A)) = A, \quad \forall A \in \mathbb{C}^{N \times N}. \tag{3.3}$$

Proposition 3.4. *Consider a fixed bounded subset $U \subset \mathbb{C}^+$. There are constants $\epsilon, C > 0$ such that if $|J(M) - g|_\infty \leq \epsilon$, then $\|F(M) - F(D(g))\|_{\alpha-1} \vee \|F(F(M)) - F(F(D(g)))\|_\alpha \leq C|J(M) - g|_\infty$ and $|F(M) - D(g)|_\infty \vee |F(F(M)) - D(g)|_\infty \leq C(|J(M) - g|_\infty + N^{-\frac{1}{2}})$, $\forall z \in U$.*

Proof. Let $A := F(D(g))$ and $R := S(M - D(g))$. Then

$$F(M) - A = \sum_{k=1}^{\infty} A(RA)^k.$$

Hence $\|F(M) - A\|_{\alpha-1} \leq \sum_{k=1}^{\infty} \|A(RA)^k\|_{\alpha-1}$. It is easy to see that $\|R\|_{\alpha-1} \leq c|J(M) - g|_\infty$ for some universal constant $c > 0$. By Lemma 3.1 we have $\|A(RA)^k\|_{\alpha-1} \lesssim (c|J(M) - g|_\infty)^k$. Therefore, taking ϵ small enough, we have $\|F(M) - F(D(g))\|_{\alpha-1} \leq C|J(M) - g|_\infty$.

Next, we define $R' = S(F(M) - F(D(g)))$, $A' = F(F(D(g)))$. Then $\|R'\|_\alpha \leq c'\epsilon$ according to the above argument. We have

$$F(F(M)) - A' = \sum_{k=1}^{\infty} A'(R'A')^k.$$

By Lemma 3.1 we have $\|A'(R'A')^k\|_\alpha \lesssim (c|J(M) - g|_\infty)^k$. Therefore, taking ϵ small enough, we have $\|F(F(M)) - F(F(D(g)))\|_\alpha \leq C|J(M) - g|_\infty$.

The last claim follows from the estimates above and Corollary 2.10. \square

3.2 Local law

Recall definition (2.31) and (2.34), for a constant $T > 0$ to be chosen, define

$$\mathcal{D} := \{z \in \mathbb{C}^+ \mid |z| \leq T, \text{Im } z \geq (\log N)^{10} N^{-1} \omega^{-4}\}. \tag{3.4}$$

Theorem 3.5 (Local law). *Define $\Lambda(z) := |D(g) - G|_\infty$. For N large enough, we have*

$$\sup_{z \in \mathcal{D}} \Lambda(z) \leq (\log N)^4 \left(\sqrt{\frac{\gamma}{N\eta}} \right) \omega^{-1},$$

with probability $1 - e^{-a_3(\log N)^2}$. If $\kappa > \rho$,

$$\sup_{z \in \mathcal{D}_N^\theta} \Lambda(z) \leq (\log N)^8 \left(\sqrt{\frac{\text{Im } m}{N\eta}} \omega^{-1} + (N\eta)^{-1} \omega^{-2} \right),$$

with probability $1 - e^{-a_3(\log N)^2}$.

Proof. Take $K := \log N$, and let $\{z_k\}$ be an N^{-4} -net of \mathcal{D} . Define

$$\Omega := \bigcup_{k=1}^{N^{10}} \{\Lambda(z) \in (K^4 \sqrt{\gamma} (N\eta)^{-\frac{1}{2}} \omega^{-1}, K^{-1} \omega)\}.$$

Then by Proposition 3.4, on Ω we have

$$F(F(G))(-S(F(F(G)) - z)) = I + \tilde{R},$$

where $\tilde{R} \lesssim |R|_\infty$. Then $J(F(F(G))(-\Psi(J(F(F(G)))) - z)) = 1 + O(|R|_\infty + N^{-1})$. By Proposition 3.4, on Ω we have $|F(F(G)) - D(g)|_\infty \lesssim K^{-1} \omega + N^{-\frac{1}{2}}$, which is $\ll \omega$.

We can then apply stability Lemma 2.11 to $J(F(F(G)))$ which approximately satisfies the self-consistent equation. This would then imply that $\|J(F(F(G))) - g\|_\infty \lesssim (|R|_\infty + N^{-1}) \omega^{-1}$. Discretizing this would give the inequality $\|F(F(G)) - D(g)\|_\infty \lesssim (|R|_\infty + N^{-1}) \omega^{-1}$. Due to the closeness of G and $F(G)$ and $F(G)$ to $F(F(G))$ from the self-consistent equation, this inequality would imply that $\lambda(z) = |G - D(g)|_\infty \lesssim (|R|_\infty + N^{-1}) \omega^{-1}$.

Therefore, on Ω we have $|R|_\infty \gtrsim K^4 \sqrt{\gamma} (N\eta)^{-\frac{1}{2}}$. By Lemma 2.15 we know $\mathbb{P}[\Omega] \leq 2N^{12} e^{-c(\log N)^2}$. On Ω^c , we either have $\inf_{z \in \mathcal{D}} |G - D(g)|_\infty \geq K^{-1} \omega / 2$ or $\sup_{z \in \mathcal{D}} |G - D(g)| \leq 2K^4 \sqrt{\gamma} (N\eta)^{-\frac{1}{2}}$. The latter is true with probability $1 - e^{-c(\log N)^2}$, since if we take the T in the definition of \mathcal{D} to be a large enough constant, then the former case holds with $O(e^{-c(\log N)^2})$ probability. \square

Corollary 3.6. *Let $a > 0$ be a small constant. Then on*

$$\mathcal{D}' := \{z \in \mathcal{D} | \kappa \geq N^{-a}\}.$$

we have

$$|\mathbb{E}[G] - D(g)|_\infty \lesssim (\log N)^{16} \left(\frac{1}{N\kappa\omega^3} + \frac{1}{(N\eta)^2\omega^5} \right).$$

Proof. By integration by parts,

$$-\mathbb{E}[GS(G)] - \mathbb{E}[G]z = I.$$

Let $R = (|G_{ij} - D(g)_{ij}|)_{1 \leq i, j \leq N}$.

$$-\mathbb{E}[G]S(\mathbb{E}[G]) - \mathbb{E}[G]z = I + \mathbb{E}[(R)S(R)] = I + \mathbb{E}[O(|D(g) - G|_\infty^2)].$$

Repeating the argument in the proof of Theorem 3.5 on $\mathbb{E}[G]$ instead of G , we have

$$|\mathbb{E}[G] - D(g)|_\infty \lesssim \mathbb{E}[O(|D(g) - G|_\infty^2)] \omega^{-1}.$$

We use Theorem 3.5 and the crude bound $\text{Im } m \leq \eta\kappa^{-2}$ to get the conclusion. \square

Remark 3.7. When we proved this local law, the only error estimates that depended strongly on the particular model we are considering are the stability results for the limiting vector equation. When considering the case of sample covariance matrices, though they are not exactly considered in the context of our proof, the stability results and the square root behavior at the right edge hold. Thus, we will be able to prove a local law for sample covariance matrices.

3.3 Upper bound of top eigenvalue

In the previous section, we have established optimal term-wise estimates on the entries of the Green's function. Estimates of the trace of the trace of the Green's function, however, are generally better due to Central Limit Theorem type cancellations.

One way to see this is to prove a Fluctuating Averaging Lemma [12], which would involve combinatorial expansions. In the Gaussian case, we can implicitly see the same effect by using a general result of the concentration of the largest eigenvalue along with our optimal term-wise estimates for the Green's function. The following lemma makes this intuition rigorous.

Lemma 3.8. *For $N \in \mathbb{N}$, consider a family of random measures $\mu_N = \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k}$ where $\lambda_1 \geq \dots \geq \lambda_N$ such that there is a deterministic $\hat{\lambda}_1$ satisfying $\lambda_1 = \hat{\lambda}_1 + o(N^{-\frac{1}{2}+\varepsilon})$ for any $\varepsilon > 0$. Assume that there exists a deterministic measure ν whose Green's function satisfies*

$$\text{Im}[m_\nu](x + i\eta) \leq C \frac{\eta}{\sqrt{\kappa + \eta}} \tag{3.5}$$

where $\kappa := \text{dist}(\text{supp}(\nu), x)$ and that

$$|\mathbb{E}[m_{\mu_N}(z)] - m_\nu(z)| = o(N^{-\frac{1}{2}-\gamma}). \tag{3.6}$$

for some $\gamma > 0$ and all $z = E + i\eta$ with $\text{dist}(E, \text{supp}(\nu)) \geq N^{-\epsilon}$ and $\eta \geq N^{-\delta-\frac{1}{2}}$ for some $\delta, \epsilon > 0$.

Then, $\text{dist}(\lambda_1, \text{supp}(\nu)) \leq N^{-\epsilon'}$ for some $\epsilon' > 0$.

Proof. Assume for contradiction that λ_1 lies outside a distance $N^{-\epsilon'}$ of $\text{supp}(\nu)$ for some ϵ' smaller than the ϵ such that condition (3.6) holds for $z = E + i\eta$ with $\text{dist}(E, \text{supp}(\nu)) \geq N^{-\epsilon}$. The exact value of ϵ' will be specified later. By our concentration result of λ_1 around $\hat{\lambda}_1$, this is equivalent to assuming that $\hat{\lambda}_1$ will be a distance $N^{-\epsilon'}$ away from $\text{supp}(\nu)$

We know that λ_1 will always be in a $N^{-1/2+\epsilon}$ neighborhood of $\hat{\lambda}_1$, we will be able to prove that the integral of the Green's function in a neighborhood around $\hat{\lambda}_1$ will always be bounded below by a constant times N^{-1} .

More specifically, we would have that

$$\frac{1}{N} \leq C \int_I \frac{1}{N} \frac{\eta}{(\lambda_1 - E)^2 + \eta^2} dE \leq C \int_I \text{Im}[m_{\mu_N}](E + i\eta) dE, \tag{3.7}$$

letting $I = [\hat{\lambda}_1 - \frac{N^{\gamma'}}{\sqrt{N}}, \hat{\lambda}_1 + \frac{N^{\gamma'}}{\sqrt{N}}]$ with $\gamma' < \gamma \wedge \delta/2$ and η to be $N^{-1/2-\delta}$. One should realize that with the above conditions, η will always be less than $N^{-1/2-\gamma'}$. The term in the middle of the above inequality is $\frac{1}{N} \text{Im}[\frac{1}{\lambda_1 - z}]$; noting that $\frac{1}{N} \text{Im}[\frac{1}{\lambda_i - z}]$ is positive in general and using the definition of the Green's function, we get the second inequality.

To see why the first inequality is true, one should first realize that a one sided η neighborhood of λ_1 will always lie in the interval I for sufficiently large N ; this is our concentration assumption $\lambda_1 = \hat{\lambda}_1 + o(N^{-\frac{1}{2}+\varepsilon})$ where we can choose ε less than γ' . Without loss of generality, we may assume that $\tilde{I} := [\lambda_1 - \eta, \lambda_1]$ is in I .

For $E \in \tilde{I}$, one would be able to bound the function $\frac{\eta}{(E-z)^2 + \eta^2}$ below by $\frac{1}{2\eta}$. The integral of this function over \tilde{I} would then clearly be bounded below by $\frac{1}{2}$. We see that we can set $C = 2$ for example.

We now take the expectation of (3.7) to get,

$$\frac{1}{N} \leq C \int_I \text{Im}[\mathbb{E}[m_{\mu_N}]](E + i\eta) dE \tag{3.8}$$

$$\leq C' \int_I \frac{o(1)}{N^{\frac{1}{2} + \gamma}} + C' \int_I \text{Im}[m_\nu](E + i\eta) dE \tag{3.9}$$

$$\leq C' \frac{o(1)}{N} + C'' \int_I \frac{\eta}{\sqrt{\text{dist}(E, \text{supp } \nu)}} \leq C' \frac{o(1)}{N} + C'' \frac{\eta N^{\gamma'}}{\sqrt{N} \sqrt{\hat{\lambda}_1}}. \tag{3.10}$$

In (3.8) we used the assumption (3.6) while in (3.10), we used the fact that ν satisfies (3.5). In the final line, we used the assumption that $\hat{\lambda}_1$ lies at a distance of $N^{-\epsilon'}$ from the support of ν , which is at a much greater scale than the length of I .

Notice that we have set $\eta = N^{-1/2-\delta}$ for δ positive and can now choose $N^{-\epsilon'} := N^{-\min(\epsilon, \delta/4)}$ and see that the error of (3.10) will be $\frac{o(1)}{N}$. This contradiction implies that for large N , $\hat{\lambda}_1$ must necessarily be of distance less than $N^{-\epsilon'}$ from the support of $\text{supp } \nu$. By concentration of λ_1 around $\hat{\lambda}_1$, we would know that all λ_1 will be less than $N^{-\epsilon'}$ from the support of ν . \square

Theorem 3.9. *For random matrix ensembles satisfying the conditions of Definition 2.1, there exists an $\epsilon > 0$ such that all eigenvalues lie within distance $N^{-\epsilon}$ from the edge of the ensemble.*

Proof. We would like to apply Lemma 3.8. First notice that by Gaussian concentration, we are able to prove that the distance of $|\lambda_1 - \mathbb{E}[\lambda_1]| \leq \frac{(\log N)^2}{\sqrt{N}}$ with probability $1 - O(N^{-c \log N})$. We thus put $\hat{\lambda}_1 = \mathbb{E}[\lambda_1]$ in the assumption of Lemma 3.8.

Then we check that the error bounds in Corollary 3.6 are sufficient for our purposes. The error that appears there is

$$|\mathbb{E}[G] - D(g)|_\infty \lesssim (\log N)^{16} \left(\frac{1}{N\kappa\omega^3} + \frac{1}{(N\eta)^2\omega^5} \right).$$

By the definition of D and the Lipschitz continuity of g , we have $|\mathbb{E}[\frac{1}{N} \text{Tr } G] - m_\nu| = O(N^{-\frac{1}{2}-\gamma})$ for some $\gamma > 0$ as long as we have $\eta \gg N^{-3/4+\delta}$ and $\kappa \sim N^{-\epsilon}$ for ϵ very small and $\delta > 0$. Since δ can be arbitrarily small, we may choose η such that $N^{-3/4+\delta} \ll \eta \ll N^{-1/2}$ and we can apply Lemma 3.8. \square

4 Universality

In the previous section, we proved a local law for m_N as well as an improved local law for $\mathbb{E}[m_N]$, and combining it with the concentration of the top eigenvalue to prove an upper bound on the top eigenvalue. According to a recent result by Landon and Yau [16] below, the local law with upper bound on the top eigenvalue is sufficient to prove universality near the edge.

Theorem 4.1. *Let $\eta^* = N^{-\phi^*}$ for some $0 < \phi^* < \frac{2}{3}$. We call a deterministic matrix V η^* -regular if it satisfies the following properties.*

1. *There exists a constant $C_V \geq 0$ such that for $-1 \leq E \leq 0, \eta^* \leq \eta \leq 10$,*

$$\frac{1}{C_V} \frac{\eta}{\sqrt{|E| + \eta}} \leq \text{Im}[m_V(E + i\eta)] \leq C_V \frac{\eta}{\sqrt{|E| + \eta}}$$

and for $0 \leq E \leq 1, (\eta^*)^{1/2}|E| + \eta^* \leq \eta \leq 10,$

$$\frac{1}{C_V} \sqrt{|E| + \eta} \leq \text{Im}[m_V(E + i\eta)] \leq C_V \sqrt{|E| + \eta}.$$

2. There exists no eigenvalue of V in the region $[-\eta^*, 0]$.
3. We have $\|V\| \leq N^{C_V}$ for some $C_V > 0$.

Consider the ensemble $V_t = V + \sqrt{t}G$. Where G is an independent GOE ensemble. Let t satisfy $N^{-\epsilon} \geq t \geq N^\epsilon \eta^*$ and let $F : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$ be a test function such that $\|F\|_\infty \leq C$ and $\|F\| \leq C$. Then there are deterministic parameters $\gamma_0 \sim 1$ and E_- such that

$$|\mathbb{E}[F(\gamma_0 N^{2/3}(\lambda_{i_0} - E_-), \dots, \gamma_0 N^{2/3}(\lambda_{i_k} - E_-))] - \mathbb{E}_{GOE}[F(N^{2/3}(\hat{\lambda}_1 + 2), \dots, N^{2/3}(\hat{\lambda}_k + 2))]| \leq N^{-c}.$$

The first expectation is with respect to the eigenvalues of the ensemble V_t with $\lambda_1 < \lambda_2 < \dots < \lambda_N$. The latter expectation is taken with respect to the eigenvalues $\hat{\lambda}_i$ of a GOE which are ordered $\hat{\lambda}_1 < \hat{\lambda}_2 < \dots < \hat{\lambda}_N$. i_0 is the first index i such that i th smallest eigenvalue of V is greater than $-\frac{1}{2}$.

Call H the ensemble with correlation structure ξ_{ijkl} . Theorem 3.9 combined with 3.5 shows that there exists a parameter $\Phi > 0$ such that with high probability a matrix M produced by H would be η^* regular for any $N^{-\phi}$ such that $\phi < \Phi$. We choose some $\phi < \Phi$ sufficiently small and set $t = N^{-\phi}$. We will use this t whenever referenced in the following sections. It will be important to choose ϕ sufficiently small in the coming sections.

In order to apply the theorem, we would like to write our matrix ensemble in the form $H = H' + \sqrt{t}G$, where G is a standard GOE matrix and H' is a matrix ensemble independent from G .

Recall the notation ξ_{ijkl} from equation (2.1). We can let $H' = (h'_{ij})$ be the auxiliary Gaussian ensemble whose correlation structure is given by

$$\xi'_{ijkl} = \xi_{ijkl} - t\delta_{ij=kl} \tag{4.1}$$

We see that when N is large enough, H' has positive semidefinite correlation matrix and so we can construct the ensemble H' . Since H' is a correlated Gaussian ensemble, we have a local law as in Theorem 3.5 well as bounds for the extremal eigenvalues as in Theorem 3.9.

We will apply (4.1) as follows. H from our original Gaussian ensemble can be written as $H' + \sqrt{t}G$ where H' is produced from our auxiliary ensemble and G is a GOE matrix independent from H' . Let U be a unitary matrix such that $V := U^*H'U$ is a diagonal matrix. Notice that U^*HU has the same eigenvalues as H and can be written as $U^*HU = V + \sqrt{t}\hat{G}$ where \hat{G} is a GOE matrix. This is possible as the GOE is invariant under unitary transformations.

We can condition on the matrix H' and apply the Theorem 4.1. The ensemble $H' + \sqrt{t}G$ with fixed H' and G a GOE has eigenvalue density near the left edge described by

$$\rho_{H'}^t(E) = \sqrt{E - E'_-(t)}(\gamma'(t))^{-1/2}(1 + t^2\mathcal{O}(|E - E'_-(t)|)) \tag{4.2}$$

for $ct^2 \geq E \geq E'_-(t)$. $E'_-(t)$ will be called the edge of this ensemble and $\gamma'(t)$ will be called the scaling factor. This is the content of lemma 2.3 of [16]. The universality result coming from applying Theorem 4.1 is:

$$\begin{aligned} & |\mathbb{E}_{H'}[F(\gamma_0 N^{2/3}(\lambda_1 - E_-), \dots, \gamma_0 N^{2/3}(\lambda_k - E_-))] \\ & - \mathbb{E}_{GOE}[F(N^{2/3}(\hat{\lambda}_1 + 2), \dots, N^{2/3}(\hat{\lambda}_k + 2))] \leq N^{-c}. \end{aligned} \tag{4.3}$$

where $\lambda_1 < \lambda_2 < \dots < \lambda_N$ are the eigenvalues of the considered matrix $H' + \sqrt{t}G$ with H' fixed and $\hat{\lambda}_1 < \hat{\lambda}_2 < \dots < \hat{\lambda}_N$ are the eigenvalues of a GOE matrix.

We used for N large enough, the smallest eigenvalue of H' is of distance less than $1/2$ from the edge, so the index i_0 is 1. The only issue with (4.3) is that γ_0 is a function of the initial data, we will make this a universal constant in the next section.

4.1 Changing the scaling factor

A priori, the scaling factor γ appearing in (4.3) is only known to be a function of the initial matrix H' used as an input to Theorem 4.1. However, we have explicit complex analytic equations determining the scaling factors depending on the initial data. By using Rouché's Theorem, the local law, and the Lipschitz continuity of the Green function, we can show that these equations are stable to small perturbations of the initial data. This allows us to show with high probability that the scaling factor will not change too much for two different initial data points and we can choose a common scaling factor γ for the entire ensemble.

Theorem 4.2. *Recall the auxiliary model H' with correlation structure as given in (4.1) which has regularity $\eta^* = N^{-\Phi}$. Let $t = N^{-\phi}$ with $0 \leq \phi \leq \Phi$ being chosen sufficiently small.*

Consider two distinct matrices H'_1 and H'_2 coming from the Gaussian ensemble with correlation structure ξ'_{ijkl} . Write the measure of the ensemble $H'_i + \sqrt{s}G$ as $\rho_i(E)$ with edges $E_-^i(s)$, scaling factors $\gamma^i(s)$ and Stieltjes transforms m_s^i . We have the relation

$$\gamma^1(t) - \gamma^2(t) = O(t) \tag{4.4}$$

The above theorem will show that any two matrices produced from H' will asymptotically have the same scaling factor.

Proof. Define z_i as follows where κ is a parameter dependent on t to be specified later.

$$z_i - tm_0^i(z_i) = E_-^i(t) + \kappa \tag{4.5}$$

The importance of the point z_i is contained in the following relation.

$$m_0^i(z_i) = m_0^i(E_-^i(t) + \kappa) \tag{4.6}$$

This is a standard property of the free convolution, one can refer to equation to equations (7.2) and (7.3) of [16] for a proof.

We can determine scaling factors using the following relation.

$$\pi[\rho'_1(E_-^1(t) + \kappa) - \rho'_2(E_-^2(t) + \kappa)] = \text{Im}[m_0^1[z_1]] - \text{Im}[m_0^2[z_2]] \tag{4.7}$$

$$= \text{Im}[m_0^1[z_1]] - \text{Im}[m_0^2[z_1]] + \text{Im}[m_0^2[z_1]] - \text{Im}[m_0^2[z_2]]. \tag{4.8}$$

In (4.8), the first term can be bounded by a sufficiently good local law. The second term can be bounded by a Lipschitz condition provided $|z_1 - z_2|$ are sufficiently close to each other.

We will now attempt to bound the quantity $|z_1 - z_2|$.

Lemma 4.3. *Recall all the definitions from the previous Theorem 4.2. Choose κ in (4.5) to be $t^{2+\epsilon}$ for $\epsilon > 0$ very small. Then there exists a small parameter $\delta > \epsilon$ such that for N large enough we have*

$$|z_1 - z_2| \leq t^{2+\delta} \tag{4.9}$$

Proof. We have that

$$z_1 - tm_t^1(z_1) - (z_2 - tm_t^2(z_2)) = (E_-^1(t) - E_-^2(t)). \tag{4.10}$$

$$(z_1 - z_2) - t(m_t^1(z_1) - m_t^1(z_2)) = (E_-^1(t) - E_-^2(t)) + t(m_t^1(z_2) - m_t^2(z_2)). \tag{4.11}$$

We will try to prove that $|z_1 - z_2|$ is sufficiently small. We will do this by appealing to Rouché’s Theorem and a Local Law bound to the second term on the RHS of (4.11).

We will now address the Local Law portion of the above estimate. We have

$$\text{Im}[z_1] = t\text{Im}[m_t^1(E_-^1(t) + \kappa)] \tag{4.12}$$

by taking imaginary parts of (4.5) and (4.6).

Using the fact that the imaginary part of the Stieltjes transform along the real axis will be the density, we have that for $\kappa \leq ct^2$

$$\text{Im}[m_t^1(E_-^1(t) + \kappa)] = O(\sqrt{\kappa}) \tag{4.13}$$

We use the expansion (4.2) and use the fact that $\gamma_t \asymp 1$. The latter fact is again an explicit calculation of the free convolution as can be seen in Lemma 2.3 of [16].

The last two equations show that $\text{Im}[z_1] \asymp t\sqrt{\kappa}$. If we take $\kappa \asymp t^{2+\epsilon}$, we get $\text{Im}[z_1] \asymp t^{2+\epsilon/2}$. Recalling that we set t of the order $N^{-\phi}$ for ϕ very small, we see that under our choice of κ we have $\text{Im}[z_1] \asymp N^{-2\phi(1+\epsilon/4)}$ and we should be in a region we can apply the local law 3.5.

To confirm this carefully, note that $\text{dist}(z_1, \text{supp } \nu) \geq \text{Im}[z_1]$ so the following should hold

$$\text{Im}[z_1] = N^{-2\phi(1+\epsilon/4)} \geq (\log N)^{10} N^{-1} (N^{2/3(-2\phi)(1+\epsilon)})^{-4} \gg (\log N)^{\log \log N} N^{-1/2}.$$

The point z_1 is in the region \mathcal{D} when when we have that ϕ is sufficiently small. Clearly, we would also have that a circle of radius $t^{2+\delta}$ around z_1 for $\delta > \epsilon/2$ would also lie in the region \mathcal{D} . This choice of δ will be important later when we apply Rouché’s theorem in a circle around z_1 . Notice that we can apply the same logic to z_2

Applying Theorem 3.5, we get

$$|m_t^1(z_2) - m_t^2(z_2)| \leq 2(\log N)^4 (\sqrt{1/(N\text{Im}(z))}) (\text{Im}(z))^{-2/3}$$

with exponentially high probability. The fact that we have a factor of \sqrt{N} in the denominator in the right hand side of the last line shows that

$$|m_t^1(z_2) - m_t^2(z_2)| \ll t^3 \tag{4.14}$$

provided ϕ is sufficiently small.

Recall that the auxiliary ensemble that produced H' is η^* regular for $\eta^* = N^{-\Phi}$. Using the bound on extreme eigenvalues from Theorem 3.9 we know that there is a deterministic E_- such that with high probability both

$$|E_-^i(t) - E_-| \leq N^{-\Phi}$$

We may set $\phi \leq \frac{\Phi}{3}$ to get the bound

$$|E_-^1(t) - E_-^2(t)| \leq t^3. \tag{4.15}$$

Using the local law on m_t^1 and the Lipschitz continuity of the limiting distribution, we know that there is a constant K such that for all pairs (a, b) in a circle of radius $O(t^{2+\delta})$ around the point z_1 we have

$$|m_t^1(a) - m_t^1(b)| \leq K|a - b| \tag{4.16}$$

for a finite constant K . Choosing $\delta > \frac{\epsilon}{2}$ ensures that since $\text{Im}[z_1] \asymp t^{2+\frac{\epsilon}{2}}$ this circle is indeed in a region where we can apply the local law.

On this circle of radius R we see that the left hand side of (4.11) has lower bound

$$|z_1 - z_2| - t|m_t^1(z_1) - m_t^1(z_2)| \geq |z_1 - z_2|(1 - Kt) \geq O(t^{2+\delta}) \tag{4.17}$$

where Kt will decay to 0 for N large enough. Here we used the Lipschitz continuity of (4.16) to bound the terms involving the Stieltjes transforms.

Combining estimates (4.14) and (4.15) we can get an upper bound on the left hand side of (4.11) as

$$|E_-^1(t) - E_-^2(t)| + t|m_t^1(z_2) - m_t^2(z_2)| \leq t^3 \tag{4.18}$$

Thus, on this circle of radius R , we see that the right hand side of (4.11) will be less than the left hand side of (4.11) in absolute value on the boundary.

If the right hand side of (4.11) were 0, then we would clearly have the unique solution $z_2 = z_1$. Rouché's theorem then shows that there is a solution such that $|z_2 - z_1| \leq R = t^{2+\delta}$. \square

Putting the estimate (4.9) back into (4.8) with $\kappa = t^{2+\epsilon}$.

$$\begin{aligned} & (\gamma^1(t))^{-1/2}t^{1+\epsilon/2}(1 + t^{-2}O(t^{2+\epsilon})) - (\gamma^2(t))^{-1/2}t^{1+\epsilon}(1 + t^{-2}O(t^{2+\epsilon})) \\ & \leq \pi^{-1}[\text{Im}[m_0^1[z_1]] - \text{Im}[m_0^2[z_1]] + \text{Im}[m_0^2[z_1]] - \text{Im}[m_0^2[z_2]]] \leq t^3 + Kt^{2+\delta}. \end{aligned} \tag{4.19}$$

For the first term in the second line, we used the local law around z_1 to bound the quantity by t^3 for the second quantity we used Lipschitz continuity of m_0^2 combined with the estimate on $|z_1 - z_2|$ coming from (4.3).

Notice that if we now have that

$$|\gamma^1(t)^{-1/2} - \gamma^2(t)^{-1/2}| \geq t \tag{4.20}$$

then it would clearly be impossible for the inequality in (4.19) to hold.

We can turn this into an $o(1)$ bound on $|\gamma^1(t) - \gamma^2(t)|$ using the facts that $\gamma^1(t)$ are $O(1)$ quantities. \square

One should note that this argument will also work to show that $\gamma^1(t)$ is of $O(t)$ distance from the scaling factor corresponding to the empirical spectral distribution of the ensemble H .

4.2 Final universality result

Using the scaling results coming from the previous section we can translate (4.3) as follows.

Theorem 4.4. *Let H be the Gaussian ensemble with correlation structure ξ_{ijkl} satisfying the assumptions of Section 2.1. There exists a deterministic scaling factor γ depending on the ensemble H such that the following inequality holds for functions $\mathcal{F} : \mathbb{R}^k \rightarrow \mathbb{R}$ such that $\|\mathcal{F}\|_\infty, \|\nabla\mathcal{F}\|_\infty \leq C$,*

$$\begin{aligned} & |\mathbb{E}_H[\mathcal{F}(\gamma N^{2/3}(\lambda_2 - \lambda_1), \dots, \gamma N^{2/3}(\lambda_k - \lambda_1))] \\ & - \mathbb{E}_{GOE}[\mathcal{F}(N^{2/3}(\hat{\lambda}_2 - \hat{\lambda}_1), \dots, N^{2/3}(\hat{\lambda}_k - \hat{\lambda}_1))] \leq N^{-c}. \end{aligned} \tag{4.21}$$

where $\lambda_1 < \lambda_2$

Proof. First, notice that we can find a function $F : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$ such that $\|F\|_\infty$ and $\|\nabla F\|_\infty$ are bounded and

$$F(x_1, \dots, x_{k+1}) = G(x_1 - x_2, \dots, x_1 - x_{k+1}).$$

Recall from earlier discussion that we can write any matrix from the ensemble H as $H' + \sqrt{t}G$ where H' is Gaussian with correlation structure with correlation structure $\xi_{abcd} - t\delta_{ab=cd}$ and G is an independent GOE matrix.

Let Ω be the set in which we know that H' has sufficiently good regularity so that (4.3) holds for the function F . On Ω , we would like to change the scaling factor γ_0 to γ , which is the scaling factor at the edge for the spectral density corresponding to H .

From Theorem 4.2, we know that the difference between the γ_0 appearing in (4.3) and the γ appearing here is of the order $t = N^{-\phi}$. Finally, one can use F is Lipschitz as well as the fact that the $N^{2/3}(\lambda_{i_k} - E_-)$ are bounded to say that

$$\begin{aligned} & |F(\gamma N^{2/3}(\lambda_1 - E_-), \dots, \gamma N^{2/3}(\lambda_k - E_-)) - \\ & F(\gamma_0 N^{2/3}(\lambda_1 - E_-), \dots, \gamma_0 N^{2/3}(\lambda_k - E_-))| \leq CkN^{-\phi/2}. \end{aligned}$$

We can take expectation of the above quantity in the ensemble $H' + \sqrt{t}G$ with G an independent GOE and apply the triangle inequality with (4.3) to prove

$$\begin{aligned} & |\mathbb{E}_{H'}[F(\gamma N^{2/3}(\lambda_1 - E_-^M), \dots, \gamma N^{2/3}(\lambda_k - E_-^M))] \\ & - \mathbb{E}_{GOE}[F(N^{2/3}(\hat{\lambda}_1 + 2), \dots, N^{2/3}(\hat{\lambda}_k + 2))] \leq N^{-c}. \end{aligned}$$

Translating this statement to G , we get on the set Ω we have

$$\begin{aligned} & |\mathbb{E}_{H'}[G(\gamma N^{2/3}(\lambda_1 - \lambda_2), \dots, \gamma N^{2/3}(\lambda_1 - \lambda_k))] \\ & - \mathbb{E}_{GOE}[G(N^{2/3}(\hat{\lambda}_1 - \hat{\lambda}_2), \dots, N^{2/3}(\hat{\lambda}_1 - \hat{\lambda}_k))] \leq N^{-c}. \end{aligned} \tag{4.22}$$

One would now like to remove the conditional expectation in the above expression. Namely, we would like to integrate (4.22) in Ω while using the trivial bound that $|\mathbb{E}_{H'}[G] - \mathbb{E}_{GOE}[G]|$ is bounded by a constant on the complement of Ω . We thus get the full universality statement

$$\begin{aligned} & |\mathbb{E}_H[G(\gamma N^{2/3}(\lambda_1 - \lambda_2), \dots, \gamma N^{2/3}(\lambda_1 - \lambda_k))] \\ & - \mathbb{E}_{GOE}[G(N^{2/3}(\hat{\lambda}_1 - \hat{\lambda}_2), \dots, N^{2/3}(\hat{\lambda}_1 - \hat{\lambda}_k))] \leq N^{-c}. \end{aligned} \tag{4.23}$$

as desired.

Remark 4.5. As long as we know that a version of the Dyson-Brownian Motion result holds for sample covariance matrices, then we will be able prove edge universality using the local law and edge upper bound for the top eigenvalue results from the previous section. \square

A Decay of inverse matrices

In this section we prove Theorem 3.2. Let $B = I - A$. Since $\|B\| < 1$, We can expand $A^{-1} = \sum_{k=0}^{\infty} B^k$. We need the following lemma to bound each term. For simplicity, we will prove the statement of polynomial decay of inverse of order 1 for matrix decay of order $2 + \delta$. The following proof can readily be generalized to show decay of inverse of order $d - 1 - \delta$, $\delta > 0$, given matrix decay of order d for $d > 2$.

Lemma A.1. *We have that*

$$\|B^n\|_{\alpha} \leq En^x \left(\frac{1 + \|B\|}{2} \right)^n. \quad (\text{A.1})$$

where E is a function that, upon fixing δ is only polynomially dependent on $\|B\|_{2+\delta}$ and $1 - \|B\|$ while x is dependent only on δ .

Proof. We want to compute the entries of $[B^n]_{jk}$. We will now define two auxiliary matrices $[\tilde{B}]_{xy} = B_{xy}\chi[|x - y| \leq \frac{j-k}{n}]$ and $[\hat{B}]_{xy} = \frac{j-k}{n} B_{xy}\chi[|x - y| \geq \frac{j-k}{n}]$.

Notice that we have the following identity

$$|j - k|[B^n]_{jk} = n \sum_{i=0}^{n-1} (\tilde{B})^i \hat{B} B^{n-i-1}. \quad (\text{A.2})$$

We now use the following interpolation identity which appears in [15].

Lemma A.2. *If $\|M\|_{l^2} \leq \infty$ and $\|N\|_{l^2} \leq \infty$, then we have that*

$$|(M\hat{B}N)_{xy}| \leq \|M\|_{l^2} \|B\|_{2+\delta} \|N\|_{l^2}. \quad (\text{A.3})$$

Proof. Notice that the decay of \hat{B} is order $1 + \delta$ with coefficient $\|B\|_{2+\delta}$. Thus we can say that \hat{B} exists in l^q for $q \geq \frac{1}{1+\delta}$. More specifically we have

$$\|\hat{B}\| \leq E \|B\|_{2+\delta} \quad (\text{A.4})$$

where E is a constant that depends on δ .

Also see that $|(M\tilde{B}N)_{xy}| = |\langle M e_x, \tilde{B} N e_y \rangle|$ where e_x is the canonical basis of our matrix space. By Young's inequality, we can say that

$$\|\tilde{B} N e_y\|_{l^2} \leq E \|B\|_{2+\delta} \|N e_y\|_{l^2} \leq E \|B\|_{2+\delta} \|N\|_{l^2}. \quad (\text{A.5})$$

The above equation is the result of Young's convolution inequality

$$\|f * g\|_r \leq \|f\|_p \|g\|_q$$

with $1 + r^{-1} = p^{-1} + q^{-1}$. Here we use $r = p = \frac{1}{2}$ and $q = 1$ with the q norm begin taken on the \hat{B} term and the p norm taken on the $N e_y$ term. We finally apply the Cauchy-Schwarz inequality to $|\langle M e_x, \tilde{B} N e_y \rangle| \leq E \|M\|_{l^2} \|B\|_{2+\delta} \|N\|_{l^2}$. \square

Applying the above lemma to each term of the form $\tilde{B}^i \hat{B} B^{n-i-1}$, we will be able to say that $[\tilde{B}^i \hat{B} B^{n-i-1}]_{ij} \leq \|\tilde{B}\|^i \|B\|_{2+\delta} \|B\|^{n-i-1}$. Finally, we would like to relate $\|\tilde{B}\|$ back to $\|B\|$. By triangle inequality, this would amount to estimating $\frac{n}{|j-i|} \|\hat{B}\|$. Notice that we have that $\|\hat{B}\| \leq E \|B\|_{2+\delta}$.

Thus, to get that $\|\hat{B}\|$ is sufficiently close to $\|B\|$, we would need to assume a few conditions on $|i - j|$. If we assume that $|j - i| > n \frac{2E\|B\|_{2+\delta}}{1-\|B\|}$, then we would know that $\|\tilde{B}\| \leq \|B\| + \frac{1-\|B\|}{2} = \frac{1+\|B\|}{2}$.

Assuming this condition on $|j - i|$, we find that

$$[\tilde{B}^i \hat{B} B^{n-i-1}]_{ij} \leq \left(\frac{1 + \|B\|}{2}\right)^{n-1} \|B\|_{2+\delta}.$$

Thus, we find that in (A.2) we have a bound of $n \left(\frac{1+\|B\|}{2}\right)^{n-1} \|B\|_{2+\delta}$. In the case that $|j - i|$ is less than $n \frac{2E\|B\|_{2+\delta}}{1-\|B\|}$, we find that we have $|j - i| [B^n]_{ij} \leq n \frac{2E\|B\|_{2+\delta}}{1-\|B\|}$. A trivial bound for $|i - j| [B^n]_{ij}$ would be a sum of the two quantities that we have derived above. \square

With the lemma in hand, we are able to say that

$$\|A\|_1 \leq \sum_{n=1}^{\infty} \|B^n\|_1 \leq E \frac{2^{x+1}}{(1 - \|B\|)^{x+1}}. \tag{A.6}$$

and we are done.

Remark A.3. If we want to show decay of inverse of order $d > \alpha > d - \frac{1}{2}$ with coefficient of decay dependent only polynomially on $\|A\|_d$ and $\|I - B\|$, then we would need a better interpolation result as appears in [15].

The main issue is that we are no longer able to estimate quantities like $\langle Me_i | \tilde{B} N e_j \rangle$ in (A.2) using the l_2 norms of M and N and instead one must use the l_p norms of M and N for p between 1 and 2.

One must then interpolate the l_p norm of M and N with the l_2 norm and the appropriate α norm like

$$\|B\|_{l^p} \leq c_p \|B\|_1^{\frac{2}{p}-1} \|B\|_{l^2}^{2-\frac{2}{p}}. \tag{A.7}$$

The bounding of $|j - k|^\alpha [B^n]_{jk}$ then becomes a recurrence relation.

$$\begin{aligned} \|B\|_\alpha &\leq C \|B\|_\alpha [\|B^{n-1}\|_\alpha^{\frac{2}{p}-1} \|B\|^{(n-1)(2-\frac{2}{p})} \\ &+ \sum_{i=1}^{n-1} (\|B^i\|_\alpha \|B^{n-i-1}\|_\alpha)^{2-\frac{2}{p}} \|B\|^{(n-1)(2-\frac{2}{p})}]. \end{aligned} \tag{A.8}$$

If one would want to prove inductively the bound that $\|B_n\|_\alpha \leq n^k R^n$, then placing this estimate inside the double product $\|B^i\|_\alpha \|B^{n-i-1}\|_\alpha$ and applying the trivial bound that $i^k (n - i - 1)^k \leq n^{2k}$ we would want $n^{2k(2-\frac{2}{p})} \leq n^k$. One notices now that this is only possible if we have that $2 - \frac{2}{p} \leq \frac{1}{2}$ or $p \leq \frac{4}{3}$. We could only choose $p < \frac{4}{3}$ if we choose $\alpha < d - \frac{1}{2}$.

If one has the comfort that $\|I - A\|$ is bounded away from 0, then one can analyze the recursion at any order $\alpha < d$ but the growth of the alpha norm in the recursion will no longer be $\|I - A\|$ but some parameter $r > \|I - A\|$.

References

- [1] Ben Adlam and Ziliang Che, *Spectral statistics of sparse random graphs with a general degree distribution*, preprint, arXiv:1509.03368 (2015).

- [2] Oskari Ajanki, László Erdős, and Torben Kruger, *Quadratic vector equations on complex upper half-plane*, preprint, arXiv:1506.05095v4 (2015). MR-3684307
- [3] Oskari Ajanki, Laszlo Erdos, and Torben Kruger, *Universality for general wigner-type matrices*, preprint, arXiv:1506.05098v2 (2015). MR-3719056
- [4] Oskari Ajanki, László Erdős, and Torben Kruger, *Local spectral statistics of gaussian matrices with correlated entries*, Journal of Statistical Physics (2016), 1–23. MR-3478311
- [5] Oskari Ajanki, Laszlo Erdos, and Torben Kruger, *Stability of the matrix dyson equation and random matrices with correlations*, preprint, arXiv:1604.08188v4 (2016). MR-3916109
- [6] Ziliang Che, *Universality of random matrices with correlated entries*, preprint, arXiv:1604.05709 (2016). MR-3629874
- [7] László Erdős, Antti Knowles, Horng Tzer Yau, and Jun Yin, *Spectral Statistics of Erdős-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues*, Communications in Mathematical Physics **314** (2012), no. 3, 587–640. MR-2964770
- [8] László Erdős, Antti Knowles, Horng Tzer Yau, and Jun Yin, *The local semicircle law for a general class of random matrices*, Electronic Journal of Probability **18** (2013). MR-3068390
- [9] László Erdős, Torbin Kruger, and Dominik Schroder, *Random Matrices with Slow Correlation Decay*, preprint, arXiv:1705.10661v2 (2017).
- [10] László Erdős, Sandrine Peche, Jose A. Ramirez, Benjamin Schlein, and Horng-Tzer Yau, *Bulk Universality for Wigner Matrices*, preprint, arXiv:0905.4176v2 (2009). MR-2662426
- [11] László Erdős and Horng-Tzer Yau, *Gap Universality of Generalized Wigner and beta-Ensembles*, preprint, arXiv:1211.3786 (2012). MR-3372074
- [12] László Erdős and Horng-Tzer Yau, *Dynamical Approach to Random Matrix Theory*, 2016. MR-3699468
- [13] László Erdős, Horng-Tzer Yau, and Jun Yin, *Universality for generalized Wigner matrices with Bernoulli distribution*, J. of Combinatorics (2011), no. 2, 15–85. MR-2847916
- [14] László Erdős, Horng-Tzer Yau, and Jun Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. (N. Y). **229** (2012), no. 3, 1435–1515. MR-2871147
- [15] S. Jaffard, *Proprietes des matrices “bien localisees” pr’es de leur diagonale et quelques applications*, Ann. Inst. H. Poincare Anal. Non Lineaire (1990), no. 7, (5)461–476. MR-1138533
- [16] Ben Landon and Horng-Tzer Yau, *Edge statistics of dyson brownian motion*, preprint (2017). MR-3687212
- [17] Terence Tao and Van Vu, *Random matrices: Universality of local eigenvalue statistics up to the edge*, Commun. Math. Phys. **298** (2010), no. 2, 549–572. MR-2669449
- [18] Terence Tao and Van Vu, *Random matrices: Universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1, 127–204. MR-2784665
- [19] Martin Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*.

Acknowledgments. We thank J. Huang and H.-T. Yau for useful discussions. The argument of Sec 3.3. came from a private communication with the two.

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>