UNIVERSALITY OF COVARIANCE MATRICES

By Natesh S. Pillai, Jun Yin

In this paper we prove the universality of covariance matrices of the form $H_{N\times N} = \frac{1}{N}X^{\dagger}X$ where $[X]_{M\times N}$ is a rectangular matrix with independent real valued entries $[x_{ij}]$ satisfying $\mathbb{E}\,x_{ij} = 0$ and $\mathbb{E}\,x_{ij}^2 = \frac{1}{M}$, $N, M \to \infty$. Furthermore it is assumed that these entries have sub-exponential tails. We will study the asymptotics in the regime $N/M = d_N \in (0,\infty), \lim_{N\to\infty} d_N \neq 1$. Our main result states that the Stieltjes transform of the empirical eigenvalue distribution of H is given by the Marcenko-Pastur law uniformly up to the edges of the spectrum with an error of order $(N\eta)^{-1}$ where η is the imaginary part of the spectral parameter in the Stieltjes transform. From this strong local Marcenko-Pastur law, we derive the following results. 1. The *rigidity of eigenvalues*: If $\gamma_j = \gamma_{j,N}$ denotes the *classical location* of the j-th eigenvalue under the Marcenko Pastur law ordered in increasing order, then the j-th eigenvalue λ_j of H is close to γ_j in the sense that for some positive constants C, c such that,

$$\mathbb{P}\Big(\exists\, j : |\lambda_j - \gamma_j| \geqslant (\log N)^{C\log\log N}\Big[\min\big(\,\min(N,M)-j, j\big)\Big]^{-1/3} N^{-2/3}\Big) \leqslant C\exp\big[-(\log N)^{c\log\log N}\big]$$

for N large enough. 2. The delocalization of the eigenvectors of the matrix $XX^{\dagger}$ uniformly both at the edge and the bulk. 3. Bulk universality, *i.e.*, n-point correlation functions of the eigenvalues of the sample covariance matrix $X^{\dagger}X$ coincide with those of the Wishart ensemble, when N goes to infinity. 4. Universality of the eigenvalues of the sample covariance matrix $X^{\dagger}X$ at *both* edges of the spectrum. Furthermore the first two results are applicable even in the case in which the entries of the column vectors of X are not independent but satisfy a certain large deviation principle. All our results hold for both real and complex valued entries.

1. Introduction. Covariance matrices are fundamental objects in modern multivariate statistics where the advance of technology has lead to high dimensional data. They have manifold applications in various applied fields; see [2, 13, 14, 15] for an extensive account on statistical applications, [12, 16] for applications in economics and [17] in population genetics to name a few. Except in special cases (under specific assumptions on the distributions of the entries of the covariance matrix such as Gaussian), the exact asymptotic distribution of the eigenvalues is not known. In this context, akin to the central limit theorem, the phenomenon

AMS 2000 subject classifications: 15B52, 82B44
Keywords and phrases: Covariance Matrix, Marcenko Pastur law, Universality, Dyson Brownian motion.

of universality helps us to obtain the asymptotic distribution of the eigenvalues, without having restrictive assumptions on the distribution on the entries. Borrowing a physical analogy, the key observation is that the eigenvalue gap distribution for a large complicated system is universal in the sense that it depends only on the symmetry class of the physical system but not on other detailed structures.

The covariance matrix formed by i.i.d standard Gaussian entries is the well studied Wishart matrix for which one has closed form expressions for many objects of interest including the joint distribution of the eigenvalues. Furthermore the empirical spectrum of the Wishart matrix converges to the Marcenko-Pastur law. In this paper we prove the universality of covariance matrices (both at the bulk and at the edges) under the assumption that the matrix entries are independent, have mean 0, variance 1 and have a sub-exponential tail decay. This implies that, asymptotically the distribution of the local statistics of eigenvalues of the covariance matrices of the above kind are identical to those of the Wishart matrix.

Over the past two decades, great progress have been made in proving the universality properties of i.i.d. matrix elements (*Standard Wigner ensembles*) (see [9] and the references there in). However results regarding universality for covariance matrices have been obtained only recently [1, 8, 18, 19, 20, 21]. Moreover these results are obtained under strong assumptions; for example in the "four moment theorem" of [23, 24], universality results are proved under the assumption that the first four moments of the matrix elements are equal to those of the standard Gaussian. In [8] the authors prove bulk unversality of covariance matrices under the assumption that distribution of matrix elements have a smooth density. These results, although quite interesting, exclude many important cases including the Bernoulli ensembles. On the other hand, we don't require the smoothness of the distribution of the matrix entries and only need the first two moments to be identical to those of the standard Gaussian. Furthermore, some of our results are applicable even in situations where the entries in same column are not independent, but satisfy a certain large deviation bound as explained below. Of course, we do require an exponential tail decay condition for the matrix entries. However in the future work, all of our results will be proved with the tail condition replaced by a uniform bound on p^{th} moment of the matrix elements (say $p = 5$ or 7), by the methods in [5].

The approach we take in this paper to prove universality is the one developed in a recent series of papers [4, 5, 6, 7, 8, 9, 10, 11]. The first step is to derive a strong *local Marcenko-Pastur law*, a precise estimate of the local eigenvalue density, which is our key technical tool for proving universality. En route to this, we also obtain precise bounds on the matrix elements of the corresponding Green function. For proving bulk universality of eigenvalues, the next step is to embed the Covariance matrix into a stochastic flow of matrices and so that the eigenvalues evolve according to a distinguished coupled system of stochastic differential equations, called the Dyson Brownian motion [3]. The central idea in the papers mentioned above is to estimate the time to local equilibrium for the Dyson Brownian motion with the introduction of a new stochastic flow, the *local relaxation flow*, which locally behaves like a Dyson Brownian motion but has a faster decay to global equilibrium. This approach [6, 8] entirely eliminates the usage of explicit formulas and it provides a unified proof for the universality. For proving edge universality of eigenvalues, we apply a "moment comparison" method based on the Green function, which is similar to the "four moment theorem" of [20, 21]. This idea has been recently used in [11] for proving edge universality of Wigner matrices.

More precisely, let $X = (x_{ij})$ be an $M \times N$ matrix with independent centered real valued entries of variance M^{-1}:

$$x_{ij} = M^{-1/2} q_{ij}, \quad \mathbb{E}\, q_{ij} = 0, \quad \mathbb{E}\, q_{ij}^2 = 1 \,. \tag{1.1}$$

Furthermore, the entries q_{ij} have a sub-exponential decay, *i.e.*, there exists a constant $\vartheta > 0$ such that for $u > 1$,

$$\mathbb{P}(|q_{ij}| > u) \leqslant \vartheta^{-1} \exp(-u^{\vartheta}) \,. \tag{1.2}$$

Notice that all our constants may depend on θ, but we will not denote this dependence. Define the Green function of $X^\dagger X$ by

$$G_{ij}(z) = \left(\frac{1}{X^\dagger X - z} \right)_{ij}, \qquad z = E + i\eta, \qquad E \in \mathbb{R}, \quad \eta > 0. \tag{1.3}$$

The Stieltjes transform of the empirical eigenvalue distribution of $X^\dagger X$ is given by

$$m(z) := \frac{1}{N} \sum_j G_{jj}(z) = \frac{1}{N} \operatorname{Tr} \frac{1}{X^\dagger X - z} \,. \tag{1.4}$$

We will be working the regime

$$d = d_N = N/M, \quad \lim_{N \to \infty} d \neq 1 \,.$$

Define

$$\lambda_\pm := \left(1 \pm \sqrt{d} \right)^2 . \tag{1.5}$$

The Marchenko-Pastur (henceforth abbreviated by MP) law is given by:

$$\varrho_W(x) = \frac{1}{2\pi d} \sqrt{\frac{\left[(\lambda_+ - x)(x - \lambda_-) \right]_+}{x^2}} \,. \tag{1.6}$$

We define $m_W(z)$, $z \in \mathbb{C}$, as the Stieltjes transform of ϱ_W, i.e.,

$$m_W(z) = \int_{\mathbb{R}} \frac{\varrho_W(x)}{(x-z)}\, dx \,. \tag{1.7}$$

The function m_W depends on d and has the closed form solution

$$m_W(z) = \frac{1-d-z+i\sqrt{(z-\lambda_-)(\lambda_+-z)}}{2\,d\,z}, \tag{1.8}$$

where $\sqrt{\ }$ denotes the square root on complex plane whose branch cut is the negative real line. One can check that $m_W(z)$ is the unique solution of

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

with $\Im\, m_W(z) > 0$ when $\Im z > 0$. Define the *normalized empirical counting function* by

$$\mathfrak{n}(E) := \frac{1}{N}\#\{\lambda_j \geqslant E\}. \tag{1.9}$$

Let

$$n_W(E) := \int_E^{\infty} \rho_W(x)\mathrm{d}x \tag{1.10}$$

so that $1 - n_W(\cdot)$ is the distribution function of the MP law.

By the singular value decomposition of X, there exist orthonormal bases $\{\mathbf{u}_1, \mathbf{u}_2 \ldots, \mathbf{u}_M\} \in \mathbb{C}^M$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\} \in \mathbb{R}^N$ such that

$$X \;=\; \sum_{\alpha=1}^{M} \sqrt{\lambda_\alpha}\, \mathbf{u}_\alpha \mathbf{v}_\alpha{}^\dagger \;=\; \sum_{\alpha=1}^{N} \sqrt{\lambda_\alpha}\, \mathbf{u}_\alpha \mathbf{v}_\alpha{}^\dagger \,, \tag{1.11}$$

where $\lambda_1 \geqslant \lambda_2 \ldots \lambda_{\max\{M,N\}} \geqslant 0$, $\lambda_\alpha = 0$ for $\min\{N,M\}+1 \leqslant \alpha \leqslant \max\{N,M\}$ and $\underline{\mathrm{v}}_\alpha = 0$, if $\alpha > N$ and $\underline{\mathrm{u}}_\alpha = 0$, for $\alpha > M$. We also define the classical location of the eigenvalues with ρ_W as follows

$$\int_{\gamma_j}^{\lambda_+} \varrho_W(x)\, dx = \int_{\gamma_j}^{+\infty} \varrho_W(x)\, dx = j/N \,. \tag{1.12}$$

Define the parameter

$$\varphi \;:=\; (\log N)^{\log\log N} \,. \tag{1.13}$$

For $\zeta \geqslant 0$, define the set

$$\underline{S}(\zeta) := \left\{ z \in \mathbb{C} : \mathbf{1}_{\mathbf{d}>\mathbf{1}}(\lambda_-/5) \leqslant E \leqslant 5\lambda_+ ,\ \varphi^{\zeta} N^{-1} \leqslant \eta \leqslant 10(1+d) \right\}. \tag{1.14}$$

Note that $m_W \sim O(1)$ in $\underline{S}(0)$.

Definition 1.1 (High probability events). *Let $\zeta > 0$. We say that an event Ω holds with* ζ-high probability *if there exists a constant $C > 0$ such that*

$$\mathbb{P}(\Omega^c) \leqslant N^C \exp(-\varphi^{\zeta}) \tag{1.15}$$

for large enough N.

Our goal is to estimate the following quantities

$$\Lambda_d := \max_k |G_{kk} - m_W|, \qquad \Lambda_o := \max_{k \neq \ell} |G_{k\ell}|, \qquad \Lambda := |m - m_W|, \tag{1.16}$$

where the subscripts refer to "diagonal" and "off-diagonal" matrix elements. All these quantities depend on the spectral parameter z and on N but for simplicity we suppress this in the notation.

The following is the main result of this paper:

Theorem 1.2 (Strong local Marchenko-Pastur law). *Let $X = [x_{ij}]$ with the entries x_{ij} satisfying (1.1) and (1.2). For any $\zeta > 0$ there exists a constant C_ζ such that the following events hold with ζ-high probability.*
(i) The Stieltjes transform of the empirical eigenvalue distribution of H satisfies

$$\bigcap_{z \in \underline{S}(C_\zeta)} \left\{ \Lambda(z) \leqslant \varphi^{C_\zeta} \frac{1}{N\eta} \right\}. \tag{1.17}$$

(ii) The individual matrix elements of the Green function satisfy that

$$\bigcap_{z \in \underline{S}(C_\zeta)} \left\{ \Lambda_o(z) + \Lambda_d \leqslant \varphi^{C_\zeta} \left(\sqrt{\frac{\Im\, m_W(z)}{N\eta}} + \frac{1}{N\eta} \right) \right\}. \tag{1.18}$$

(iii) The smallest non zero and largest eigenvalue of $X^\dagger X$ satisfy

$$\lambda_- - N^{-2/3}\varphi^{C_\zeta} \leqslant \min_{j \leqslant \min\{M,N\}} \lambda_j \leqslant \max_j \lambda_j \leqslant \lambda_+ + N^{-2/3}\varphi^{C_\zeta}. \tag{1.19}$$

(iv) Delocalization of the eigenvectors of $X^\dagger X$:

$$\max_{\alpha : \lambda_\alpha \neq 0} \|\mathbf{v}_\alpha\|_\infty \leqslant \varphi^{C_\zeta} N^{-1/2}. \tag{1.20}$$

The main theorem above is then used to the following results:

THEOREM 1.3 (Rigidity of the eigenvalues of covariance matrix). *Recall* γ_j *in* (1.12). *Let* $X = [x_{ij}]$ *with the entries* x_{ij} *satisfying* (1.1) *and* (1.2). *For any* $1 \leqslant j \leqslant N$, *let*

$$\widetilde{j} = \min\Big\{ \min\{N, M\} - j,\ j \Big\}.$$

For any $\zeta > 0$ *there exists a constant* C_ζ *such that*

$$|\lambda_j - \gamma_j| \leqslant \varphi^{C_\zeta} N^{-2/3} \widetilde{j}^{-1/3} \tag{1.21}$$

and

$$|\mathfrak{n}(E) - n_W(E)| \leqslant \varphi^{C_\zeta} N^{-1} \tag{1.22}$$

hold with ζ*-high probability for any* $1 \leqslant j \leqslant N$.

The above two results are stated under the assumption that the matrix entries are independent. The independence assumption (of the elements in each column vector of X) required in Theorems 1.2 and 1.3 may be replaced with the following large deviation criteria.

Let us first recall the following large deviation lemma for independent random variables (see [9], Appendix B for a proof).

LEMMA 1.4. *(Large Deviation Lemma) Suppose* a_i *be independent, mean* 0 *complex variables, with* $\mathbb{E}|a_i|^2 = \sigma^2$ *and have a sub-exponential decay as in* (1.2). *Then there exists a constant* $\rho \equiv \rho(\vartheta) > 1$ *such that, for any* $\zeta > 0$ *and for any* $A_i \in \mathbb{C}$ *and* $B_{ij} \in \mathbb{C}$, *the bounds*

$$\sum_{i=1}^{M} a_i A_i \leqslant (\log M)^{\rho\zeta} \sigma \|A\| \tag{1.23}$$

$$|\sum_{i=1}^{M} \bar{a}_i B_{ii} a_i - \sum_{i=1}^{M} \sigma^2 B_{ii}| \leqslant (\log M)^{\rho\zeta} \sigma^2 (\sum_{i=1}^{M} |B_{ii}|^2)^{1/2} \tag{1.24}$$

$$|\sum_{i \neq j} \bar{a}_i B_{ij} a_j| \leqslant (\log M)^{\rho\zeta} \sigma^2 (\sum_{i \neq j} |B_{ij}|^2)^{1/2} \tag{1.25}$$

hold with ζ*-high probability.*

Next we extend Theorems 1.2 and 1.3 by relaxing the independence assumption:

THEOREM 1.5. *Let $X = (x_{ij})$ be a random matrix with the entries satisfying (1.1) and assume that the column vectors of the matrix X are mutually independent. Furthermore, suppose that for any fixed $j \leqslant N$, the random variables defined by $a_i = x_{ij}, 1 \leqslant i \leqslant M$ satisfy the large deviation bounds (1.23), (1.24) and (1.25), for any $A_i \in \mathbb{C}$ and $B_{ij} \in \mathbb{C}$ and for any $\zeta > 0$. Then the conclusions of Theorem 1.2 and 1.3 hold for the random matrix X.*

Thus the above Theorem extends the universality results to a large class of matrix ensembles. For instance, let h_{ij} be a sequence of i.i.d random variables from a symmetric distribution and set

$$x_{ij} = \frac{h_{ij}}{\sqrt{\sum_{i=1}^{M} h_{ij}^2}}, \quad 1 \leqslant i \leqslant M, 1 \leqslant j \leqslant N . \tag{1.26}$$

Thus the entries of the column vector $(x_{1j}, x_{2j}, \cdots, x_{Mj})$ are not independent, but exchangable. Clearly $\mathbb{E}(x_{ij}) = 0, \mathbb{E}(x_{ij}^2) = \frac{1}{M}$. The random variables x_{ij} given by (1.26), are called self normalized sums and arise in various statistical applications.

Proof of Theorem (1.5): Actually in the proof of Theorem 1.2 and 1.3, we only use the large deviation properties of $a_i = x_{ij}$ instead of independence and sub-expotential decay. Therefore, the proof of Theorem 1.2 and 1.3 is already enough for Theorem (1.5).

□

THEOREM 1.6 (Universality of eigenvalues in Bulk). *Let $X^{\mathbf{v}} = [x_{ij}^{\mathbf{v}}]$ with the independent entries satisfying (1.1) and (1.2), as so $X^{\mathbf{w}}$. Let $E \in [\lambda_- + c, \lambda_+ - c]$ with some $c > 0$. Then for any $\varepsilon > 0$, $N^{-1+\varepsilon} < b < c/2$, any integer $n \geqslant 1$ and for any compactly supported continuous test function $O : \mathbb{R}^n \to \mathbb{R}$ we have*

$$\lim_{N\to\infty} \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int_{\mathbb{R}^n} O(\alpha_1, \dots, \alpha_n) \Big(p_{\mathbf{v}N}^{(n)} - p_{\mathbf{w},N}^{(n)}\Big) \Big(E' + \frac{\alpha_1}{N\rho_W(E)}, \dots, E' + \frac{\alpha_n}{N\rho_W(E)}\Big) \prod_i \frac{\mathrm{d}\alpha_i}{\rho_W(E)} = 0 \tag{1.27}$$

where $p_{\mathbf{v}N}^{(n)}$ and $p_{\mathbf{w},N}^{(n)}$, are the n-points correlation functions of the eigenvalues of $(X^{\mathbf{v}})^\dagger X^{\mathbf{v}}$ and $(X^{\mathbf{w}})^\dagger X^{\mathbf{w}}$, respectively.

THEOREM 1.7 (Universality of extreme eigenvalues). *Let $X^{\mathbf{v}} = [x_{ij}^{\mathbf{v}}]$ with independent entries satisfying (1.1) and (1.2), as so $X^{\mathbf{w}}$. Then there is an $\varepsilon > 0$ and $\delta > 0$ such that for any real number s (which may depend on N) we have*

$$\mathbb{P}^{\mathbf{v}}(N^{2/3}(\lambda_N - \lambda_+) \leqslant s - N^{-\varepsilon}) - N^{-\delta} \leqslant \mathbb{P}^{\mathbf{w}}(N^{2/3}(\lambda_N - \lambda_+) \leqslant s) \leqslant \mathbb{P}^{\mathbf{v}}(N^{2/3}(\lambda_N - \lambda_+) \leqslant s + N^{-\varepsilon}) + N^{-\delta}$$

(1.28)

for $N \geqslant N_0$ sufficiently large, where N_0 is independent of s. Analogous result hold for the smallest eigenvalue λ_1.

Theorem 1.7 can be extended to finite correlation functions of extreme eigenvalues. For example, we have the following extension to (1.28):

$$\begin{aligned}
&\mathbb{P}^{\mathbf{v}}\Big(N^{2/3}(\lambda_N-\lambda_+)\leqslant s_1-N^{-\varepsilon},\dots,N^{2/3}(\lambda_{N-k}-\lambda_+)\leqslant s_{k+1}-N^{-\varepsilon}\Big)-N^{-\delta}\\
&\leqslant \mathbb{P}^{\mathbf{w}}\Big(N^{2/3}(\lambda_N-\lambda_+)\leqslant s_1,\dots,N^{2/3}(\lambda_{N-k}-\lambda_+)\leqslant s_{k+1}\Big)\\
&\leqslant \mathbb{P}^{\mathbf{v}}\Big(N^{2/3}(\lambda_N-\lambda_+)\leqslant s_1+N^{-\varepsilon},\dots,N^{2/3}(\lambda_{N-k}-\lambda_+)\leqslant s_{k+1}+N^{-\varepsilon}\Big)+N^{-\delta}
\end{aligned} \tag{1.29}$$

for all k fixed and N sufficiently large. The proof of (1.29) is similar to that of (1.28) and we will not provide details except stating the general form of the Green function comparison theorem (Theorem 6.4) needed in this case. We remark that edge universality is usually formulated in terms of joint distributions of edge eigenvalues in the form (1.29) with fixed parameters $s_1, s_2, \dots$ *etc.* Our result holds uniformly in these parameters, *i.e.*, they may depend on N. However, the interesting regime is $|s_j| \leqslant \varphi^{O(1)}$, otherwise the rigidity estimate (1.21) gives a stronger control than (1.29).

The rest of the paper is organized as follows. In Sections 2-4 we establish the strong version of the Marcenko-Pastur law, rigidity and delocalization of eigenvalues. In Section 6-7, we respectively prove the bulk and edge universality results.

2. Apriori bound for the strong local Marcenko-Pastur law. We first prove a weaker form of Theorem 1.2, and in Section 4 we will use this apriori bound to obtain the stronger form as claimed in Theorem 1.2.

THEOREM 2.1. *Let $X = [x_{ij}]$ with the entries x_{ij} satisfying (1.1) and (1.2). For any $\zeta > 0$ there exists a constant C_ζ such that the following events hold with ζ-high probability.*

$$\bigcap_{z \in \underline{S}(C_\zeta)} \left\{ \Lambda_d(z) + \Lambda_o(z) \leqslant \varphi^{C_\zeta} \frac{1}{(N\eta)^{1/4}} \right\} \tag{2.1}$$

Before proceeding, let us introduce some notations. Define

$$\begin{aligned}
H := X^\dagger X, \quad G(z) := (H-z)^{-1} = (X^\dagger X - z)^{-1}, \quad m(z) := \frac{1}{N}\operatorname{Tr} G(z)\\
\mathcal{G}(z) := (XX^\dagger - z)^{-1}\,.
\end{aligned} \tag{2.2}$$

We know that the non-zero eigenvalues of $XX^\dagger$ and $X^\dagger X$ are identical and $XX^\dagger$ has $M-N$ more (or $N-M$ less) zero eigenvalues. We then have the identity

$$\operatorname{Tr} G(z) - \operatorname{Tr} \mathcal{G}(z) = \frac{M-N}{z} \,. \tag{2.3}$$

We shall often need to consider minors of X, which are the content of the following definition.

DEFINITION 2.2 (Minors). *Let $\mathbb{T} \subset \{1, \dots, N\}$. Then we define $X^{(\mathbb{T})}$ as the $(M \times (N - |\mathbb{T}|))$ minor of X obtained by removing all columns of X indexed by $i \in \mathbb{T}$. Note that we keep the names of indices of X when defining $X^{(\mathbb{T})}$.*

$$(X^{(\mathbb{T})})_{ij} \ := \ \underline{1}(j \notin \mathbb{T}) X_{ij} \,.$$

The quantities $G^{(\mathbb{T})}(z)$, $\mathcal{G}^{(\mathbb{T})}(z)$, $\lambda_\alpha^{(\mathbb{T})}$, $\underline{\mathrm{u}}_\alpha^{(\mathbb{T})}$, $\underline{\mathrm{v}}_\alpha^{(\mathbb{T})}$etc. are defined in the obvious way using $X^{(\mathbb{T})}$. Furthermore, we write abbreviate $(i) = (\{i\})$ as well as $(i\mathbb{T}) = (\{i\} \cup \mathbb{T})$. We also set

$$m^{(\mathbb{T})}(z) \ := \ \frac{1}{N} \sum_{i \notin \mathbb{T}} G_{ii}^{(\mathbb{T})}(z) \,. \tag{2.4}$$

We denote by $\mathbf{x}_i$ as the i-th column of X, which is a $M \times 1$ vector.

2.1. *Preliminary Lemmas.* We start with the following elementary lemma whose proof is standard:

LEMMA 2.3. *For any rectangular matrix M, and partition matrices, A, B and D of M given by $M = \begin{pmatrix} A & B \\ B^\dagger & D \end{pmatrix}$, we have the following identity*

$$M^{-1} = \begin{pmatrix} G^{-1} & -G^{-1} B D^{-1} \\ -D^{-1} B^\dagger G^{-1} & D^{-1} + D^{-1} B^\dagger G^{-1} B D^{-1} \end{pmatrix}, \quad G = A - B D^{-1} B^\dagger \,.$$

LEMMA 2.4. *For any z not in the spectrum of $X^\dagger X$,*

$$X (X^\dagger X - z)^{-1} X^\dagger = I + z (X X^\dagger - z)^{-1}$$

PROOF. Indeed from the SVD decomposition given in (1.11) we have

$$\begin{aligned} X (X^\dagger X - z)^{-1} X^\dagger &= \sum_\alpha \frac{\lambda_\alpha}{\lambda_\alpha - z} \underline{\mathrm{u}}_\alpha \underline{\mathrm{u}}_\alpha{}^\dagger \\ &= \sum_\alpha \Big(1 + \frac{z}{\lambda_\alpha - z}\Big) \underline{\mathrm{u}}_\alpha \underline{\mathrm{u}}_\alpha{}^\dagger = I + z (X X^\dagger - z)^{-1} \end{aligned}$$

and the lemma is proved. □

The next lemma collects the main identities of the resolvent matrix elements $G_{ij}^{(\mathbb{T})}$ and $\mathcal{G}_{ij}^{(\mathbb{T})}(z)$.

LEMMA 2.5 (Resolvent identities).

$$G_{ii}(z) = \frac{1}{-z - z\,\langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\,\mathbf{x}_i\rangle}, \quad i.e., \quad \langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\,\mathbf{x}_i\rangle = \frac{-1}{z\,G_{ii}(z)} - 1\,, \tag{2.5}$$

$$G_{ij}(z) = z\,G_{ii}(z)\,G_{jj}^{(i)}(z)\,\langle \mathbf{x}_i, \mathcal{G}^{(ij)}(z)\,\mathbf{x}_j\rangle, \quad i \neq j \tag{2.6}$$

$$G_{ij}(z) = \ G_{ij}^{(k)}(z) + \frac{G_{ik}(z)G_{kj}(z)}{G_{kk}(z)}, \quad i,j \neq k\ . \tag{2.7}$$

PROOF. First we show (2.5) with $i = 1$. Let $a = \mathbf{x}_1$ and $B = (X^{(1)})$. We have $X^\dagger = \begin{pmatrix} a^\dagger \\ B^\dagger \end{pmatrix}$, so that

$$X^\dagger X - z \ = \ \begin{pmatrix} a^\dagger a - z & a^\dagger B \\ B^\dagger a & B^\dagger B - z \end{pmatrix}\,.$$

By Lemmas 2.3 and 2.4

$$\begin{aligned} G_{11}(z) &= \left(\frac{1}{a^\dagger a - z - a^\dagger B(B^\dagger B - z)^{-1}B^\dagger a}\right)_{11} = \frac{1}{a^\dagger a - z - a^\dagger(1 + z(BB^\dagger - z)^{-1})\,a} \\ &= \frac{1}{-z - z\,a^\dagger(BB^\dagger - z)^{-1}\,a}\,. \end{aligned} \tag{2.8}$$

On the other hand, we have

$$\langle \underline{\mathrm{x}}_1, \mathcal{G}^{(1)}(z)\,\underline{\mathrm{x}}_1\rangle = a^\dagger(BB^\dagger - z)^{-1}\,a$$

which together with (2.8) implies (2.5). Next we prove (2.6). From Lemma 3.2 of [11] and $H = X^\dagger X$, we have the identity

$$G_{ij} = -G_{ii}G_{jj}^{(i)}\left(h_{ij} - \sum_{k,l\neq i,j} h_{ik}G_{kl}^{(ij)}h_{lj}\right) \tag{2.9}$$

i.e.,

$$G_{ij}(z) = -G_{ii}(z)\,G_{jj}^{(i)}(z)\,(h_{ij} - Z_{ij}), \quad Z_{ij} = \underline{\mathrm{x}}_i^\dagger X^{(ij)}G^{(ij)}X^{(ij)\dagger}\,\underline{\mathrm{x}}_j = \underline{\mathrm{x}}_i^\dagger(I + z\,\mathcal{G}^{(ij)})\,\underline{\mathrm{x}}_j, \tag{2.10}$$

where the last equality follows from an application of Lemma 2.4. Now (2.6) follows from (2.10). Finally (2.7) is proved in Lemma 3.2 of [11]. □

Set

$$\kappa := \min(|\lambda_+ - E|, |E - \lambda_-|). \tag{2.11}$$

LEMMA 2.6 (Properties of m_W). *Based on the definition of m_W, for $z \in \mathbf{S}(0)$, (see (1.14)) we have the following bounds:*

$$|m_W(z)| \sim 1, \qquad |1 - m_W^2(z)| \sim \sqrt{\kappa + \eta} \tag{2.12}$$

$$\Im\, m_W(z) \sim \begin{cases} \frac{\eta}{\sqrt{\kappa+\eta}} & \text{if } \kappa \geqslant \eta \text{ and } |E| \notin [\lambda_-, \lambda_+] \\ \sqrt{\kappa+\eta} & \text{if } \kappa \leqslant \eta \text{ or } |E| \in [\lambda_-, \lambda_+]. \end{cases} \tag{2.13}$$

where $A \sim B$ denotes $C^{-1}B \leqslant A \leqslant CB$ for some constants C. Furthermore

$$\frac{\Im\, m_W}{N\eta} \geqslant O(\frac{1}{N}) \quad \text{and} \quad \partial_\eta \frac{\Im m_W}{\eta} \leqslant 0 \tag{2.14}$$

For $z \in \underline{\mathrm{S}}(0)$, define the event

$$\underline{\mathrm{B}}(z) := \left\{ \Lambda_o(z) + \Lambda_d(z) > (\log N)^{-1} \right\}. \tag{2.15}$$

LEMMA 2.7 (Rough bounds of $\Lambda_o^{(\mathbb{T})}$ and $\Lambda_d^{(\mathbb{T})}$). *Fix $\mathbb{T} \subset \{1, 2, \cdots, N\}$. For $z \in \mathbf{S}(0)$, there exists a constant $C = C_{\mathbb{T}}$ such that the following estimates hold in $\underline{B}^c$:*

$$\max_{k \notin \mathbb{T}} |G_{kk}^{(\mathbb{T})} - G_{kk}| \leqslant C\Lambda_o^2 \tag{2.16}$$

$$\frac{1}{C} \leqslant |G_{kk}^{(\mathbb{T})}| \leqslant C \tag{2.17}$$

$$\Lambda_o^{(\mathbb{T})} \leqslant C\Lambda_o \tag{2.18}$$

PROOF. For $\mathbb{T} = \emptyset$, (2.16) , (2.18) follow from definition, (2.17) follows the definition of $\underline{\mathrm{B}}(z)$ and $m_W \sim 1$ in (2.12). For nonempty $\mathbb{T}$, one can prove the lemma using an induction on $|\mathbb{T}|$. For example, for $|\mathbb{T}| = 1$, using (2.7) we can show that

$$|G_{kk}(z) - G_{kk}^{(\mathbb{T})}(z)| \leqslant C\Lambda_o^2\,, \tag{2.19}$$

which implies the bound (2.16). A similar argument will yield (2.17), (2.18). □

On the other hand, in the case of $\eta = O(1)$, similar result of (2.17) holds without the assumption of $\underline{\mathrm{B}}^c$.

LEMMA 2.8 (Rough bounds for G_{kk} in large η case). *For any $z \in \mathbf{S}(0)$ and $\eta = O(1)$, we have the bound*

$$|G_{ii}(z)| \leqslant C \ ,$$

for some $C > 0$.

Proof: By definition

$$|G_{ii}| = \left| \sum_\alpha \frac{\mathbf{u}_\alpha(i) \overline{\mathbf{u}}_\alpha(i)}{\lambda_\alpha - z} \right| \leqslant \frac{1}{\eta} \sum_\alpha \mathbf{u}_\alpha(i) \overline{\mathbf{u}}_\alpha(i) \leqslant \frac{1}{\eta} \leqslant C$$

where we used $|\lambda_\alpha - z| \geqslant \Im z = \eta$. □

Define the quantity

$$\Psi := \sqrt{\frac{\Im m_W + \Lambda}{N\eta}} \ . \tag{2.20}$$

and

$$Z_i := z \langle \underline{\mathrm{x}}_i, \mathcal{G}^{(i)} \underline{\mathrm{x}}_i \rangle - \frac{z}{M} \operatorname{Tr} \mathcal{G}^{(i)} \ . \tag{2.21}$$

REMARK 2.9. *Note that if $m_W \leqslant O(1)$ and $\Lambda \leqslant O(1)$, then*

$$\Psi \leqslant O(N\eta)^{-1/2} \ . \tag{2.22}$$

We now identify the "bad sets" (improbable events) and show that they indeed have small probability. Define, for fixed z, the events

$$\Omega_o(z, K) := \Big\{ \Lambda_o(z) \geqslant K \Psi(z) \Big\} \tag{2.23}$$
$$\Omega_d(z, K) := \Big\{ \max_{ij} |G_{ii}(z) - m(z)| \geqslant K \Psi(z) \Big\} \ .$$

LEMMA 2.10. *Let $\Omega(z, K)^c$ be the good set where*

$$\Omega(z, K) := \Omega_d(z, K) \cup \Omega_o(z, K) \tag{2.24}$$

and

$$\Gamma(z,K)=\Omega(z,K)^c\cup \underline{B}(z)$$

For any $\zeta>0$ *there exists a constant* C_ζ *such that*

$$\bigcap_{z\in \underline{S}(C_\zeta)}\Gamma(z,\varphi^{C_\zeta}) \tag{2.25}$$

holds with ζ*-high probability.*

Proof. We only need to prove that there exists a uniform constant C_ζ such that for any $z\in \underline{\mathrm{S}}(C_\zeta)$ the event

$$\Gamma(z,\varphi^{C_\zeta}) \tag{2.26}$$

holds with ζ-high probability. It is clear that (2.25) follows from (2.26) and the fact that

$$|\partial_z G_{ij}|\leqslant N^C\ ,\quad \eta>N^{-1}\ . \tag{2.27}$$

Note $\Gamma(z,K)=(\Omega_o^c\cup B)\cap(\Omega_d^c\cup B)$. First we prove $\Omega_o^c\cup B$ holds with ζ-high probability. Using Lemma 1.4, Equation (2.6) and the fact that $|G|^2=G^*G$, we infer that there exists a constant C_ζ such that with ζ-high probability,

$$\begin{aligned}\Lambda_o\leqslant C|z|\max_{i\neq j}\left|\langle \underline{\mathrm{x}}_i,\mathcal{G}^{(ij)}\underline{\mathrm{x}}_j\rangle\right|&\leqslant \varphi^{C_\zeta}\frac{|z|}{N}\left(\sum_{kl}|\mathcal{G}^{(ij)}_{kl}|^2\right)^{1/2}\leqslant\varphi^{C_\zeta}\frac{|z|}{N}\left(\mathrm{Tr}\,|\mathcal{G}^{(ij)}|^2\right)^{1/2}\\ &\leqslant\varphi^{C_\zeta}|z|\sqrt{\frac{\Im\,\mathrm{Tr}\,\mathcal{G}^{(ij)}}{N^2\eta}}\ ,\qquad\text{in}\quad \underline{\mathrm{B}}^c\end{aligned} \tag{2.28}$$

where in the last step we used the identity $\frac{1}{\eta}\Im\,\mathrm{Tr}\,\mathcal{G}^{(ij)}=\mathrm{Tr}\,|\mathcal{G}^{(ij)}|^2$. Using the identity

$$\mathrm{Tr}\,G^{(\mathbb{T})}(z)-\mathrm{Tr}\,\mathcal{G}^{(\mathbb{T})}(z)=\frac{M-N+|\mathbb{T}|}{z}, \tag{2.29}$$

Equation (2.16) and $\Im(z^{-1})=\eta|z|^{-2}$ we have that with ζ-high probability

$$\Lambda_o\leqslant\varphi^{C_\zeta}\sqrt{\frac{\Im\, m_W+\Lambda+\Lambda_o^2}{N\eta}+\frac{1}{N}}\qquad\text{in}\quad\underline{\mathrm{B}}^c\ .$$

For the above choice of C_ζ, for $z \in \underline{S}(3C_\zeta)$, we have that with ζ-high probability

$$\Lambda_o \leqslant \varphi^{C_\zeta} \sqrt{\frac{\Im\, m_W + \Lambda}{N\eta} + \frac{1}{N}} + o(\Lambda_o) \qquad \text{in} \quad \underline{B}^c \tag{2.30}$$

Together, with (2.14), we have $\Omega_o^c \cup \underline{B}$ holds with ζ-high probability.

A similar argument using Lemma 1.4 will give

$$|Z_i| = |z| \left| \langle \underline{x}_i, \mathcal{G}^{(i)} \underline{x}_i \rangle - \frac{1}{M} \operatorname{Tr} \mathcal{G}^{(i)} \right| \leqslant \varphi^{C_\zeta} \Psi \,, \qquad \text{in} \quad \underline{B}^c \tag{2.31}$$

hold with ζ-high probability. Notice that $\max_i |G_{ii} - m| \leqslant \max_{i \neq j} |G_{ii} - G_{jj}|$. From (2.5) we obtain

$$\begin{aligned} |G_{ii} - G_{jj}| &\leqslant \Big| \frac{1}{-z - z \langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\, \mathbf{x}_i \rangle} - \frac{1}{-z - z \langle \mathbf{x}_j, \mathcal{G}^{(j)}(z)\, \mathbf{x}_j \rangle} \Big| \\ &\leqslant |G_{ii} G_{jj}| \left(|Z_i - Z_j| + \frac{|z|}{M} | \operatorname{Tr} \mathcal{G}^{(i)} - \operatorname{Tr} \mathcal{G}^{(j)} | \right) \\ &\leqslant C(\varphi^{C_\zeta} \Psi + \Lambda_o^2 + N^{-1}) \qquad \text{in} \quad \underline{B}^c \end{aligned}$$

hold with ζ-high probability, where the last inequality follows from (2.31), (2.3), (2.16) and (2.17). The lemma now follows from (2.30) and (2.22). □

On the other hand, in the case of $\eta = O(1)$, similar result holds without the assumption of $\underline{B}^c$.

Lemma 2.11. *Let $\Omega_o(z)$ and $\Omega_d(z)$ be as in (2.23). For any $\zeta > 0$, there exists a constant C_ζ such that the event*

$$\bigcap_{z \in \underline{S}(0), \eta \geqslant 1} \left(\Omega_d(z, \varphi^{C_\zeta}) \cup \Omega_o(z, \varphi^{C_\zeta}) \right)^c \bigcap \{ \max_i |Z_i| \leqslant \varphi^{C_\zeta} \Psi \} \tag{2.32}$$

holds with ζ-high probability.

Proof. With (2.27), we only need to prove (2.32) for fixed z. First we note in this case, we have $\Im m_W \sim 1$ and $\Lambda = O(1)$ and therefore

$$\Psi \sim N^{-1/2} \,. \tag{2.33}$$

It follows from (2.28) and Lemma 2.8 we have

$$\Lambda_o \leqslant \varphi^{C_\zeta} \sqrt{\frac{\Im \operatorname{Tr} \mathcal{G}^{(ij)}}{N^2}} \leqslant \varphi^{C_\zeta} N^{-1/2} \leqslant \varphi^{C_\zeta} \Psi.$$

The Z_i part can be proved as in (2.31), with Lemma 2.8. The Ω_d part can also be proved similarly as in above proof, where for $\operatorname{Tr}\mathcal{G}^{(i)} - \operatorname{Tr}\mathcal{G}^{(j)}$, we used

$$\operatorname{Tr}\mathcal{G}^{(i)} - \operatorname{Tr}\mathcal{G}^{(j)} = \operatorname{Tr}G^{(i)} - \operatorname{Tr}G^{(j)} = O(\eta)^{-1}$$

which follows the interlacing theorem of the eigenvalue, i.e.,

$$|m - m^{(i)}| \leqslant (N\eta)^{-1} \tag{2.34}$$

□

2.2. *Self consistent equations.* In last subsection, we have obtained the bound of Λ_o and $\max_i(G_{ii} - m)$ in term of m_W, η and Λ in B^c. In this subsection, we will give the desired bound for Λ and show B^c holds with ζ-high probability.

First we give the bound for Λ in the case of $\eta = O(1)$.

Lemma 2.12. *For any $\zeta > 0$, there exists a constant C_ζ such that*

$$\bigcap_{z\in \underline{S}(0),\eta=10(1+d)} \Lambda(z) \leqslant \varphi^{C_\zeta} N^{-1/4} \tag{2.35}$$

hold with ζ-high probability.

Proof. Recall (2.33). By definition and (2.5),

$$m(z) = \frac{1}{N}\sum_i G_{ii}(z) = \frac{1}{N}\sum_i \frac{1}{-z - z\frac{1}{M}\operatorname{Tr}\mathcal{G}^{(i)} - Z_i}\ .$$

Using (2.29) and (2.34), we obtain

$$\left| z\frac{1}{M}\operatorname{Tr}\mathcal{G}^{(i)} - zd\,m(z) + 1 - d \right| \leqslant CN^{-1}. \tag{2.36}$$

Together with $|Z_i| \leqslant \varphi^{C_\zeta}\Psi$ (see (2.32)), we have

$$m(z) = \frac{1}{N}\sum_i \frac{1}{1 - z - d - zdm(z) + Y_i}, \quad \max_i |Y_i| \leqslant \varphi^{C_\zeta}\Psi\ .$$

Since $|m| \leqslant \eta^{-1}$, we have $1 - z - d - zdm(z) \geqslant O(1)$, then

$$m(z) = \frac{1}{1 - z - d - zdm(z)} + O(\varphi^{C_\zeta}\Psi).$$

which implies (2.35). □

Now combining (2.35) with (2.32), we have proved that For any $\zeta > 0$, there exists a constant C_ζ such that, in the case $\eta = 10(1+d)$, (2.1) hold with ζ-high probability. It immediately implies that

$$\bigcap_{z \in \underline{\mathrm{S}}(0), \eta = 10(1+d)} B^c(z) \tag{2.37}$$

hold with ζ-high probability.

Now we prove (2.1) for general η. For a function $u(z)$, define its "deviance" to be

$$\mathcal{D}(u)(z) := (u^{-1}(z) + zd\, u(z)) - ({m_W}^{-1}(z) + zd\, m_W(z)) \tag{2.38}$$

Clearly, $\mathcal{D}(m_W) = 0$. Recall Z_i from (2.21) and define

$$[Z] = \frac{1}{N} \sum_{i=1}^{N} Z_i\,. \tag{2.39}$$

Recall the set $\underline{\mathrm{B}}(z)$ from (2.15) and $\Gamma(z)$ from Lemma 2.10.

LEMMA 2.13. *Let* $1 \leqslant K \leqslant (\log N)^{-1} (N\eta)^{-1/2}$, *on the set* $\Gamma(z, K)$ *(see* (2.24)*),*

$$|\mathcal{D}(m)| \leqslant O([Z]) + O(K^2 \Psi^2) + \infty\, 1_{\underline{B}(z)}$$

PROOF. Using (2.5), (2.16), (2.29) and the definition of m_W, we have that on the set $\Gamma(z, K)$

$$G_{ii}(z)^{-1} = m_W(z)^{-1} + zd\,[m_W(z) - m(z)] + O(K^2\Psi^2) + O(Z_i) + O(N^{-1}) \quad \text{in} \quad \underline{\mathrm{B}}^c \cap \Omega^c\,.$$

Then

$$G_{ii}^{-1} - m^{-1} = \mathcal{D}(m) + O(K^2\ \Psi^2) + O(Z_i) + O(N^{-1}) \quad \text{in} \quad \underline{\mathrm{B}}^c \cap \Omega^c \tag{2.40}$$

and summation over i yields

$$\frac{1}{N} \sum_{i=1}^{N} (G_{ii}^{-1} - m^{-1}) = \mathcal{D}(m) + O(K^2\Psi^2) + O(Z_i) + O(N^{-1}) \quad \text{in} \quad \underline{\mathrm{B}}^c \cap \Omega^c\,.$$

It follows from the assumptions $K \ll (N\eta)^{-1/2} \ll \Psi$ that $G_{ii} - m = o(1)$. Expanding the left hand side and using the facts that $\sum_i (G_{ii} - m) = 0$,

$$\sum_{i=1}^{N} (G_{ii}^{-1} - m^{-1}) = \sum_{i=1}^{N} \frac{G_{ii} - m}{G_{ii}\, m} = \frac{1}{m^3} \sum_{i=1}^{N} (G_{ii} - m)^2 + \sum_{i=1}^{N} O(\frac{(G_{ii} - m)^3}{m^4}) \quad \text{in} \quad \underline{\mathrm{B}}^c \cap \Omega^c$$

Together with (2.17) and (2.23), it follows that

$$\frac{1}{N}\sum_{i=1}^{N}(G_{ii}^{-1}-m^{-1}) = C\big(K\Psi\big)^2(1+K\Psi) \quad \text{in} \quad \underline{\mathrm{B}}^c\cap\Omega^c \tag{2.41}$$

Now the lemma follows from (2.40), (2.41) and the assumptions $K \ll (N\eta)^{-1/2} \leqslant O(\Psi)$. □

The two solutions m_1, m_2 of the equation $\mathcal{D}(m) = \delta(z)$ for a given $\delta(\cdot)$ are given by

$$\begin{aligned} m_{1,2} &= \frac{\delta(z)+1-d-z\pm i\sqrt{(z-\lambda_{-,\delta})(\lambda_{+,\delta}-z)}}{2\,d\,z} \\ \lambda_{\pm,\delta} &= 1+d\pm 2\sqrt{d-\delta(z)}-\delta(z)\,, \qquad |\lambda_{\pm,\delta}-\lambda_{\pm}| = O(\delta)\;. \end{aligned} \tag{2.42}$$

LEMMA 2.14. *Let $K, L > 0$, such that $\varphi^L \geqslant K^2(\log N)^4$, where L and K may depend on N. In any subset A of*

$$\bigcap_{z\in S(L)}\Gamma(z,K)\cap\bigcap_{z\in S(L),\eta=10(1+d)}B^c(z) \tag{2.43}$$

suppose we have the bound

$$|\mathcal{D}(m)(z)| \leqslant \delta(z)+\infty\,1_{\underline{B}(z)} \quad \forall z\in\underline{S}(L)$$

where $\delta : \mathbb{C}\mapsto\mathbb{R}_+$ is a continuous function, decreasing in $\Im z$ and $|\delta(z)| \leqslant (\log N)^{-8}$. Then for some uniform $C > 0$

$$|m(z)-m_W(z)| = \Lambda \leqslant C(\log N)\frac{\delta(z)}{\sqrt{\kappa+\eta+\delta}} \quad \forall z\in\underline{S}(L)\;. \tag{2.44}$$

holds in A and

$$A\subset\bigcap_{z\in S(L)}B^c \tag{2.45}$$

Note: The difficulty in the proof is that the bound $\mathcal{D}(m) \leqslant \delta(z)$ only in the set $\underline{\mathrm{B}}$ but we need to prove (2.45)

PROOF. Let us first fix E and define the set

$$I_E = \Big\{\eta : \Lambda_o(E+i\hat\eta)+\Lambda_d(E+i\hat\eta) \leqslant \frac{1}{\log N}, \quad \forall\hat\eta\geqslant\eta\,, E+i\hat\eta\in\underline{\mathrm{S}}(L)\Big\}\;.$$

We first prove (2.44) for all $z = E + i\eta$ with $\eta \in I_E$. Define

$$\eta_1 = \sup_{\eta \in I_E} \left\{ \eta : \delta(E + i\eta) \geqslant (\log N)^{-1}(\kappa + \eta) \right\} .$$

Since δ is a continuos decreasing function of η by assumption, $\delta(E+i\eta) \leqslant (\log N)^{-1}(\kappa+\eta_1)$ for $\eta \geqslant \eta_1$. Let m_1 and m_2 be the two solutions of the equation $\mathcal{D}(m) = \delta(z)$ as given in (2.42). (Note that since we are in $\underline{\mathrm{B}}$ by assumption we do have $\mathcal{D}(m) \leqslant \delta(z)$.) Then it can be easily verified that

$$\begin{aligned} |m_1 - m_2| &\geqslant C\sqrt{\kappa + \eta}, & \eta \geqslant \eta_1 \\ &\leqslant C(\log N)\sqrt{\delta(z)}, & \eta \leqslant \eta_1. \end{aligned} \tag{2.46}$$

The difficulty here is that we don't know which of the two solutions m_1, m_2 is equal to m. However for $\eta = O(1)$, we claim that $m = m_1$. With assumption, $|m - m_W| = \Lambda \leqslant \Lambda_d \ll 1$. Also a direct calculation using (2.42) gives

$$|m_1 - m_W| = C\frac{\delta(z)}{\sqrt{\kappa + \eta}} \ll \frac{1}{\log N} . \tag{2.47}$$

Since $|m_1 - m_2| \geqslant C\sqrt{\kappa+\eta}$ for $\eta = O(1)$(see (2.46)), it immediately follows that $m = m_1$ for $\eta = O(1)$. Furthermore since the functions m_1, m_2 and m are continuous and since $m_1 \neq m_2$, it follows that $m = m_1$ for $\eta \geqslant \eta_1$. Thus $\eta \geqslant \eta_1$,

$$|m(z) - m_W(z)| = |m_1(z) - m_W(z)| \leqslant C\frac{\delta(z)}{\sqrt{\kappa+\eta}} \leqslant C\frac{\delta(z)}{\sqrt{\kappa+\eta+\delta}}$$

where in the last step we have used $\delta \leqslant \kappa + \eta$.

For $\eta \leqslant \eta_1$, we take advantage of the fact that the difference $|m_1 - m_2|$ is the same order as in (2.47). Indeed, for $\eta \leqslant \eta_1$, if $m = m_2$ (say), then

$$|m - m_W| \leqslant |m_2 - m_1| + |m_1 - m_W| \leqslant (\log N)\sqrt{\delta(z)} \leqslant C(\log N)\frac{\delta(z)}{\sqrt{\kappa+\eta+\delta}}$$

verifying (2.44) for $\eta \in I_E$.

Now we prove that I_E equals to the desired region $[\varphi^L N^{-1}, 5]$, i.e. (2.45). We argue by contradiction. If not, let $\eta_0 = \inf I_E$, with continuity, we have

$$\Lambda_o(z_0) + \Lambda_d(z_0) = (\log N)^{-1} , z_0 = E + i\eta_0 \tag{2.48}$$

and thus $\Lambda(z_0) \leqslant \Lambda_d(z_0) \leqslant (\log N)^{-1}$. On the other hand, with above result, i.e, (2.44) holds for $\eta \in I_E$, we have

$$\Lambda(z_0) \leqslant (\log N)^{-3} \tag{2.49}$$

By definition

$$\left\{\Lambda_o(z_0)+\Lambda_d(z_0)=(\log N)^{-1}\right\}\cap\Gamma(z_0)=(\Omega_o(z_0)\cup\Omega_d(z_0))^c\;,$$

and therefore

$$\Lambda_o(z_0)+\max_k|G_{kk}(z_0)-m(z_0)|\leqslant CK\,\Psi(z_0)\;.$$

With assumptions $\varphi^L\geqslant K^2(\log N)^4$, we have $\Psi(z_0)\leqslant\sqrt{\frac{\Im m_W}{N\eta}+\frac{\Lambda(z_0)}{N\eta}}\ll K^{-1}(\log N)^{-2}$ which immediately implies that $\Lambda_o(z_0)+\max_k|G_{kk}(z_0)-m(z_0)|\ll(\log N)^{-1}$. Using this estimate and (2.49) we deduce that

$$\Lambda_o(z_0)+\Lambda_d(z_0)\leqslant\Lambda_o(z_0)+\max_k|G_{kk}(z_0)-m(z_0)|+\Lambda\ll\log N^{-1}$$

which contradicts (2.48) and concludes the proof of the lemma. □

Proof of Lemma 2.1: Now we complete the proof of Lemma 2.1. It follows from (2.31), Lemma 2.10 and 2.13 that for any $\zeta>0$, there exists C_ζ, D_ζ and $\widetilde{C}_\zeta$ that

$$\bigcap_{z\in\underline{\mathrm{S}}(C_\zeta)}|\mathcal{D}(m)(z)|\leqslant\varphi^{\widetilde{C}_\zeta}\Psi+\infty\,1_{\underline{\mathrm{B}}(z)}$$

holds on

$$\bigcap_{z\in\underline{\mathrm{S}}(C_\zeta)}\Gamma(z,\varphi^{D_\zeta})\tag{2.50}$$

which is with ζ-high probability. Choosing larger C_ζ, applying Lemma 2.14 with choosing A being (2.43), with (2.14), we obtain that for some C_ζ,

$$\Lambda(z)\leqslant\varphi^{C_\zeta}\Psi^{1/2},\quad\forall z\in\underline{\mathrm{S}}(C_\zeta)\tag{2.51}$$

holds on (2.43). Using (2.50) and (2.35), we obtain that for any $\zeta>0$, there exists C_ζ such that (2.51) holds with ζ-high probability. Furthermore, (2.45) implies

$$\bigcap_{z\in\underline{\mathrm{S}}(C_\zeta)}B^c(z)$$

is with ζ-high probability. Together with (2.25), (2.51), we obtain (2.1) and complete the proof of Lemma 2.1.

3. Strong bound on $[Z]$. For proving Theorem 1.2 and 1.3, the key input is the following lemma which gives a much stronger bound of $[Z]$. The following is the main result of this section:

LEMMA 3.1. *Let $K, L > 0$, such that $\varphi^L \geqslant K^2(\log N)^4$. Suppose for some event*

$$\Xi \subset \cap_{z \in \underline{S}(L)} \left(\Gamma(z, K) \cap B^c(z)\right),$$

we have

$$\Lambda(z) \leqslant \tilde{\Lambda}(z), \quad \forall z \in \underline{S}(L)$$

with some deterministic number $\tilde{\Lambda}(z)$ and $\mathbb{P}(\Xi^c) \leqslant e^{-p(\log N)^2}$ where p depends on N and

$$1 \ll p \ll (\log NK)^{-1} \varphi^{L/2}. \tag{3.1}$$

Then there exists a subset Ξ' of Ξ such that $\mathbb{P}(\Xi \setminus \Xi') \leqslant e^{-p}$ and for any $z \in \underline{S}(L)$,

$$\left|z[Z]\right| \leqslant C p^5 K^2 \widetilde{\Psi}^2, \quad \widetilde{\Psi} := \sqrt{\frac{\Im\, m_W + \tilde{\Lambda}}{N\eta}}, \quad \text{in} \quad \Xi' \tag{3.2}$$

Note: In the application of this lemma, p_N and $K = O(\varphi^{O(1)})$. First, we are going to introduce the abstract Z lemma, which is similar to Theorem 5.6 of [4]. Also see [11] for a similar lemma for generalized Wigner matrices.

THEOREM 3.2 (Abstract decoupling lemma). *Let $\mathcal{I}$ be finite set which may depend on N and*

$$\mathcal{I}_i \subset \mathcal{I}, \quad 1 \leqslant i \leqslant N$$

Let $\mathcal{Z}_1, \ldots, \mathcal{Z}_N$ be random variables which depend on the independent random variables $\{x_\alpha, \alpha \in \mathcal{I}\}$. Let $\mathbb{E}_i$ denote the expectation value respect to $\{x_\alpha, \alpha \in \mathcal{I}_i\}$ and $\mathbb{IE}_i = 1 - \mathbb{E}_i$. Define the commuting projection operators

$$Q_i = \mathbb{IE}_i, \; P_i = \mathbb{E}_i, \quad P_i^2 = P_i, \; Q_i^2 = Q_i, \quad [Q_i, P_j] = [P_i, P_j] = [Q_i, Q_j] = 0$$

and for $A \subset \{1, 2, \ldots, N\}$

$$Q_A := \prod_{i \in A} Q_i, \quad P_A := \prod_{i \in A} P_i$$

We use the notation

$$[\mathcal{Q}\mathcal{Z}] \; = \; \frac{1}{N} \sum_{i=1}^{N} Q_i \mathcal{Z}_i \,.$$

Let Ξ be an event and p an even integer. Suppose following assumptions hold with some constants C_0, $c_0 > 0$.

(i) (Bound on $Q_A\mathcal{Z}_i$ in Ξ). There exist deterministic positive numbers $X<1$ and Y such that for any set $A\subset\{1,2,\dots,N\}$ with $i\in A$ and $|A|\leqslant p$, $Q_A\mathcal{Z}_i$ in Ξ can be written as the sum of two new random variables

$$\mathbf{1}(\Xi)(Q_A\mathcal{Z}_i)=\mathcal{Z}_{i,A}+\mathbf{1}(\Xi)Q_A\mathbf{1}(\Xi^c)\widetilde{\mathcal{Z}}_{i,A} \tag{3.3}$$

and

$$|\mathcal{Z}_{i,A}|\leqslant Y\big(C_0X|A|\big)^{|A|},\quad |\widetilde{\mathcal{Z}}_{i,A}|\leqslant YC_0^{|A|}N^{C_0} \tag{3.4}$$

(ii) (Rough bound on $\mathcal{Z}_i$).

$$\max_i|\mathcal{Z}_i|\ \leqslant\ YN^{C_0}\,. \tag{3.5}$$

(iii) (Ξ has high probability).

$$\mathbb{P}[\Xi^c]\ \leqslant\ \mathrm{e}^{-c_0(\log N)^{3/2}p}\,. \tag{3.6}$$

Then, under the assumptions (i) – (iii), we have

$$\mathbb{E}\Bigg[\mathbf{1}(\Xi)[\mathcal{Q}\mathcal{Z}]^p\Bigg]\leqslant(Cp)^{4p}\big[X^2+N^{-1}\big]^pY^p \tag{3.7}$$

for some $C>0$ and any sufficiently large N.

Before we give the proof, we introduce a trivial but useful identity

$$\prod_{i=1}^n(x_i+y_i)=\sum_{s=1}^{n+1}\left[\left(\prod_{i=1}^{s-1}x_i\right)\prod_{i=s}y_{i(=s)}\left(\prod_{i=s+1}^n(x_i+y_i)\right)\right] \tag{3.8}$$

with the convention that $\prod_{i\in\emptyset}=1$. It implies that

$$\left|\prod_{i=1}^n(x_i+y_i)-\prod_{i=1}^n(x_i)\right|\leqslant n\max_i|y_i|\left(\max_i|x_i+y_i|+\max_i|x_i|\right)$$

For any $1\leqslant k\leqslant n$, it follows from $\prod_{i=1}^n(x_i+y_i)=(x_k+y_k)\prod_{i\neq k}(x_i+y_i)$ and (3.8) that

$$\prod_{i=1}^n(x_i+y_i)=\sum_{s\neq k,s=1}^n(x_k+y_k)\left[\left(\prod_{i\neq k,i=1}^{s-1}x_i\right)\prod_{i=s}y_{i(=s)}\left(\prod_{i\neq k,i=s+1}^n(x_i+y_i)\right)\right] \tag{3.9}$$

Proof of Lemma 3.2 First, by definition, we have

$$\mathbb{E}\left[\mathbf{1}(\Xi)[\mathcal{Q}\mathcal{Z}]^p\right] = \frac{1}{N^p}\sum_{j_1,\ldots,j_p}\mathbb{E}\mathbf{1}(\Xi)\prod_{\alpha=1}^{p} Q_{j_\alpha}\mathcal{Z}_{j_\alpha} \tag{3.10}$$

For fixed $j_1,\ldots,j_p$, let $T_\alpha = Q_{j_\alpha}\mathcal{Z}_{j_\alpha}$. Now using (3.9) with choosing $k=1$, $x_i, = P_{j_1}T_i$ and $y_i = Q_{j_1}T_i$ in (3.9) (Note: here $x_i + y_i = T_i$), we have

$$\prod_{\alpha=1}^{p} T_\alpha = \sum_{s=2}^{p+1} T_1\left[\left(\prod_{\alpha<s,\alpha\neq 1} P_{j_1}T_\alpha\right)(Q_{j_1}T_s)\left(\prod_{\alpha>s,\alpha\neq 1} T_\alpha\right)\right]$$

We define $A_{\alpha,s} := \mathbf{1}_{\{\alpha<s,\alpha\neq 1\}}\{j_1\}$ and $B_{\alpha,s} := \mathbf{1}_{\alpha=s}\{j_1\}$, i.e., $B_{\alpha,s} = \{j_1\}$ if $\alpha = s$ otherwise $A_{\alpha,s} = \emptyset$. It is clear that $A_{1,s} = B_{1,s} = \emptyset$. Then

$$\prod_{\alpha=1}^{p} T_\alpha = \sum_{s=2}^{p+1}\prod_{\alpha} P_{A_{\alpha,s}}Q_{B_{\alpha,s}}T_\alpha$$

For generalization, we replace s with s_1 and write it as

$$\prod_{\alpha=1}^{p} T_\alpha = \sum_{s_1=1}^{p+1}\mathbf{1}(s_1\neq 1)\prod_{\alpha} P_{A_{\alpha,s_1}}Q_{B_{\alpha,s_1}}T_\alpha$$

and

$$A_{\alpha,s_1} = \{j_1 :\ \alpha < s_1, \alpha\neq 1\},\quad B_{\alpha,s_1} = \{j_1 :\ s_1 = \alpha\} \tag{3.11}$$

Iterating for $1\leqslant j_1, j_2,\ldots,j_p\leqslant N$, we have

$$\prod_{\alpha=1}^{p} T_\alpha = \sum_{s_1,s_2,\ldots,s_p=1}^{p+1}\prod_{i}\mathbf{1}(s_i\neq i)\prod_{\alpha} P_{A_{\alpha,\mathbf{s}}}Q_{B_{\alpha,\mathbf{s}}}T_\alpha,$$

where $\mathbf{s}$ denotes $s_1, s_2\cdots, s_p$ and $A_{\alpha,\mathbf{s}}$ and $B_{\alpha,\mathbf{s}}$ are defined as

$$A_{\alpha,\mathbf{s}} = \{j_i :\ \alpha < s_i, \alpha\neq i\},\quad B_{\alpha,\mathbf{s}} = \{j_i :\ s_i = \alpha\}$$

Then

$$\left|\mathbb{E}\mathbf{1}(\Xi)\prod_{\alpha=1}^{p} Q_{j_\alpha}\mathcal{Z}_{j_\alpha}\right| \leqslant (2p)^p\max_{\mathbf{s}}\prod_{i}\mathbf{1}(s_i\neq i)\left|\mathbb{E}\mathbf{1}(\Xi)\prod_{\alpha} P_{A_{\alpha,\mathbf{s}}}Q_{B_{\alpha,\mathbf{s}}}T_\alpha\right|.$$

Now to prove (3.7), it only remains to show that for any $\{j_1, \ldots, j_p\}$ and $\mathbf{s} = \{s_1, s_2, \cdots, s_p\}$ such that $s_i \neq i$, we have

$$\left| \mathbb{E}\mathbf{1}(\Xi) \prod_\alpha P_{A_{\alpha,\mathbf{s}}} Q_{B_{\alpha,\mathbf{s}}} T_\alpha \right| \leqslant (Cp)^{2p} Y^p X^{2t}, \quad t := |\{j_1, \ldots, j_p\}| \tag{3.12}$$

for simplicity, we denote $A_{\alpha,\mathbf{s}}$ and $B_{\alpha,\mathbf{s}}$ by A_α and B_α and similarly we denote the characteristic function $\mathbf{1}(\Xi)$ by Ξ, i.e., we need to prove

$$\left| \mathbb{E}(\Xi) \prod_\alpha P_{A_\alpha} Q_{B_\alpha} T_\alpha \right| \leqslant (Cp)^{2p} Y^p X^{2t}, \quad t := |\{j_1, \ldots, j_p\}| \tag{3.13}$$

Since $T_1 = Q_{j_1} T_1$ and P and Q's are communitive, we have

$$\mathbb{E}(\Xi) \prod_\alpha P_{A_\alpha} Q_{B_\alpha} T_\alpha = \mathbb{E}(Q_{j_1} P_{A_1} Q_{B_1} T_1) \left(\Xi \prod_{\alpha=2}^p (P_{A_\alpha} Q_{B_\alpha}) T_\alpha \right) \tag{3.14}$$

Recall $\mathcal{I}$ and $\mathcal{I}_i$ in assumption. For $\alpha = 2, \ldots, p$, if $j_1 \in \cap_{\alpha \neq 1} A_\alpha$, then $(P_{A_\alpha} Q_{B_\alpha}) T_{j_\alpha}$ are independent of $\mathcal{I}_{j_1}$, i.e., any x_t, $t \in \mathcal{I}_{j_1}$. For any function f independent of $\mathcal{I}_{j_1}$, since $Q = Q_{j_1}$ is an projection operator we have

$$|\mathbb{E}(Qg)(\Xi f)| = |\mathbb{E}(Qg)(Q\Xi) f| \leqslant \|(Qg) f\|_2 \|Q\Xi\|_2 = \|(Qg) f\|_2 \sqrt{\mathbb{E}|Q\Xi^c|^2} \leqslant \|(Qg) f\|_2 \sqrt{\mathbb{P}(\Xi^c)}$$

where we have used the Schwarz inequality. In our application,

$$Q = Q_{j_1}, \quad f = \prod_{\alpha=2}^p (P_{A_\alpha} Q_{B_\alpha}) T_\alpha, \quad g = P_{A_1} Q_{B_1} T_1$$

Since P_i and Q_i's are projections, we have

$$\|(Qg) f\|_2 \leqslant (CY)^p N^{Cp}$$

and obtain that (3.14) is bounded above by $Y^p N^{Cp} \exp[-c(\log N)^{3/2} p]$ with (3.6). We note that this part can be neglected in proving (3.13).

Hence we can assume that $j_1 \notin \cap_{\alpha \neq 1} A_\alpha$, i.e., $1 < s_1 \leqslant p$, (see (3.11)) i.e., $j_1 \in \cup_{\alpha \neq 1} B_\alpha$. Similarly for j_i, we have $j_i \in \cup_{\alpha \neq i} B_\alpha$ here $i = 2, \ldots p$. Recall that $j_\alpha \notin B_\alpha$. With these two conditions, B_α's satisfy the inequality

$$p + t \geqslant \sum_\alpha |B_\alpha \cup \{j_\alpha\}| \geqslant 2t, \quad t := |\{j_1, \ldots, j_p\}| \tag{3.15}$$

Now it only remains to prove (3.13) under the condition (3.15). First, we write

$$\mathbb{E}\mathbf{1}(\Xi)\prod_{\alpha} P_{A_\alpha}Q_{B_\alpha}T_\alpha = \mathbb{E}\,\Xi\prod_{\alpha=1}^{p}(P_{A_\alpha}Q_{\widetilde{B}_\alpha}\mathcal{Z}_{j_\alpha}), \quad \widetilde{B}_\alpha := B_\alpha\cup\{j_\alpha\}$$

Using (3.8) with $x = P\Xi Q\mathcal{Z}$ and $y = P\Xi^c Q\mathcal{Z}$ ($x+y = PQ\mathcal{Z}$), we have

$$\mathbb{E}\Xi\prod_{\alpha=1}^{p}(P_{A_\alpha}Q_{\widetilde{B}_\alpha}\mathcal{Z}_{j_\alpha}) = \sum_{s=1}^{p+1}\left(\mathbb{E}\Xi\prod_{i=1}^{s-1}(P_{A_i}(\Xi)Q_{\widetilde{B}_i}\mathcal{Z}_{j_i})\Big(P_{A_s}(\Xi^c)Q_{\widetilde{B}_s}\mathcal{Z}_{j_s}\Big)\prod_{i=s+1}^{p}(P_{A_i}Q_{\widetilde{B}_i}\mathcal{Z}_{j_i})\right) \tag{3.16}$$

First for $s\leqslant p$, one can use the following fomular: for any f and h

$$\mathbb{E}|h(P\Xi^c Qf)| \leqslant \|h\|_\infty\|(\Xi^c Qf)\|_2 \leqslant \sqrt{\mathbb{P}(\Xi^c)}\;\|f\|_\infty\|h\|_\infty$$

Let

$$h = \prod_{i=1}^{s-1}(P_{A_i}(\Xi)Q_{\widetilde{B}_i}\mathcal{Z}_{j_i})\prod_{i=s+1}^{p}(P_{A_i}Q_{\widetilde{B}_i}\mathcal{Z}_{j_i}), \quad f = \mathcal{Z}_{j_s}, \quad P = P_{A_s}, \quad Q = Q_{\widetilde{B}_s}$$

By (3.5) and $p\geqslant 1$, we have

$$|h| \leqslant Y^{p-1}N^{Cp}, \quad |f|\leqslant YN^C$$

Then with (3.6), we proved that $\sum_{s=1}^{p}$ of r.h.s of (3.16) is bounded above by $Y^pN^{Cp}\exp[-c(\log N)^{3/2}p]$ which can be neglected in proving (3.13). Then it only remains to bound the r.h.s of (3.16) in the case $s = p+1$, i.e., to prove

$$\left|\mathbb{E}\Xi\prod_{\alpha=1}^{p}(P_{A_\alpha}\Xi Q_{\widetilde{B}_\alpha}\mathcal{Z}_{j_\alpha})\right| \leqslant (Cp)^{2p}Y^pX^{2t}, \quad t := |\{j_1,\ldots,j_p\}| \tag{3.17}$$

under the assumption (3.15). Using (3.3) and (3.8), with $x = P\Xi\mathcal{Z}$ and $y = P\Xi Q\Xi^c\widetilde{\mathcal{Z}}$ we can write the l.h.s. of (3.17) as

$$\begin{aligned}(3.18)\qquad &\mathbb{E}\Xi\prod_{\alpha=1}^{p}(P_{A_\alpha}\Xi Q_{\widetilde{B}_\alpha}\mathcal{Z}_{j_\alpha})\\ &= \sum_{s=1}^{p+1}\left(\mathbb{E}\Xi\prod_{i=1}^{s-1}\Big(P_{A_i}(\Xi)\mathcal{Z}_{j_i,\widetilde{B}_i}\Big)\Big(P_{A_s}(\Xi)Q_{\widetilde{B}_s}(\Xi^c)\widetilde{\mathcal{Z}}_{j_s,\widetilde{B}_s}\Big)\prod_{i=s+1}^{p}(P_{A_i}\Xi Q_{\widetilde{B}_i}\mathcal{Z}_{j_i})\right)\end{aligned}$$

Now we repeat the argument for (3.16). For $s \leqslant p$, one can use the following fomula: for any f and h

$$\mathbb{E}|h(P\Xi Q\Xi^c f)| \leqslant \|h\|_\infty \|(\Xi^c f)\|_2 \leqslant \sqrt{\mathbb{P}(\Xi^c)}\, \|f\|_\infty \|h\|_\infty$$

Let

$$h = \Xi \prod_{i=1}^{s-1} \left(P_{A_i}(\Xi) \mathcal{Z}_{j_i, \widetilde{B}_i} \right) \prod_{i=s+1}^{p} \left(P_{A_i} \Xi Q_{\widetilde{B}_i} \mathcal{Z}_{j_i} \right), \quad f = \widetilde{\mathcal{Z}}_{j_s, \widetilde{B}_s}, \quad P = P_{A_s}, \quad Q = Q_{\widetilde{B}_s}$$

With the assumption in (3.4), we know $\sum_{s=1}^{p}$ of r.h.s of (3.18) is bounded above by

$$Y^p N^{Cp} \exp[-c(\log N)^{3/2} p]$$

which can be neglected in proving (3.17). For the main term, i.e., $s = p+1$ in r.h.s. of (3.18), using (3.4) and (3.15), we have

$$\mathbb{E}\Xi \prod_{\alpha=1}^{p} (P_{A_\alpha} \Xi \mathcal{Z}_{j_\alpha, \widetilde{B}_\alpha}) \leqslant (CY)^p (C_0 X p)^{2t} \leqslant (CYp^2)^p X^{2t}$$

and complete the proof of Theorem 3.2.

□

We note that, with(2.5) and (2.21), we can write zZ as

$$zZ_i = Q_i \Big[\frac{-1}{G_{ii}} \Big], \quad Q_i := 1 - P_i, \quad P_i := \mathbb{E}_{\mathbf{x}_i} \tag{3.19}$$

LEMMA 3.3. *Let $\mathcal{Z}_i = (G_{ii})^{-1}$, P_i and Q_i defined as in (3.19). We assume that $\eta = \Im z \geqslant N^{-C}$ for some $C > 0$. Suppose there exists p and Ξ, such that $\mathbb{P}(\Xi^c) \leqslant e^{-p(\log N)^{3/2}}$ and in Ξ*

$$\max_i |Q_i \mathcal{Z}_i| \leqslant CYX, \quad \frac{\Lambda_o(z)}{\min_i |G_{ii}(z)|} \leqslant CX \ll 1, \quad \min_i |G_{ii}(z)| \geqslant Y^{-1}, \quad p \leqslant \frac{C}{(\log N) X} \tag{3.20}$$

where X and Y are deterministic numbers. Then there exists $\Xi' \subset \Xi$ with $\mathbb{P}((\Xi')^c) \leqslant e^{-p}$ and

$$\left| \frac{1}{N} \sum_i Q_i \mathcal{Z}_i \right| \leqslant Cp^5 \left(X^2 + N^{-1} \right) Y \tag{3.21}$$

Proof: We are going to apply Theorem 3.2. Then (3.21) follows from (3.7) and Markov inequality. First one can easily prove (3.5) and (3.6). It only remains to show that for $i \in A \subset \{1, 2 \cdots, N\}$ and $|A| \leqslant p$ there exists $\mathcal{Z}_{i,A}$ and $\widetilde{\mathcal{Z}}_{i,A}$

$$\mathbf{1}(\Xi)(Q_A \mathcal{Z}_i) = \mathcal{Z}_{i,A} + \mathbf{1}(\Xi) Q_A(\Xi^c)\widetilde{\mathcal{Z}}_{i,A}, \quad \mathcal{Z}_{i,A} \leqslant Y\big(C_0 X |A|\big)^{|A|}, \quad \widetilde{\mathcal{Z}}_{i,A} \leqslant Y C^{|A|} N^{C} \tag{3.22}$$

It holds in the case $A = \{i\}$ by the assumption. Then we assume that $|A| \geqslant 2$. As in Lemma 5.1 in [4] let $\mathcal{A} = \mathcal{A}(H) = \mathcal{A}(X^\dagger X)$ be a quantity defined with $X^\dagger X$, we define

$$(\mathcal{A})^{S,U} := \sum_{S/U \subset V \subset S} (-1)^{|V|} \mathcal{A}^{(V)}, \quad A^{(V)} := A\left((X^{(V)})^\dagger (X^{(V)})\right)$$

then

$$\mathcal{A} = \sum_{U \subset S} (\mathcal{A})^{S,U}$$

By definition, $(\mathcal{A})^{S,U}$ is independent of the j-th column of X if $j \in S/U$. Therefore,

$$Q_S \mathcal{A} = Q_S (\mathcal{A})^{S,S}$$

In our case,

$$Q_{A/\{i\}} \mathcal{Z}_i = Q_A \left(\frac{1}{G_{ii}}\right)^{A/\{i\}, A/\{i\}},$$

Then we choose

$$\mathcal{Z}_{i,A} := \mathbf{1}(\Xi) Q_A \Xi \left(\frac{1}{G_{ii}}\right)^{A/\{i\}, A/\{i\}}, \quad \widetilde{\mathcal{Z}}_{i,A} := \left(\frac{1}{G_{ii}}\right)^{A/\{i\}, A/\{i\}}$$

It is easy to prove the bound for $\widetilde{\mathcal{Z}}_{i,A}$ in (3.22) with definition. For the bound of $\mathcal{Z}_{i,A}$, it only remains to prove that for $2 \leqslant |A| \leqslant p_N$

$$\left| \mathbf{1}(\Xi) \left(\frac{1}{G_{ii}}\right)^{A/\{i\}, A/\{i\}} \right| \leqslant Y\big(C_0 X |A|\big)^{|A|} \tag{3.23}$$

To prove it, we first show that for $|T| \leqslant p$,

$$\max_{i,j \notin T} |G_{ij}^{(T)}| \leqslant C \max_{i,j} |G_{ij}| \quad \min_{i \notin T} |G_{ii}^{(T)}| \geqslant c \min_i |G_{ii}|, \tag{3.24}$$

We start from $|T| = 1$, i.e., $T = \{k\}$. First using (2.7) and the assumptions of this lemma, we have

$$(G_{ii})^{-1} = \frac{-G_{ij}G_{ji}}{G_{ii}G_{jj}G_{ii}^{(j)}} + (G_{ii}^{(j)})^{-1} = (1 + O(X^2))(G_{ii}^{(j)})^{-1}$$

and

$$|G_{ij}^{(k)}| = \left| G_{ij} - \frac{G_{ik}G_{kj}}{G_{kk}} \right| \leqslant \Lambda_o(1 + O(X))$$

It shows that

$$\max_{i,j \neq k} |G_{ij}^{(k)}| \leqslant (1 + O(X)) \max_{i,j} |G_{ij}|, \quad \min_{i \neq k} |G_{ii}^{(k)}| \geqslant (1 - O(X)) \min_{i} |G_{ii}|$$

Then with induction on $|T|$ and the assumption $Xp \ll 1$, we obtain the desired result (3.24).

Now we return to prove (3.23) from the case $|A| = 2$. If $i \neq j$, with (2.7), (3.24) and (3.20), we have

$$\left(\frac{1}{G_{ii}} \right)^{j,j} = (G_{ii})^{-1} - (G_{ii}^{(j)})^{-1} = \frac{-G_{ij}G_{ji}}{G_{ii}G_{jj}G_{ii}^{(j)}} \leqslant O(YX^2)$$

The general case has been proved in Lemma 5.11 of [4], which gives that

$$\left(\frac{1}{G_{ii}} \right)^{A/\{i\},A/\{i\}} \leqslant (C|A|)^{|A|} \frac{\left(\max_{i,j \notin T, T \subset A/\{i\}} |G_{ij}^{(T)}| \right)^{|A|}}{\left(\min_{j \notin T, T \subset A/\{i\}} |G_{jj}^{(T)}| \right)^{|A|+1}}$$

Together with (3.24) and (3.20), we obtain (3.23) and complete the proof. At last we need to point out that the definition of $G_{ij}^{(V)}$ $(ij \notin V)$ in [4] is different from this paper, though they are equivalent. The one we use in this paper is defined as

$$G^{(V)} = \left((X^{(V)})^{\dagger} (X^{(V)}) - z \right)^{-1}$$

In [4], it is defined as

$$G^{(V)} = \left(H^{(V)} - z \right)^{-1}$$

where $H^{(V)}$ is the minor of H obtained by removing all $i - th$ row and columns of H indexed by $i \in V$. But one can see that if $H = X^{\dagger}X$ then $H^{(V)} = (X^{(V)})^{\dagger}(X^{(V)})$. □

Proof of Lemma 3.1: It is a special case of Lemma 3.3 with $X = K\widetilde{\Psi}$ and $Y = C$ for some large C. First $\max_i |Q_i \mathcal{Z}_i| \leqslant CYX$ is proved in (2.31). By assumptions, $\Xi \subset \cap_{z\in \underline{S}(L)} (\Gamma(z,K) \cap B^c(z))$, then

$$\Lambda_o, \Lambda_d \leqslant K\Psi \leqslant K\widetilde{\Psi} = X \leqslant CK(N\eta)^{-1/2} \ll 1$$

in Ξ. Then we have

$$\frac{\Lambda_o(z)}{\min_i |G_{ii}(z)|} \leqslant CX \ll 1, \quad \min_i |G_{ii}(z)| \geqslant Y^{-1}$$

Furthermore, (3.1) and $\eta \geqslant N^{-1}\varphi^L$ (since $z \in S(L)$) imply $p \leqslant C((\log N)X)^{-1}$ and completes the proof.

□

4. Strong Marcenko Pastur law and rigidity of eigenvalues.

4.1. *Proof of Theorem 1.2 .* First, we assume $\zeta \geqslant 1$. With lemma 2.10 and lemma 2.1, for any $\zeta > 0$, there exists C_ζ, such that

$$\Xi_1 \subset \bigcap_{z\in \underline{S}(C_\zeta)} B^c(z) \cap \Gamma(z, C_\zeta) \tag{4.1}$$

holds with $(\zeta+2)$-high probability. Then with Lemma 2.13, for $z \in \underline{S}(3C_\zeta)$, we have

$$|\mathcal{D}(m)(z)| \leqslant \varphi^{2C_\zeta}\Psi^2 + O[Z], \quad \text{in} \quad \Xi_1$$

Let $\Lambda_1 = 1$, then $\Lambda \leqslant \Lambda_1$ in Ξ_1. Therefore, we can apply Lemma 3.1 with

$$p = p_1 = -\log[1 - \mathbb{P}(\Xi_1)]/(\log N)^2$$

Without loss of generality, we can assume that $\mathbb{P}(\Xi_1)$ is not too close to 1, otherwise, we can choose Ξ as a subset of itself. Then with Definition 1.1, we have

$$p_1 = C\varphi^{\zeta+2}/(\log N)^2$$

We assume that $C_\zeta \geqslant 6\zeta$ then (3.1) holds and (3.2) gives that for $z \in \underline{S}(3C_\zeta)$, we have that for some

$$\Xi_2 \subset \Xi_1, \quad \mathbb{P}(\Xi_2) = e^{-p_1}$$

and

$$\big|[Z]\big| \leqslant \varphi^{2C_\zeta+11\zeta}\Psi_1^2, \quad \Psi_1 := \sqrt{\frac{\Im\, m_W + \Lambda_1}{N\eta}}, \quad \text{in} \quad \Xi_2$$

Since in $\Xi_2 \subset \Xi_1$, by assumption, $\Lambda \leqslant \Lambda_1$, then $\Psi \leqslant \Psi_1$ in $\Xi_2 \subset \Xi_1$, i.e.,

$$|\mathcal{D}(m)(z)| \leqslant \varphi^{2C_\zeta+11}\frac{\Im\, m_W + \Lambda_1}{N\eta}, \quad \text{in} \quad \Xi_2 \tag{4.2}$$

Then applying Lemma 2.14, (2.45) shows that for $z \in \underline{\mathrm{S}}(3C_\zeta)$

$$\Lambda(z) \leqslant \Lambda_2(z) := \varphi^{C_\zeta+6\zeta}\Lambda_1^{1/2}(N\eta)^{-1/2}, \quad \text{in} \quad \Xi_2$$

Repeating this process, by choosing

$$p_2 = -\log[1 - \mathbb{P}(\Xi_2)]/(\log N)^2 = C\varphi^{\zeta+2}/(\log N)^4$$

we have that there exists Ξ_3 such that

$$\Xi_3 \subset \Xi_2, \quad \mathbb{P}(\Xi_3) = e^{-p_2}$$

for $z \in \underline{\mathrm{S}}(3C_\zeta)$

$$\Lambda(z) \leqslant \Lambda_3(z) := \varphi^{C_\zeta+6\zeta}\Lambda_2^{1/2}(N\eta)^{-1/2} \leqslant \varphi^{2C_\zeta+12\zeta}(N\eta)^{-3/4}, \quad \text{in} \quad \Xi_3$$

Now we iterate this process K times, $K := \log\log N/\log 1.9$. For $k \leqslant K$, we have that for some

$$\Xi_k \subset \Xi_{k-1}, \quad \mathbb{P}(\Xi_k) = e^{-p_{k-1}}$$

where

$$p_{N,k} = -\log[1 - \mathbb{P}(\Xi_{k-1})]/(\log N)^2 = C\varphi^{\zeta+2}/(\log N)^{2k} \geqslant \varphi^\zeta$$

and for $z \in \underline{\mathrm{S}}(3C_\zeta)$

$$\Lambda(z) \leqslant \Lambda_{k+1}(z) := \varphi^{C_\zeta+6\zeta}\Lambda_k^{1/2}(N\eta)^{-1/2} \leqslant \varphi^{2C_\zeta+12\zeta}(N\eta)^{-1+(1/2)^k}, \quad \text{in} \quad \Xi_{k+1}$$

Note:

$$N^{(1/2)^K} \leqslant \varphi$$

Then for $k=K$, for $z \in \underline{\mathbf{S}}(3C_\zeta)$

$$\Lambda(z) \leqslant \Lambda_{k+1}(z) \leqslant \varphi^{2C_\zeta+12\zeta}(N\eta)^{-1+(1/2)^K} \leqslant \varphi^{2C_\zeta+12\zeta+1}(N\eta)^{-1} \tag{4.3}$$

holds with ζ-high probability and complete the proof of (1.17). Furthermore, since $\Xi_{K+1} \subset \Xi_1$ with (4.1), we obtain (1.18).

Now we suppose (1.19) holds and prove (1.20) first. Using (1.18), we have

$$\max_{\lambda_-/5 \leqslant E \leqslant 5\lambda_+} \Im G_{ii}(E + i\varphi^{C_\zeta}N^{-1}) \leqslant C \tag{4.4}$$

By definition,

$$\Im G_{ii} = \sum_\alpha \frac{|\mathbf{v}_\alpha(i)|^2\eta}{(\lambda_\alpha - E)^2 + \eta^2}$$

Then choosing $E = \lambda_\alpha$ and $\eta = \varphi^{C_\zeta}N^{-1}$, with (4.4), we have for any α

$$|\mathbf{v}_\alpha(i)|^2 \leqslant \eta = \varphi^{C_\zeta}N^{-1}$$

which implies (1.20). Here the (1.19) guarantees $\lambda_-/5 \leqslant E \leqslant 5\lambda_+$.

Now it only remains to prove (1.19). The proof proceeds via the following four steps.

Step one of proof on (1.19): First we prove

$$\begin{aligned} \lambda_- - N^{-2/3}\varphi^{C_\zeta} &\leqslant \min\{\lambda_j : 1_{d>1}\lambda_-/5 \leqslant \lambda_j \leqslant 5\lambda_+\} \\ &\leqslant \max\{\lambda_j : 1_{d>1}\lambda_-/5 \leqslant \lambda_j \leqslant 5\lambda_+\} \leqslant \lambda_+ + N^{-2/3}\varphi^{C_\zeta} \end{aligned} \tag{4.5}$$

By repeating the iteration one more time, i.e., replace Λ_1 in (4.2) with Λ_{k+1} in (4.3), we have

$$|\mathcal{D}(m)(z)| \leqslant \varphi^{C_\zeta}\frac{\Im m_W + \frac{1}{N\eta}}{N\eta}$$

for some large C_ζ. With (2.44) again, we obtain that for some $D_\zeta \geqslant 1$

$$\Lambda(z) \leqslant \varphi^{D_\zeta}\frac{\delta}{\sqrt{\kappa+\eta+\delta}}, \quad \delta := \left(\frac{\Im m_W}{N\eta} + \frac{1}{(N\eta)^2}\right) \tag{4.6}$$

For any $E : E \geqslant \lambda_+ + N^{-2/3}\varphi^{4D_\zeta}$, we choose $z = E + i\eta$ and

$$\eta := \varphi^{-D_\zeta}N^{-1/2}\kappa^{1/4}, \quad \kappa = E - \lambda_+$$

then it is easy to check, with $\kappa \geqslant N^{-2/3}\varphi^{4D_\zeta}$, that

$$\kappa \gg \varphi^{D_\zeta}\eta, \quad N\eta\sqrt{\kappa} \gg \varphi^{D_\zeta}, \quad \frac{\sqrt{\kappa}}{N\eta^2} \gg 1 \tag{4.7}$$

With (2.13) and $\kappa \geqslant \eta$, we have

$$\Im\, m_W(z) = C\frac{\eta_E}{\sqrt{\kappa}} \tag{4.8}$$

which implies

$$\delta \leqslant \frac{C}{N\sqrt{\kappa}} + (N\eta)^{-2}\,.$$

Therefore, $\kappa \geqslant \delta$. Together with (4.6) and (4.7), we have

$$\Lambda(z) \leqslant C\varphi^{D\zeta}\left(\frac{\eta}{\kappa} + \frac{1}{N\eta\sqrt{\kappa}}\right)\frac{1}{N\eta} \ll \frac{1}{N\eta}$$

Combining (4.8) and the last inequality of (4.7), we have

$$\Im\, m_W(z) \ll \frac{1}{N\eta}$$

Therefore, we obtain

$$\Im\, m(z) \ll \frac{1}{N\eta}$$

Note: if $\Im\, m(z) < (2N\eta)^{-1}$, $(z = E + i\eta)$ then the number of the eigenvalue in $[E-\eta, E+\eta]$ is zero, which follows from

$$\Im\, m = \frac{1}{N}\sum_\alpha \frac{\eta}{(\lambda_\alpha - E)^2 + \eta^2} \geqslant \sum_{\alpha:|\lambda_\alpha - E|\leqslant\eta} \frac{1}{2N\eta} \tag{4.9}$$

Since it holds for any $E \geqslant \lambda_+ + N^{-2/3}\varphi^{4D_\zeta}$. Then we have proved that for any $\zeta > 0$, there exists some $D_\zeta > 0$ such that

$$\max\{\lambda_j : \lambda_j \leqslant 5\lambda_+\} \leqslant \lambda_+ + N^{-2/3}\varphi^{4D_\zeta}$$

hold with ζ-high probability. An analogous bound for the smallest eigenvalue can be proved similarly.

Step two of proof on (1.19): Recall $\mathfrak{n}(E)$ in (1.9) and $n_W(E)$ in (1.10). Then we prove that

$$\Big|\,(\mathfrak{n}(E_1)-\mathfrak{n}(E_2))-(n_W(E_1)-n_W(E_2))\,\Big| \leqslant \frac{C(\log N)\varphi^{C_\zeta}}{N}, \quad E_1,E_2\in[\mathbf{1_{d>1}}\lambda_-/4,\ 4\lambda_+] \tag{4.10}$$

which implies that

$$\#\left\{j:\lambda_j\notin[\mathbf{1_{d>1}}\lambda_-/5,\ 5\lambda_+]\right\}\leqslant\varphi^{C_\zeta} \tag{4.11}$$

We note that though we only need (4.11) for (1.19), but (4.10) is very useful for Theorem 1.3. Then we prove it at here. The proof is similar to the one of Theorem 2.2 in [11]. We now translate the information on the Stieltjes transform obtained in Theorem 1.2 to prove (4.10) on the location of the eigenvalues.

Lemma 4.1. *Let ϱ^Δ be a signed measure on the real line. For any $E_1,E_2\in[A_1,\ A_2]$, with $|A_{1,2}|\leqslant O(1)$ and $\eta=N^{-1}$ we define $f(\lambda)=f_{E_1,E_2,\eta}(\lambda)$ to be a characteristic function of $[E_1,E_2]$ smoothed on scale η, i.e., $f\equiv1$ on $[E_1+\eta,E_2-\eta]$, $f\equiv0$ on $\mathbb{R}\setminus[E_1,E_2]$ and $|f'|\leqslant C\eta^{-1}$, $|f''|\leqslant C\eta^{-2}$. Let m^Δ be the Stieltjes transform of ϱ^Δ. Suppose for some positive number U (may depend on N) we have*

$$|m^\Delta(x+iy)|\leqslant\frac{CU}{Ny}\qquad\text{for}\quad y<1,\quad x\in[A_1,\ A_2], \tag{4.12}$$

Then

$$\left|\int_{\mathbb{R}}f_{E_1,E_2,\eta}(\lambda)\varrho^\Delta(\lambda)\mathrm{d}\lambda\right|\leqslant\frac{CU|\log\eta|}{N} \tag{4.13}$$

with some constant C.

Proof of Lemma 4.1: For simplicity, we drop the Δ superscript in the proof. Using Helffer-Sjostrand functional calculus, let $\chi(y)$ be a smooth cutoff function with support in $[-1,1]$, with $\chi(y)=1$ for $|y|\leqslant1/2$ and with bouded derivatives.

$$f(\lambda)=\frac{1}{2\pi}\int_{\mathbb{R}^2}\frac{iyf''(x)\chi(y)+i(f(x)+iyf'(x))\chi'(y)}{\lambda-x-iy}$$

Since f and χ are real,

$$\begin{aligned}\left|\int f(\lambda)\varrho(\lambda)\mathrm{d}\lambda\right|\leqslant\quad & C\int_{\mathbb{R}^2}(|f(x)|+|y||f'(x)|)|\chi'(y)||m(x+iy)|\mathrm{d}x\mathrm{d}y\\ &+C\left|\int_{|y|\leqslant\eta}\int yf''(x)\chi(y)\,\Im\, m(x+iy)\mathrm{d}x\mathrm{d}y\right|\\ &+C\left|\int_{|y|\geqslant\eta}\int_{\mathbb{R}}yf''(x)\chi(y)\,\Im\, m(x+iy)\mathrm{d}x\mathrm{d}y\right|,\end{aligned} \tag{4.14}$$

The first term is estimated by, with (4.12),

$$\int_{\mathbb{R}^2} (|f(x)| + |y||f'(x)|)|\chi'(y)||m(x+iy)|\mathrm{d}x\mathrm{d}y \leqslant CU. \tag{4.15}$$

For the second term in r.h.s of (4.14) we use that from (4.12) it follows for any $1 \geqslant y > 0$ that

$$y|\Im m(x+iy)| \leqslant CU. \tag{4.16}$$

With $|f''| \leqslant C\eta^{-2}$ and

$$\mathrm{supp} f'(x) \subset \{|x - E_1| \leqslant \eta\} \cup \{|x - E_2| \leqslant \eta\}, \tag{4.17}$$

we get

$$\text{second term in r.h.s of (4.14)} \leqslant CU.$$

Now we integrate the third term in (4.14) by parts first in x, then in y. Then we bound it with absolute value by

$$C\int_{\mathbb{R}} \eta|f'(x)||\operatorname{Re} m(x+i\eta)|\mathrm{d}x + C\int_{\mathbb{R}^2} y|f'(x)\chi'(y)\operatorname{Re} m(x+iy)| \tag{4.18}$$
$$+\frac{C}{\eta}\int_{\eta\leqslant y\leqslant 1}\int_{\mathrm{supp} f'} |\operatorname{Re} m(x+iy)|\mathrm{d}x\mathrm{d}y.$$

By using (4.12) and (4.17) in the first term, (4.15) in the second and (4.12) in the third, we have

$$(4.18) \leqslant CU + CU\eta^{-1}\int_{\mathrm{supp} f'} \mathrm{d}x \int_{\eta\leqslant y\leqslant 1} \frac{1}{yN}\mathrm{d}y \leqslant CU|\log\eta|$$

□

We will apply this lemma with $[A_1, A_2] = [\mathbf{1_{d>1}}\lambda_-/4,\ 4\lambda_+]$ and the choice that the signed measure is the difference of the empirical density and the WP law,

$$\varrho^{\Delta}(\mathrm{d}\lambda) = \varrho(\mathrm{d}\lambda) - \varrho_W(\lambda)\mathrm{d}\lambda, \qquad \varrho(\mathrm{d}\lambda) := \frac{1}{N}\sum_i \delta(\lambda_i - \lambda).$$

Now we prove that (4.10) holds. By Theorem 1.2, the assumptions of Lemma 4.1 hold for the difference $m^{\Delta} = m - m_W$ $U = \varphi^{C_\zeta}$ if $y \geqslant y_0 := \varphi^{C_\zeta}/N$. For $y \leqslant y_0$, set $z = x + iy$, $z_0 = x + iy_0$ and estimate

$$|m(z) - m_W(z)| \leqslant |m(z_0) - m_W(z_0)| + \int_y^{y_0} \left|\partial_\eta\big(m(x+i\eta) - m_{sc}(x+i\eta)\big)\right|\mathrm{d}\eta. \tag{4.19}$$

Note that

$$
\begin{aligned}
|\partial_\eta m(x+i\eta)| =& \Big|\frac{1}{N}\sum_j \partial_\eta G_{jj}(x+i\eta)\Big| \\
\leqslant& \frac{1}{N}\sum_{jk} |G_{jk}(x+i\eta)|^2 = \frac{1}{N\eta}\sum_j \Im\, G_{jj}(x+i\eta) = \frac{1}{\eta}\Im\, m(x+i\eta),
\end{aligned}
$$

and similarly

$$
|\partial_\eta m_W(x+i\eta)| = \Big|\int \frac{\varrho_W(s)}{(s-x-i\eta)^2}\mathrm{d}s\Big| \leqslant \int \frac{\varrho_W(s)}{|s-x-i\eta|^2}\mathrm{d}s = \frac{1}{\eta}\Im\, m_W(x+i\eta).
$$

Now we use the fact that the functions $y \to y\,\Im\, m(x+iy)$ and $y \to y\,\Im\, m_W(x+iy)$ are monotone increasing for any $y>0$ since both are Stieltjes transforms of a positive measure. Therefore the integral in (4.19) can be bounded by ($z_0 = x+iy_0$)

$$
\int_y^{y_0} \frac{\mathrm{d}\eta}{\eta}\big[\,\Im\, m(x+i\eta) + \Im\, m_W(x+i\eta)\big] \leqslant y_0\big[\,\Im\, m(z_0) + \Im\, m_W(z_0)\big]\int_y^{y_0}\frac{\mathrm{d}\eta}{\eta^2} \tag{4.20}
$$

By definition, $\Im\, m_{sc}(x+iy_0) \leqslant |m_{sc}(x+iy_0)| \leqslant C$. By the choice of y_0 and Theorem 1.2, we have

$$
\Im\; m(x+iy_0) \leqslant \Im\, m_{sc}(x+iy_0) + \frac{\varphi^{C_\zeta}}{Ny_0} \leqslant C \tag{4.21}
$$

with very high probability. Together with (4.20) and (4.19), this proves that (4.12) holds for $y \leqslant y_0$ as well if U is increased to $U = C\varphi^{C_\zeta}$.

The application of Lemma 4.1 shows that for any $\eta \geqslant 1/N$

$$
\left|\int_{\mathbb{R}} f_{E_1,E_2,\eta}(\lambda)\varrho(\lambda)\mathrm{d}\lambda - \int_{\mathbb{R}} f_{E_1,E_2,\eta}(\lambda)\varrho_{sc}(\lambda)\mathrm{d}\lambda\right| \leqslant \frac{C(\log N)\varphi^{C_\zeta}}{N}. \tag{4.22}
$$

With the fact $y \to y\,\Im\, m(x+iy)$ is monotone increasing for any $y>0$, (4.21) implies a crude upper bound on the empirical density. Indeed, for any interval $I := [x-\eta, x+\eta]$, with $\eta = 1/N$, we have

$$
\mathfrak{n}(x+\eta) - \mathfrak{n}(x-\eta) \leqslant C\eta\;\Im\; m\big(x+i\eta\big) \leqslant Cy_0\;\Im\; m\big(x+iy_0\big) \leqslant \frac{C\varphi^{C_\zeta}}{N}. \tag{4.23}
$$

Together with (4.22), we have proved (4.10).

Step three of proof on (1.19): Now we prove $\lambda_j \leqslant 5\lambda_+$, holds with ζ-high probability. Since in previous proof, 5 is not a special number, then it only remains to prove that for some large K, the following bound holds with ζ-high probability,

$$\lambda_j \leqslant K\lambda_+.$$

Let

$$z = E + i\eta, \quad E \geqslant K\lambda_+, \quad \eta = EN^{-2/3} \tag{4.24}$$

With (4.10) with $E_1 = \lambda_-$ and $E_2 = K\lambda_+$, we have proved that there are at least $\varphi^{O(1)}$ eigenvalues larger than $K\lambda_+$. Then by definition,

$$\Im\, m(z) \leqslant \frac{C\eta}{E^2} + \frac{\varphi^{C_\zeta}}{N\eta}, \quad |\operatorname{Re} m(z)| \leqslant CE^{-1} + \frac{\varphi^{C_\zeta}}{N\eta} \leqslant O(E^{-1}), \quad E \geqslant K\lambda_+ \tag{4.25}$$

so as $m^{(\mathbb{T})}$ for $|\mathbb{T}| = O(1)$. Now using Lemma 1.4, as in (2.28) and (2.31), we have

$$|Z_i| \leqslant |E| \left(E^{-1}N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta} \right), \quad \langle \mathbf{x}_i | \mathcal{G}^{(i,j)} | \mathbf{x}_j \rangle \leqslant E^{-1}N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta} \tag{4.26}$$

First we estimate G_{ii}, with (2.5)

$$|G_{ii}| = \left| \frac{1}{-1 - z - d - zdm^{(i)}(z) + Z_i} \right| \leqslant CE^{-1}$$

and G_{ij} with (2.6)

$$G_{ij} \leqslant E^{-1} \left(\frac{\varphi^{C_\zeta}}{N\eta} + E^{-1}N^{-1/2} \right) \tag{4.27}$$

where we used (4.25), (4.26) and the fact E is large enough and $\eta = EN^{-2/3}$. Furthermore

$$m^{(i)} - m \leqslant \frac{1}{N} \sum \frac{G_{ji}G_{ij}}{G_{ii}} \leqslant E^{-1} \left(\frac{\varphi^{C_\zeta}}{N\eta} + E^{-1}N^{-1/2} \right)^2$$

where we used (2.6)

$$\frac{G_{ji}}{G_{ii}} = zG_{jj}^{(i)} \langle \mathbf{x}_i | \mathcal{G}^{(i,j)} | \mathbf{x}_j \rangle \leqslant E^{-1}N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta} \tag{4.28}$$

Using these bounds,

$$G_{ii} = \frac{1}{-1-z-d-zdm} + E^{-1}O(m^{(i)}-m) + E^{-2}O(Z_i)$$

Summing up G_{ii}

$$m = \frac{1}{-1-z-d-zdm} + O(E^{-2})\left(\frac{\varphi^{C_\zeta}}{N\eta} + E^{-1}N^{-1/2}\right)^2 + O(E^{-2})[Z] \tag{4.29}$$

Since the real part of $-1-z-d-zdm$ is much larger than its image part, then,

$$\Im \frac{1}{1-z-d-zdm} \leqslant CE^{-2}\eta + \frac{1}{2}\Im m \tag{4.30}$$

Together with (4.29)

$$\Im m \leqslant CE^{-2}\eta + E^{-1}\left(\frac{\varphi^{C_\zeta}}{N\eta} + E^{-1}N^{-1/2}\right) = \left(\frac{N\eta^2}{E^2} + \frac{\eta N^{1/2}}{E^2} + \frac{\varphi^{C_\zeta}}{E}\right)\frac{1}{N\eta} \tag{4.31}$$

If $E \geqslant N^\varepsilon$ for some $\varepsilon > 0$, we have

$$\Im m \ll \frac{1}{N\eta} \tag{4.32}$$

With (4.9), it implies that there is no eigenvalues locating in $[E-\eta, E+\eta]$ holds with ζ-high probability, i.e., there is no eigenvalues locating larger than N^ε.

Now, it only remains to prove (4.32) for $K\lambda_+ \leqslant E \leqslant N^\varepsilon$. Around (4.32), we have proved that $\max_j \lambda_j \leqslant N^\varepsilon$, then

$$|G_{ii}| \geqslant N^{-2\varepsilon}$$

Therefore, applying (3.21) and (3.19) with choosing $X = N^\varepsilon\left(N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta}\right)$, $Y = N^2\varepsilon$ and $p = N^\varepsilon$, by using (4.24), (4.27), (4.26), (4.28), we have

$$O[Z] \leqslant N^{C\varepsilon}\left(N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta}\right)^2$$

Inserting it in (4.29), with (4.25), (4.30), we obtain that (4.32). Again with (4.9), it implies that there is no eigenvalues locating in $[K\lambda_+, N^\varepsilon]$. □

Step four of proof on (1.19): Now we prove the last component of the proof for (1.19), i.e., in the case of $d > 1$, i.e, $N > M$, we have $\lambda_M \geqslant \lambda_-/5$, Since in previous proof, 5 is not a special number, then it only remains to prove that for some large K, the following bound holds with ζ-high probability,

$$\lambda_j \geqslant \lambda_-/K. \tag{4.33}$$

Let

$$z = E + i\eta, \quad 0 \leqslant E \leqslant \lambda_-/K, \quad \eta = N^{-1/2-\varepsilon} \tag{4.34}$$

for some small enough $\varepsilon > 0$. Since we have proved that among λ_i, $i \leqslant M$, there are at least $\varphi^{O(1)}$ eigenvalues less than λ_-/K, then for some $C, c \geqslant 0$ (Here $\frac{\varphi^{C_\zeta}}{N\eta}$ is contributed by these $\varphi^{O(1)}$ eigenvalues)

$$\Im \frac{1}{N} \operatorname{Tr} \mathcal{G}(z) \leqslant C\eta + \frac{\varphi^{C_\zeta}}{N\eta}, \quad c - \frac{\varphi^{C_\zeta}}{N\eta} \leqslant \operatorname{Re} \frac{1}{N} \operatorname{Tr} \mathcal{G}(z) \leqslant C + \frac{\varphi^{C_\zeta}}{N\eta} \tag{4.35}$$

so as $\mathcal{G}^{(\mathbb{T})}$ for $|\mathbb{T}| = O(1)$. Then using Lemma 1.4,

$$|Z_i| \leqslant |z| \left(N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta} \right) \leqslant |z| N^{-1/2+2\varepsilon}, \quad \langle \mathbf{x}_i | \mathcal{G}^{(i,j)} | \mathbf{x}_j \rangle \leqslant N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta} \leqslant N^{-1/2+2\varepsilon} \tag{4.36}$$

First we estimate G_{ii}, with (2.5), we obtain,

$$G_{ii} = \left(-z - zd \frac{1}{N} \operatorname{Tr} \mathcal{G}^{(i)}(z) + Z_i \right)^{-1} \tag{4.37}$$

Then using (4.35) we have

$$c|z|^{-1} \leqslant |G_{ii}| \leqslant C|z|^{-1} \tag{4.38}$$

Similarly with (2.6), we have

$$|G_{ij}| \leqslant |z|^{-1} N^{-1/2+C\varepsilon} \tag{4.39}$$

As (2.3), we have

$$\operatorname{Tr} G^{(i)}(z) - \operatorname{Tr} \mathcal{G}^{(i)}(z) = \frac{M - N + 1}{z} = \operatorname{Tr} G(z) - \operatorname{Tr} \mathcal{G}(z) + \frac{1}{z} \, .$$

Together with (4.37),

$$G_{ii} = \left(-z - zd\frac{1}{N}\operatorname{Tr}\mathcal{G}(z) - zd\left(m^{(i)} - m - \frac{1}{Nz}\right) + Z_i\right)^{-1}$$

Using the fact: $c|z| \leqslant |-z - zd\frac{1}{N}\operatorname{Tr}\mathcal{G}(z)| \leqslant C|z|$, (4.36) and $|m^{(i)} - m| \leqslant (N\eta)^{-1}$, we can sum up G_{ii}, with Taylor expansion, and obtain

$$m = \frac{1}{1 - z - d - zdm(z)} + \delta \tag{4.40}$$

$$\delta := |z|^{-1}O\left(\sum_i (m^{(i)} - m) + (Nz)^{-1}\right) + |z|^{-2}O([Z]) + |z|^{-1}N^{-1+C\varepsilon}$$

Similarly, by estimating $G_{ii} - G_{jj}$, we have

$$|G_{ii} - m| \leqslant |z|^{-1}N^{-1/2+C\varepsilon}$$

First we estimate $m^{(i)} - m$ in (4.40),

$$m^{(i)} - m = \frac{-1}{N}\sum_j \frac{G_{ji}G_{ij}}{G_{ii}} = \frac{-1}{N}\frac{[G^2]_{ii}}{G_{ii}} = \frac{-1}{N}\frac{[G^2]_{ii}}{m} + O(|z|N^{-3/2+C\varepsilon})\left|[G^2]_{ii}\right|$$

Averaging $m^{(i)} - m$, we obtain that

$$\frac{1}{N}\sum_i \left(m^{(i)} - m\right) = \frac{-1}{N^2}\frac{\operatorname{Tr}[G^2]}{m} + O(|z|N^{-5/2+C\varepsilon})\sum_i \left|[G^2]_{ii}\right| \tag{4.41}$$

Since we have proved that there are at least $\varphi^{O(1)}$ non-zero eigenvalues less than λ_-/K, then

$$\operatorname{Tr}[G^2] = \sum_\alpha \frac{1}{(\lambda_\alpha - z)^2} = \frac{N - M}{z^2} + O(\varphi^{C_\zeta})\eta^{-2} + O(N) \tag{4.42}$$

These three terms come from zero eigenvalues, small eigenvalues and normal eigenvalues, i.e., around $[\lambda_-, \lambda_+]$ respectively. We denote these three parts as T_0, T_s and T_n. Similarly, we have

$$Nm = \operatorname{Tr}[G] = \frac{N - M}{-z} + O(\varphi^{C_\zeta})\eta^{-1} + O(N) = \frac{N - M}{-z}\left(1 + O\left(\frac{z}{N\eta}\right) + O(z)\right) \tag{4.43}$$

(Note: here $z \leqslant O(1)$ is small enough) and

$$\left|(G^2)_{ii}\right| \leqslant \left|\sum_\alpha \frac{|u_\alpha(i)|^2}{(\lambda_\alpha - z)^2}\right| = C\sum_{\alpha\in T_0}\frac{|u_\alpha(i)|^2}{|z^2|} + C\sum_{\alpha\in T_s}\frac{|u_\alpha(i)|^2}{\eta^2} + C\sum_{\alpha\in T_n}|u_\alpha(i)|^2$$

The last one implies that

$$\sum_i \left|(G^2)_{ii}\right| \leqslant C\frac{N}{|z|^2} + O(\varphi^{C_\zeta})\eta^{-2} + O(N)$$

Together with (4.41), we have

$$\frac{1}{N}\sum_i \left(m^{(i)} - m\right) = \frac{-1}{N^2}\frac{\mathrm{Tr}[G^2]}{m} + O(|z|^{-1}N^{-3/2+C\varepsilon}) \tag{4.44}$$

Using (4.42) and (4.43), we have

$$\frac{\mathrm{Tr}[G^2]}{Nm} = \frac{-1}{z} + O(zN^{2\varepsilon}) + O(\frac{1}{N\eta}) + O(1) \tag{4.45}$$

Now combining (4.44) and (4.45) with (4.40), we obtain

$$\delta \leqslant |z|^{-1}O\left(|z^{-1}|N^{-3/2+C\varepsilon} + N^{-1+C\varepsilon}\right) + |z|^{-2}O([Z]) \tag{4.46}$$

Now we apply Lemma 3.3 to estimate $[Z]$, with choosing $X = N^{-1/2+C\varepsilon}$, $Y = C|z|$ and $p = N^\varepsilon$, using (4.36), (4.38) and (4.39), we have

$$|z|^{-2}|[Z]| \leqslant |z|^{-1}N^{-1+C\varepsilon}$$

Together with (4.46), we have obtained

$$\delta \leqslant |z|^{-1}O\left(|z^{-1}|N^{-3/2+C\varepsilon} + N^{-1+C\varepsilon}\right) \tag{4.47}$$

We note if $m = \frac{1}{1-z-d-zdm(z)} + \delta$, then

$$m - m_W = \frac{1}{1-z-d-zdm(z)} - \frac{1}{1-z-d-zdm_W(z)} + \delta$$

which implies that

$$\left(\frac{zd}{(1-z-d-zdm(z))(1-z-d-zdm_W(z))} - 1\right)(m - m_W) = \delta$$

As above, we have $c|z| \leqslant |1-z-d-zdm(z)|$, $|1-z-d-zdm_W(z)| \leqslant C|z|$ for small enough $z \leqslant O(1)$. Therefore, we have

$$|m - m_W| \leqslant |z\delta|$$

Using (4.47), we have

$$|m - m_W| \leqslant O\left(|z^{-1}|N^{-3/2+C\varepsilon} + N^{-1+C\varepsilon}\right) \ll (N\eta)^{-1}$$

Furthermore, it is easy to prove that

$$\Im\left(m_W - \frac{1-d^{-1}}{-z}\right) = O(\eta) \ll (N\eta)^{-1}$$

Together with $\operatorname{Tr} G = \operatorname{Tr} \mathcal{G} - z^{-1}(N-M)$, we have

$$\Im \operatorname{Tr} \mathcal{G}(z) \ll \frac{1}{\eta},$$

As (4.9), we have $\lambda_\alpha \notin [E-\eta, E+\eta]$ for $E \in [0, \lambda_-/K]$ with large enough $K = O(1)$ and complete the proof of (4.33) and complete the proof of Theorem 1.2.

□

4.2. *Proof of Theorem 1.3.* First, we prove (1.22). Recall (4.10) and the fact that there is no eigenvalue in $(0, \lambda_-/4] \cup [4\lambda_+, +\infty]$, We have that

$$\max_{E\in\mathbb{R}} \left|\mathfrak{n}(E) - n_W(E)\right| \leqslant \frac{C(\log N)\varphi^{C_\zeta}}{N}, \tag{4.48}$$

holds with ζ-high probability. The supremium over E is a standard argument for extremely small events and we omit the details.

Now we turn to the proof of (1.21). The proof is very similarly to the one for generalized Wigner matrix in [11]. For reader and self containing, we repeat the argument at here again. By symmetry, we assume that $1 \leqslant j \leqslant N/2$ and let $E = \gamma_j$, $E' = \lambda_j$. Setting $t_N = (\log N)\varphi^{C_\zeta}$ for simplicity, from (4.48) we have

$$n_W(E) = \mathfrak{n}(E') = n_W(E') + O(t_N/N). \tag{4.49}$$

Clearly $E \geqslant \lambda_C := (\lambda_+ + 3\lambda_-)/4$, and using (4.48) $E' \geqslant \lambda_C$ also holds with an overwhelming probability. First, using (1.19) and

$$n_W(x) \sim (\lambda_+ - x)^{3/2}, \ \text{ for } \ \lambda_C \leqslant x \leqslant \lambda_+, \tag{4.50}$$

i.e.

$$n_W(E)=n_W(\gamma_j)=\frac{j}{N}\sim(\lambda_+-E)^{3/2},$$

we know that (1.21) holds (with a possibly increased power) if

$$E,E'\geqslant\lambda_+-t_NN^{-2/3}.$$

Hence, we can assume that one of E and E' is in the interval $[\lambda_C,\lambda_+-t_NN^{-2/3}]$. With (4.50), this assumption implies that at least one of $n_W(E)$ and $n_W(E')$ is larger than $t_N^{3/2}/N$. Inserting this information into (4.49), we obtain that both $n_W(E)$ and $n_W(E')$ are positive and

$$n_W(E)=n_W(E')\big[1+O(t_N^{-1/2})\big],$$

in particular, $\lambda_+-E\sim\lambda_+-E'$. Using that $n_W'(x)\sim(\lambda_+-x)^{1/2}$ for $\lambda_C\leqslant x\leqslant\lambda_+$, we obtain that $n_W'(E)\sim n_W'(E')$, and in fact $n_W'(E)$ is comparable with $n_W'(E'')$ for any E'' between E and E'. Then with Taylor's expansion, we have

$$|n_W(E')-n_W(E)|\leqslant C|n_W'(E)||E'-E|. \tag{4.51}$$

Since $n_W'(E)=\rho_W(E)\sim\sqrt{\kappa}$ and $n_W(E)\sim\kappa^{3/2}$, moreover, by $E=\gamma_j$ we also have $n_W(E)=j/N$, we obtain from (4.49) and (4.51) that

$$|E'-E|\leqslant\frac{C|n_W(E')-n_W(E)|}{n_W'(E)}\leqslant\frac{Ct_N}{Nn_W'(E)}\leqslant\frac{Ct_N}{N(n_W(E))^{1/3}}\leqslant\frac{Ct_N}{N^{2/3}j^{1/3}},$$

which proves (1.21), again, after increasing power.

□

5. Universality of eigenvalues in Bulk. In this section, we are going to prove Theorem 1.6. As mentioned in the introduction, our arguments are valid for both real and complex valued entries. First, we consider a flow of random matrices X_t satisfying the following matrix valued stochastic differential equation

$$\mathrm{d}X_t=\frac{1}{\sqrt{M}}\mathrm{d}\beta_t-\frac{1}{2}X_t\,\mathrm{d}t, \tag{5.1}$$

where β_t is a real matrix valued process whose elements are standard real valued independent Brownian motions. The initial condition $X_0=X=[x_{ij}]$ satisfying (1.1) and (1.2). For any fixed $t\geqslant0$, the distribution of X_t coincides with that of

$$X_t\overset{d}{=}e^{-t/2}X_0+(1-e^{-t})^{1/2}\,V, \tag{5.2}$$

where V is a real matrix with Gaussian entries which have mean 0 and variance $1/M$. The singular values of the matrix X_t also satisfy a system of coupled SDEs which is also called the Dyson Brownian motion (with a drift in our case). More precisely, let

$$\mu = \mu_N(\mathrm{d}\mathbf{w}) = \frac{e^{-\mathcal{H}_W^\beta(\mathbf{w})}}{Z_\beta}\mathrm{d}\mathbf{w}$$
$$\mathcal{H}_W^\beta(\mathbf{w}) = \beta\left[\sum_{i=1}^N \frac{w_i^2}{2d} - \frac{1}{N}\sum_{i<j}\log|w_j^2 - w_i^2| - \Big(\frac{1}{d} - 1 + \frac{1-\beta^{-1}}{N}\Big)\sum_{i=1}^N \log|w_i|\right], \tag{5.3}$$

denote the joint distribution of the singular values of X when the matrix X has independent Gaussian entries (so that $X^\dagger X$ is a Wishart random matrix). In Equation 5.3, the constant β takes values $\{1,2\}$ with $\beta = 2$ for complex entries and $\beta = 1$ for real valued entries. Also, Z_β is the normalization constant so that μ is probability measure. Denote the distribution of the singular values at time t by $f_t(\mathbf{w})\mu(\mathrm{d}\mathbf{w})$. Then f_t satisfies

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

where

$$L^W = L^W_{\beta,N} = \sum_{i=1}^N \frac{1}{2N}\partial_i^2 + \sum_{i=1}^N \left(- \frac{\beta w_i}{2d} + \frac{\beta}{N}\sum_{j\neq i}\frac{w_i}{w_i^2 - w_j^2} + \frac{1}{2}\Big(\beta\big(\frac{1}{d} - 1\big) + \frac{\beta - 1}{N}\Big)\frac{1}{w_j}\right)\partial_i\,. \tag{5.5}$$

For any $n \geqslant 1$ we define the n-point correlation functions (marginals) of the probability measure $f_t\mathrm{d}\mu$ by

$$p_{t,N}^{(n)}(w_1, w_2, \ldots, w_n) = \int_{\mathbb{R}^{N-n}} f_t(\mathbf{w})\mu(\mathbf{w})\mathrm{d}w_{n+1}\ldots\mathrm{d}w_N. \tag{5.6}$$

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

$$p_{\mu,N}^{(n)}(w_1, w_2, \ldots, w_n) = \int_{\mathbb{R}^{N-n}} \mu(\mathbf{w})\mathrm{d}w_{n+1}\ldots\mathrm{d}w_N. \tag{5.7}$$

Now we are ready to prove the *strong local ergodicity of the Dyson Brownian motion* which states that the correlation functions of the Dyson Browian motion $p_{t,N}^{(n)}$ and those of the equilibrium measure $p_{\mu,N}^{(n)}$ are close:

THEOREM 5.1. *Let $X = [x_{ij}]$ with the entries x_{ij} satisfying (1.1) and (1.2). Let $E \in [\lambda_- + c, \lambda_+ - c]$ with some $c > 0$. Then for any $\varepsilon' > 0$, $\delta > 0$, $0 < b = b_N < c/2$, any integer $n \geqslant 1$ and for any compactly supported continuous test function $O : \mathbb{R}^n \to \mathbb{R}$ we have*

$$\sup_{t \geqslant N^{-1+\delta+\varepsilon'}} \Big| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int_{\mathbb{R}^n} \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_n \, O(\alpha_1, \ldots, \alpha_n) \frac{1}{\varrho_W(E)^n} \\ \times \Big(p^{(n)}_{t,N} - p^{(n)}_{\mu,N}\Big)\Big(E' + \frac{\alpha_1}{N\varrho_W(E)}, \ldots, E' + \frac{\alpha_n}{N\varrho_W(E)}\Big)\Big| \leqslant C_n N^{2\varepsilon'} \Big[b^{-1}N^{-1+\varepsilon'} + b^{-1/2}N^{-\delta/2}\Big], \tag{5.8}$$

where $p^{(n)}_{t,N}$ and $p^{(n)}_{\mu,N}$, (5.6)–(5.7), are the correlation functions of the eigenvalues of the Dyson Brownian motion flow (5.2) and those of the equilibrium measure respectively and C_n is a constant.

REMARK 5.2. *Notice that if we choose $\delta = 1 - 2\varepsilon'$ and thus $t = N^{-\varepsilon'}$, then we can set $b \sim N^{-1+8\varepsilon'}$ so that the right hand side of (5.8) vanishes as $N \to \infty$. From the MP law we know that the spacing of the eigenvalues in the bulk is $O(N^{-1})$ and thus we see that Theorem 5.1 yields universality with essentially no averaging in E.*

PROOF OF THEOREM 5.1. The proof follows from the main result in [8] (Theorem 2.1) which states that the local ergodicity of Dyson Brownian motion ((5.8)) holds for $t \geqslant N^{-2\mathfrak{a}+\delta}$ for any $\delta > 0$ provided that there exists an $\mathfrak{a} > 0$ such that

$$\sup_{t \geqslant N^{-2\mathfrak{a}}} \frac{1}{N} \mathbb{E} \sum_{j=1}^{N} (\lambda_j(t) - \gamma_j)^2 \leqslant CN^{-1-2\mathfrak{a}} \tag{5.9}$$

holds with a constant C uniformly in N. Here $\sqrt{\lambda_j(t)}$ is the singular value of the matrix X_t given in (5.2). The condition (5.9) is a simple consequence of (1.21) as long as $\mathfrak{a} < 1/2$.

Strictly speaking, there are four assumptions in the hypothesis of Theorem 2.1 in [8]. Assumptions I and II of Theorem 2.1 in [8] are automatically satisfied in the setting that the Dyson Brownian motion is generated by flows on the Covariance matrix ensembles. Assumption IV of Theorem of [8] states that the local density of the singular values of X_t in the scale larger than N^{-1+c} is bounded above by a constant. As in [8] this follows from the large deviation estimate (1.17) since a bound on $\Im m(z)$, $z = E + i\eta$, can be easily used to prove an upper bound on the local density of eigenvalues in a window of size η about E. As usual, the additional condition in [8] on the entropy $S_\mu(f_{t_0}) \leqslant CN^m$ for some constant m for $t_0 = N^{-2\mathfrak{a}}$, holds due to the regularization property of the Ornstein-Uhlenbeck process. Thus for a given $0 < \varepsilon' < 1$, choosing $\mathfrak{a} = 1/2 - \varepsilon'/2$, $A = \varepsilon'$ in the second part of Theorem 2.1 in [8] and using (1.21), we obtain (5.9) and the proof is finished. □

For any $\varepsilon > 0$, applying Theorem 5.1 with $\delta = 1 - 2\varepsilon$, $\varepsilon' = \varepsilon$ and $b = -1 + 8\varepsilon$, we obtain universality for all ensembles with the matrix elements distributed according to $M^{-1/2}\xi_t$ with

$$\xi_t = e^{-t/2}\xi_0 + (1 - e^{-t})^{1/2}\,\xi_G, \tag{5.10}$$

where the matrix ξ_G has independent Gaussian random variables with mean 0 and variance 1 and $t \sim N^{-\varepsilon}$ and the initial condition ξ_0 has entries satisfying our conditions (1.1) and (1.2). In other words, for $t \sim N^{-\varepsilon}$ the random matrices ξ_t which are distributed according to (5.10) have the same correlation functions as that of the matrix with Gaussian entries, averaged on a length of $O(N^{-1+8\varepsilon})$. Thus in order to prove Theorem 1.6, it remains to find a random matrix $\widetilde{\xi}_t$ of the form (5.10) (with time $t = N^{-\varepsilon}$) whose eigenvalue correlation functions well approximate that of the spectrum of the given matrix X satisfying (1.1) and (1.2).

The requirements on the entries of the matrix $\tilde{\xi}_t$ are just mean zero, variance one and subexponential decay; however it turns out that for any fixed X and ε, one may find a $\widetilde{\xi}_0$ such that $\widetilde{\xi}_t$ satisfies (5.10), with $t \sim N^{-\varepsilon}$, and the entries $(\widetilde{\xi}_t)_{ij}$ have mean 0, variance 1 and the *same* third moment as those of the initial condition X. Moreover $\widetilde{\xi}_t$ can be chosen in such a way so that its entries have fourth moment very close to those of X. More precisely, Lemma 3.4 in [10] yields that for any given matrix X satisfying (1.1) and (1.2) and $t \sim N^{-\varepsilon}$, there exists a matrix $\widetilde{\xi}_t$ of the form (5.10) such that for $1 \leqslant k \leqslant 3$,

$$\mathbb{E}\,\sqrt{M}x_{ij}^k = \mathbb{E}\,(\widetilde{\xi}_t)_{ij}^k, \qquad \Big|\mathbb{E}\,(\sqrt{M}x_{ij})^4 - \mathbb{E}\,(\widetilde{\xi}_t)_{ij}^4\Big| \leqslant Ct \sim N^{-\varepsilon}\;.$$

Now to finish the proof of Theorem 1.6, it only remains to show that that the correlation functions of the eigenvalues of two matrix ensembles at a fixed energy (*i.e.*, for a fixed value of $E = \Re(z)$) are identical up to the scale $1/N$ provided that the first four moments of the matrix elements of these two ensembles are almost identical. To achieve this, as shown for the Wigner matrices [9] (see Section (8.6-8.13) of [9]), it is enough to show that the corresponding Green functions are close for these two matrix ensembles. This is the content of the following theorem which we call, following [9], the Green function comparison theorem.

Let $X^{\mathbf{v}} = [x_{ij}^{\mathbf{v}}]$, with the entries $x_{ij}^{\mathbf{v}}$ satisfying (1.1) and (1.2), and let $G^{\mathbf{v}}(z) = (X^{\mathbf{v}\dagger}X^{\mathbf{v}} - z)^{-1} = (H^{\mathbf{v}} - z)^{-1}$ be the Green function corresponding to $X^{\mathbf{v}}$. Define the matrix $X^{\mathbf{w}}$ and the Green function $G^{\mathbf{w}}(z)$ analogously.

Theorem 5.3. *Assume that the first three moments of $x_{ij}^{\mathbf{v}}$ and $x_{ij}^{\mathbf{w}}$ are identical, i.e.,*

$$\mathbb{E}(x_{ij}^{\mathbf{v}})^u = \mathbb{E}(x_{ij}^{\mathbf{w}})^u, \qquad 0 \leqslant u \leqslant 3,$$

and the difference between the fourth moments of $x_{ij}^{\mathbf{v}}$ and $x_{ij}^{\mathbf{w}}$ is much less than 1, say

$$\Big|\mathbb{E}\,(\sqrt{M}x_{ij}^{\mathbf{v}})^4 - \mathbb{E}\,(\sqrt{M}x_{ij}^{\mathbf{w}})^4\Big| \leqslant N^{-\delta}, \tag{5.11}$$

for some given $\delta > 0$. Let $\varepsilon > 0$ be arbitrary and choose an η with $N^{-1-\varepsilon} \leqslant \eta \leqslant N^{-1}$. For any sequence of positive integers $k_1, \ldots, k_n$, set complex parameters

$$z_j^m = E_j^m \pm i\eta, \quad j = 1, \ldots k_i, \quad m = 1, \ldots, n$$

with an arbitrary choice of the $\pm$ signs and $\lambda_- + \kappa \leqslant |E_j^m| \leqslant \lambda_+ - \kappa$ for some $c > 0$. Let $F(x_1, \ldots, x_n)$ be a function such that for any multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $1 \leqslant |\alpha| \leqslant 5$ and for any $\varepsilon' > 0$ sufficiently small, we have

$$\max\left\{|\partial^\alpha F(x_1, \ldots, x_n)| : \max_j |x_j| \leqslant N^{\varepsilon'}\right\} \leqslant N^{C_0 \varepsilon'} \tag{5.12}$$

$$\max\left\{|\partial^\alpha F(x_1, \ldots, x_n)| : \max_j |x_j| \leqslant N^2\right\} \leqslant N^{C_0} \tag{5.13}$$

for some constant C_0.

Then, there is a constant C_1, depending on α, $\sum_i k_i$ and C_0 such that for any η with $N^{-1-\varepsilon} \leqslant \eta \leqslant N^{-1}$ and for any choices of the signs in the imaginary part of z_j^m

$$\left|\mathbb{E}F\left(\frac{1}{N^{k_1}} \operatorname{Tr}\left[\prod_{j=1}^{k_1} G^{\mathbf{v}}(z_j^1)\right], \ldots, \frac{1}{N^{k_n}} \operatorname{Tr}\left[\prod_{j=1}^{k_n} G^{\mathbf{v}}(z_j^n)\right]\right) - \mathbb{E}F\left(G^{\mathbf{v}} \to G^{\mathbf{w}}\right)\right|$$
$$\leqslant C_1 N^{-1/2 + C_1\varepsilon} + C_1 N^{-\delta + C_1 \varepsilon}, \tag{5.14}$$

where in the second term the arguments of F are changed from the Green functions of $H^{\mathbf{v}}$ to $H^{\mathbf{w}}$ and all other parameters remain unchanged.

Once again we note the equivalence of (5.8) and (6.14) as discussed in [9] (Sections 8.6-8.13). The only difference is that in [9], the equivalence is proved for Wigner matrices, but the arguments are easily adapted for covariance matrices. Thus to complete the proof of Theorem 1.6, all that remains to be done is the proof of Theorem 5.3 which we give below.

Proof of Theorem 5.3. The proof is very similar to Lemma 2.3 of [9]. The only differences are a few simple linear algebraic identities. Therefore, we will only prove the simple case of $k = 1$ and $n = 1$.

Fix a bijective ordering map on the index set of the independent matrix elements,

$$\phi : \{(i, j) : 1 \leqslant i \leqslant M, 1 \leqslant j \leqslant N\} \to \Big\{1, \ldots, MN\Big\},$$

and define the family of random matrices X_γ, $0 \leqslant \gamma \leqslant MN$,

$$[X_\gamma]_{ij} = [X^{\mathbf{v}}]_{ij} \quad \phi(i, j) > \gamma,$$

$$= [X^{\mathbf{w}}]_{ij} \quad \phi(i,j) \leqslant \gamma \ .$$

In particular we have $X_0 = X^{\mathbf{v}}$ and $X_{MN} = X^{\mathbf{w}}$. Denote H_γ, G_γ and $\mathcal{G}_\gamma$ as

$$H_\gamma = X_\gamma^\dagger X_\gamma, \quad G_\gamma = (H_\gamma - z)^{-1}, \quad \mathcal{G}_\gamma = (X_\gamma X_\gamma^\dagger - z)^{-1} \ .$$

First, using the delocalization result (1.20) and the rigidity of eigenvalues (1.21), it is easy to have the following estimate on the matrix elements of the resolvent:

$$\max_\gamma \max_{k\, l} \max_{\eta \geqslant N^{-1-\varepsilon}} \max_{\kappa \geqslant c} \left|[G_\gamma(z)]_{kl}\right| + \left|[\mathcal{G}_\gamma(z)]_{kl}\right| \leqslant N^{C\varepsilon} \tag{5.15}$$

holds with ζ-high probability for any $\zeta = O(1)$. For instance, for $\gamma = 0$, we have the identity $G_0(z) = \sum_{\alpha=1}^N \frac{\mathbf{v}_\alpha{}^\dagger \mathbf{v}_\alpha}{\lambda_\alpha - z}$ where $\lambda_\alpha, \mathbf{v}_\alpha$ are the eigenvalues and eigenvectors of H_0. By the delocalisation result (1.20) we obtain

$$|G_0(z)| \leqslant \frac{\varphi^{C_\zeta}}{N} \sum_{\alpha=1}^N \frac{1}{|\lambda_\alpha - z|} \ .$$

We write the above sum as

$$\sum_\alpha \frac{1}{|\lambda_\alpha - z|} = \sum_k \sum_{\alpha \in I_k} \frac{1}{|\lambda_\alpha - z|} \leqslant \sum_k |I_k| \frac{1}{|\lambda_\alpha - E| + \eta} \ , \tag{5.16}$$

where I_k is the set that

$$N^{-1} 2^{K-1} \leqslant (\lambda_\alpha - E) \leqslant N^{-1} 2^K \ .$$

By the rigidity of eigenvalues we obtain that $|I_K| \leqslant C 2^K$ with ζ-high probability. Substituting this bound in (5.16) yields the estimate (5.15).

For $1 \leqslant i \leqslant N$, it is easy to check that

$$\mathcal{G}_{kl}^{(i)} = \mathcal{G}_{kl} + \frac{(\mathcal{G}\mathbf{x}_i)_k\, (\mathbf{x}_i^\dagger \mathcal{G})_l}{1 - \langle \mathbf{x}_i, \mathcal{G}(z)\, \mathbf{x}_i \rangle}, \quad \mathcal{G}_{kl} = \mathcal{G}_{kl}^{(i)} - \frac{(\mathcal{G}^{(i)}\mathbf{x}_i)_k (\mathbf{x}_i^\dagger \mathcal{G}^{(i)})_l}{1 + \langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\, \mathbf{x}_i \rangle} \ . \tag{5.17}$$

With (2.5), we have

$$\langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\, \mathbf{x}_i \rangle = -1 + \frac{-1}{z\, G_{ii}}, \quad \langle \mathbf{x}_i, \mathcal{G}(z)\, \mathbf{x}_i \rangle = 1 + z\, G_{ii} \tag{5.18}$$

$$\mathcal{G}\mathbf{x}_i = \frac{\mathcal{G}^{(i)}\mathbf{x}_i}{1 + \langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\mathbf{x}_i \rangle} = -z\, G_{ii}\, \mathcal{G}^{(i)}\, \mathbf{x}_i \ . \tag{5.19}$$

Furthermore, with (2.6), we have

$$\langle \mathbf{x}_i, \mathcal{G}^{(i)} \mathbf{x}_j \rangle = \langle \mathbf{x}_i, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle - \frac{\langle \mathbf{x}_i, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle \langle \mathbf{x}_j, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle}{1 + \langle \mathbf{x}_j, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle}$$
$$= \frac{\langle \mathbf{x}_i, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle}{1 + \langle \mathbf{x}_j, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle} = -z G_{jj}^{(i)} \langle \mathbf{x}_i, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle = -\frac{G_{ij}}{G_{ii}}$$

and similarly

$$\langle \mathbf{x}_i, \mathcal{G} \mathbf{x}_j \rangle = -z G_{ii} \langle \mathbf{x}_i, \mathcal{G}^{(i)} \mathbf{x}_j \rangle = z G_{ij} \tag{5.20}$$

which implies that

$$\langle \mathbf{x}_i, \mathcal{G}^{(i)} \, \mathbf{x}_j \rangle = \frac{G_{ij}}{G_{ii}}, \quad \langle \mathbf{x}_i, \mathcal{G} \mathbf{x}_j \rangle = -z G_{ij} \; . \tag{5.21}$$

Let x_i be the i^{th} row of X (recall $\mathbf{x}_i$ is the i th column of X). By symmetry, the above identities also hold if one switches G and $\mathcal{G}$, $\mathbf{x}_i$ and x_i.

Combining the above identities with (5.15), we obtain

$$\max_{\gamma} \max_{k\,l} \max_{\eta \geqslant N^{-1-\varepsilon}} \max_{|\kappa| \geqslant c} \left| [G_\gamma(z)]_{kl} \right| + \left| [X_\gamma G_\gamma(z)]_{kl} \right| + \left| \left[G_\gamma X_\gamma^\dagger(z) \right]_{kl} \right| + \left| \left[X_\gamma G_\gamma X_\gamma^\dagger(z) \right]_{kl} \right| \leqslant N^{C\varepsilon} \; . \tag{5.22}$$

Consider the telescopic sum of differences of expectations

$$\mathbb{E}\, F\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H^{\mathbf{w}} - z} \right) - \mathbb{E}\, F\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H^{\mathbf{v}} - z} \right)$$
$$= \sum_{\gamma=1}^{MN} \left[\mathbb{E}\, F\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H_\gamma - z} \right) - \mathbb{E}\, F\left(\frac{1}{N} \operatorname{Tr} \frac{1}{H_{\gamma-1} - z} \right) \right]. \tag{5.23}$$

Let $E^{(ij)}$ denote the matrix whose matrix elements are zero everywhere except at the (i,j) position, where it is 1, i.e., $E^{(ij)}_{k\ell} = \delta_{ik}\delta_{j\ell}$. Fix an $\gamma \geqslant 1$ and let (i,j) be determined by $\phi(i,j) = \gamma$. We will compare $H_{\gamma-1}$ with H_γ. Note that these two matrices differ only in the (i,j) matrix element and they can be written as

$$X_{\gamma-1} = Q + V, \qquad V := x_{ij}^{\mathbf{v}} \, E^{(ij)}, \qquad X_\gamma = Q + W, \qquad W := x_{ij}^{\mathbf{w}} \, E^{(ij)}$$

with a matrix Q that has zero matrix element at the (i,j) position. Define the Green functions

$$R = \frac{1}{Q^\dagger Q - z}, \qquad S = \frac{1}{H_{\gamma-1} - z}, \qquad T = \frac{1}{H_\gamma - z} \; .$$

The following lemma is at the heart of the Green function comparison first established in [9] (then also used in [10, 11, 5]) which states that the difference of smooth functionals of Green functions of two matrices which differ from a single entry can be bounded above as a function of its first four moments.

Lemma 5.4. *Let m_k be the k^{th} moment of $\sqrt{M}x^{\mathbf{v}}_{ij}$, then*

$$\mathbb{E}\left[F\left(\frac{1}{N}\operatorname{Tr}S\right)-F\left(\frac{1}{N}\operatorname{Tr}R\right)\right]=A(Q,m_1,m_2,m_3)+N^{-5/2+C\varepsilon}+\widetilde{A}(Q)m_4 \tag{5.24}$$

for a functional $A(Q,m_1,m_2,m_3)$ which only depends on the distribution of Q and m_1,m_2,m_3. The constant $\widetilde{A}(Q)$ depends only on the distribution of Q and satisfies the bound

$$|\widetilde{A}(Q)|\leqslant N^{-2+C\varepsilon}\,.$$

Before giving the proof of Lemma 5.4, let us it to conclude the forgoing argument in the proof of Theorem 5.3. Note that the matrices H_γ and Q also differ by one entry, and therefore applying Lemma 5.4 yields

$$\mathbb{E}\left[F\left(\frac{1}{N}\operatorname{Tr}T\right)-F\left(\frac{1}{N}\operatorname{Tr}R\right)\right]=A(Q,m_1,m_2,m_3)+N^{-5/2+C\varepsilon}+\widetilde{A}(Q)m_4' \tag{5.25}$$

where m_4' is the fourth moment of $\sqrt{M}x^{\mathbf{w}}_{ij}$ (by hypothesis, the first three moments of $x^{\mathbf{w}}_{ij}$ are identical to those of $x^{\mathbf{v}}_{ij}$.) Since $|m_4'-m_4|\leqslant N^{-\delta}$ by hypothesis, we have

$$\mathbb{E}\,F\left(\frac{1}{N}\operatorname{Tr}\frac{1}{H_\gamma-z}\right)-\mathbb{E}\,F\left(\frac{1}{N}\operatorname{Tr}\frac{1}{H_{\gamma-1}-z}\right)\leqslant CN^{-5/2+C\varepsilon}+CN^{-2-\delta+C\varepsilon}.$$

After summing up in (5.23) we have thus proved that

$$\mathbb{E}\,F\left(\frac{1}{N}\operatorname{Tr}\frac{1}{H^{\mathbf{v}}-z}\right)-\mathbb{E}\,F\left(\frac{1}{N}\operatorname{Tr}\frac{1}{H^{\mathbf{w}}-z}\right)\leqslant CN^{-1/2+C\varepsilon}+CN^{-\delta+C\varepsilon},$$

obtaining precisely what we set out to show in (5.14). The proof can be easily generalized to functions of several variables. Thus to conclude the proof of of Theorem 5.3, we just need to give the proof of Lemma 5.4.

Proof of Lemma 5.4. We first claim that the estimate (5.15) holds for the Green function R as well. To see this, we have, from the resolvent expansion,

$$R=S+S(V^\dagger X+X^\dagger V+V^\dagger V)S+\ldots+[S(V^\dagger X+X^\dagger V+V^\dagger V)]^9S+[S(V^\dagger X+X^\dagger V+V^\dagger V)]^{10}R\,.$$

Since the matrix V has only at most one non-zero entry, when computing the (k,ℓ) matrix element of this matrix identity, each term is a finite sum involving matrix elements of S, XS, $SX^\dagger$, $XSX^\dagger$ or R (only for the last term) and $x^{\mathbf{v}}_{ij}$. Using the bound (5.22) for the S matrix elements, the subexponential decay for $x^{\mathbf{v}}_{ij}$ and the trivial bound $|R_{ij}| \leqslant \eta^{-1}$, we obtain that the estimate (5.15) holds for R. Similarly by expanding XR, RX and XRX, we can obtain (5.22) for XR, RX and XRX, QR, RQ and QRQ.

Now we prove (5.24). By the resolvent expansion,

$$S = R - R(V^\dagger Q + Q^\dagger V + V^\dagger V)R + \ldots - [R(V^\dagger Q + Q^\dagger V + V^\dagger V)]^9 R + O(N^{-4}) \quad (5.26)$$

holds with extremely high probability. Thus we may write

$$\frac{1}{N}\operatorname{Tr} S = \frac{1}{N}\operatorname{Tr} R + \sum_{k\leqslant 20} y_k + O(N^{-4}),$$

where y_k is the sum of the terms in (5.26), in which there are exactly k V's. Recall m_k as the k-th moment of $\sqrt{M}x_{ij}$, which is order one if $k = O(1)$. The terms y_k satisfy the bound (with $K = (k_1, k_2, \cdots, k_n)$ and $|K| := \sum_i k_i$)

$$|y_k| \leqslant N^{C\varepsilon} N^{-k/2}, \quad \mathbb{E}_{\mathbf{v}}\, y_{k_1} y_{k_2} \cdots y_{k_n} = N^{-|K|/2}\, m_{|K|}\, z_K(Q), \quad |z_K(Q)| \leqslant N^{C\varepsilon} \quad (5.27)$$

for some $z_K(Q)$ only depends on the distribution Q and the last inequality holds with with ζ-high probability. Here $\mathbb{E}_{\mathbf{v}}$ is the expectation value with respect to the distribution of the entries of the matrix $X^{\mathbf{v}}$. Then we have

$$\mathbb{E} F\left(\frac{1}{N}\operatorname{Tr}\frac{1}{H_{\gamma-1} - z}\right) = \mathbb{E}\sum_{n=0}^{4}\frac{1}{n!}F^{(n)}\Big(\frac{1}{N}\operatorname{Tr} R\Big)\left(\sum_{k\leqslant 20} y_k\right)^n + O(N^{-5/2+C\varepsilon})\,. \quad (5.28)$$

and with (5.27)

(5.29)

$$\begin{aligned}\mathbb{E} F\left(\frac{1}{N}\operatorname{Tr}\frac{1}{H_{\gamma-1} - z}\right) &= \mathbb{E}\sum_{n=0}^{4}\frac{1}{n!}F^{(n)}\Big(\frac{1}{N}\operatorname{Tr} R\Big)\left(\sum_{k_1,\ldots,k_n} N^{-|K|/2}\, m_{|K|}\, z_K(Q)\right) + O(N^{-5/2+C\varepsilon})\\ &= B + O(N^{-5/2+C\varepsilon}) + A(Q, m_1, m_2, m_3) + \widetilde{A}(Q) m_4\,.\end{aligned}$$

where $A(Q, m_1, m_2, m_3)$ only depends on the distribution of Q and m_1, m_2, m_3 and

$$B = \mathbb{E}\sum_{n=0}^{4}\frac{1}{n!}F^{(n)}\Big(\frac{1}{N}\operatorname{Tr} R\Big)\left(\sum_{k_1,\ldots,k_n : |K|\geqslant 5,\, k_i\leqslant 20} N^{-|K|/2}\, m_{|K|}\, z_K(Q)\right)$$

$$\widetilde{A} = \mathbb{E} \sum_{n=0}^{4} \frac{1}{n!} F^{(n)} \Big(\frac{1}{N} \operatorname{Tr} R \Big) \left(\sum_{k_1,\ldots,k_n : |K|=4} N^{-2} \, z_K(Q) \right)$$

Now it only remains to prove

$$|B| \leqslant O(N^{-5/2+C\varepsilon}), \quad \widetilde{A} \leqslant O(N^{-2+C\varepsilon})$$

Using the estimate (5.22) for R and the derivative bounds (5.12) for the typical values of $\frac{1}{N} \operatorname{Tr} R$, we see that $F^{(n)}\Big(\frac{1}{N} \operatorname{Tr} R\Big)$ $(n \leqslant 4)$ are bounded by $N^{C\varepsilon}$ with ζ-high probability, where C is an explicit constant. Similarly z_K $(k_i \leqslant 20)$ is also bounded by $N^{C\varepsilon}$ for some $C > 0$ with ζ-high probability. Now we define Ξ_g as the good set where these quantities are bounded by $N^{C\varepsilon}$. Furthermore, using (5.13) and definition of z_K, we know that $F^{(n)}\Big(\frac{1}{N} \operatorname{Tr} R\Big)$ and z_K are bounded by N^C for some $C > 0$ in A^c. Since A^c has a very small probability by (5.22), we have

$$\widetilde{A} = \mathbb{E}_{\Xi_g} \sum_{n=0}^{4} \frac{1}{n!} F^{(n)} \Big(\frac{1}{N} \operatorname{Tr} R \Big) \left(\sum_{k_1,\ldots,k_n : |K|=4} N^{-2} \, z_K(Q) \right) + O(N^{-5/2+C\varepsilon}) \, . \tag{5.30}$$

Then with the bounds on $F^{(n)}$ and z_K in Ξ_g, we obtain $\widetilde{A} \leqslant O(N^{-2+C\varepsilon})$. Similarly with $m_{|K|} \leqslant O(1)$, we have $\widetilde{B} \leqslant O(N^{-5/2+C\varepsilon})$ and complete the proof of Lemma 5.4 and thereby also finishing the proof of Theorem 5.3. □

6. Universality of eigenvalues at Edge. In this section, we are going to proof the edge universality stated in Theorem 1.7. The proof is based on Theorem 2.4 of [11] which is an analogous result for Wigner matrices. Here we consider the largest eigenvalue λ_1, but the same argument applies to the lowest non-zero eigenvalue as well. Also for the rest of this section, let us fix a constant $\zeta > 0$.

For any $E_1 \leqslant E_2$ let

$$\mathcal{N}(E_1, E_2) := \#\{E_1 \leqslant \lambda_j \leqslant E_2\}$$

denote the number of eigenvalues of the covariance matrix $\frac{1}{N} X^\dagger X$ in $[E_1, E_2]$ where X is a random matrix whose entries satisfy (1.1) and (1.2). By Theorem 1.2 and 1.3 (rigidity of eigenvalues), there exist positive constants C_ζ such that

$$|\lambda_1 - \lambda_+| \leqslant \varphi^{C_\zeta} N^{-2/3} \tag{6.1}$$

$$\mathcal{N}\left(\lambda_+ - 2\varphi^{C_\zeta} N^{-2/3}, \; \lambda_+ + 2\varphi^{C_\zeta} N^{-2/3}\right) \leqslant \varphi^{2C_\zeta} \tag{6.2}$$

holds with ζ-high probability. Using these estimates, we can assume that s in (1.28) satisfies

$$-\varphi^{C_\zeta} \leqslant s \leqslant \varphi^{C_\zeta} \tag{6.3}$$

Set

$$E_\zeta := \lambda_+ + 2\varphi^{C_\zeta} N^{-2/3} \tag{6.4}$$

and for any $E \leqslant E_\zeta$ define

$$\chi_E := \mathbf{1}_{[E,E_\zeta]}$$

to be the characteristic function of the interval $[E, E_\zeta]$. For any $\eta > 0$ we define

$$\theta_\eta(x) := \frac{\eta}{\pi(x^2+\eta^2)} = \frac{1}{\pi}\Im\frac{1}{x-i\eta} \tag{6.5}$$

to be an approximate delta function on scale η. In the following elementary lemma we compare the sharp counting function $\mathcal{N}(E, E_\zeta) = \operatorname{Tr}\chi_E(H)$ by its approximation smoothed on scale η.

Lemma 6.1. *For any $\varepsilon > 0$, set $\ell_1 := N^{-2/3-3\varepsilon}$ and $\eta := N^{-2/3-9\varepsilon}$. Then there exist constants C, c such that for any E satisfying*

$$|E - \lambda_+| \leqslant \frac{3}{2}\varphi^{C_\zeta} N^{-2/3} \tag{6.6}$$

with the C_ζ in (6.1)-(6.4), we have

$$|\operatorname{Tr}\chi_E(H) - \operatorname{Tr}\chi_E * \theta_\eta(H)| \leqslant C\left(N^{-2\varepsilon} + \mathcal{N}(E-\ell_1, E+\ell_1)\right) \tag{6.7}$$

holds with ζ-high probability.

Proof of Lemma 6.1. From equations (6.13) and (6.17) of [11] we obtain

$$\begin{aligned} |\operatorname{Tr}\chi_E(H) - \operatorname{Tr}\chi_E * \theta_\eta(H)| &\leqslant C\left(\mathcal{N}(E-\ell_1, E+\ell_1) + N^{-5\varepsilon}\right) \\ &\quad + C\, N\eta\,(E_\zeta - E)\int_{\mathbb{R}} \frac{1}{y^2+\ell_1^2}\Im\, m(E - y + i\ell_1)\mathrm{d}y \end{aligned} \tag{6.8}$$

Using the rigidity of eigenvalues (1.21), one can prove that

$$\int_{E-y\geqslant 5\lambda_+} \frac{1}{y^2+\ell_1^2}\,\Im m(E-y+i\ell_1)\,\mathrm{d}y + \int_{E-y\leqslant \lambda_-/5} \frac{1}{y^2+\ell_1^2}\,\Im m(E-y+i\ell_1)\,\mathrm{d}y = O(1)$$

with ζ-high probability. On the interval $\lambda_-/5 \leqslant E-y \leqslant 5\lambda_+$ we use (1.17), *i.e.*,

$$\Im\, m(E-y+i\ell_1) \leqslant \Im\, m_W(E-y+i\ell_1) + \frac{\varphi^{C_\zeta}}{N\ell_1}$$

and the elementary estimate $\Im\, m_W(E-y+i\ell_1) \leqslant C\sqrt{\ell_1 + |E-y-\lambda_+|}$. Using the definitions of ℓ_1 and η it can be shown that (see Equation (6.18) of [11])

$$N\eta\,(E_\zeta - E)\int_{\mathbb{R}} \frac{1}{y^2+\ell_1^2}\,\Im\, m(E-y+i\ell_1)\,\mathrm{d}y \leqslant N^{-2\varepsilon}\,.$$

Now the Lemma follows from (6.8). □

Let $q:\mathbb{R}\to\mathbb{R}_+$ be a smooth cutoff function such that

$$q(x)=1 \quad \text{if} \quad |x|\leqslant 1/9, \qquad q(x)=0 \quad \text{if} \quad |x|\geqslant 2/9,$$

and we assume that $q(x)$ is decreasing for $x\geqslant 0$. Then we have the following corollary for Lemma 6.1:

Corollary 6.2. *Let ℓ_1 be as in Lemma 6.1 and set $\ell := \frac{1}{2}\ell_1 N^{2\varepsilon} = \frac{1}{2}N^{-2/3-\varepsilon}$. Then for all E such that*

$$|E-\lambda_+| \leqslant \varphi^{C_\zeta} N^{-2/3} \tag{6.9}$$

with the C_ζ in (6.1)-(6.4), the following inequality

$$\operatorname{Tr}\chi_{E+\ell} * \theta_\eta(H) - N^{-\varepsilon} \leqslant \mathcal{N}(E,\infty) \leqslant \operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H) + N^{-\varepsilon} \tag{6.10}$$

holds with ζ-high probability. Furthermore, we have

$$\mathbb{E}\, q\left(\operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H)\right) \leqslant \mathbb{P}(\mathcal{N}(E,\infty)=0) \leqslant \mathbb{E}\, q\left(\operatorname{Tr}\chi_{E+\ell} * \theta_\eta(H)\right) + Ce^{-\varphi^{C_\zeta}} \tag{6.11}$$

for sufficiently large N independent of E.

Proof. For any E satisfying (6.9) we have $E_\zeta - E \gg \ell$ thus $|E-\lambda_+-\ell|N^{2/3} \leqslant \frac{3}{2}\varphi^{C_\zeta}$ (see (6.6)), therefore (6.7) holds for E replaced with $y\in[E-\ell,E]$ as well. We thus obtain

$$\begin{aligned}
\operatorname{Tr}\chi_E(H) &\leqslant \ell^{-1}\int_{E-\ell}^{E} \mathrm{d}y\ \operatorname{Tr}\chi_y(H) \\
&\leqslant \ell^{-1}\int_{E-\ell}^{E} \mathrm{d}y\ \operatorname{Tr}\chi_y * \theta_\eta(H) + C\ell^{-1}\int_{E-\ell}^{E}\mathrm{d}y\left[N^{-2\varepsilon} + \mathcal{N}(y-\ell_1, y+\ell_1)\right] \\
&\leqslant \operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H) + CN^{-2\varepsilon} + C\frac{\ell_1}{\ell}\mathcal{N}(E-2\ell, E+\ell)
\end{aligned}$$

holds with ζ-high probability. From (1.22), (6.9), $\ell_1/\ell = 2N^{-2\varepsilon}$ and $\ell \leqslant N^{-2/3}$, we gather that

$$\frac{\ell_1}{\ell}\mathcal{N}(E-2\ell, E+\ell) \leqslant N^{1-2\varepsilon}\int_{E-2\ell}^{E+\ell} \varrho_W(x)\mathrm{d}x + N^{-2\varepsilon}(\log N)^{L_1} \leqslant \frac{1}{2}N^{-\varepsilon}$$

holds with ζ-high probability, where we estimated the explicit integral using that the integration domain is in a $CN^{-2/3}\varphi^{C_\zeta}$-vicinity of the edge at λ_+. We have thus proved

$$\mathcal{N}(E, E_\zeta) = \operatorname{Tr}\chi_E(H) \leqslant \operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H) + N^{-\varepsilon}.$$

Using (6.1) we see that one can replace $\mathcal{N}(E, E_\zeta)$ by $\mathcal{N}(E,\infty)$ with a change of probability of at most $Ce^{-\varphi^{C_\zeta}}$. This proves the upper bound of (6.10) and the lower bound can be proved similarly.

On the event that (6.10) holds, the condition $\mathcal{N}(E,\infty) = 0$ implies that $\operatorname{Tr}\chi_{E+\ell} * \theta_\eta(H) \leqslant 1/9$. Thus we have

$$\mathbb{P}\left(\mathcal{N}(E,\infty) = 0\right) \leqslant \mathbb{P}\left(\operatorname{Tr}\chi_{E+\ell} * \theta_\eta(H) \leqslant 1/9\right) + Ce^{-\varphi^{C_\zeta}}. \tag{6.12}$$

Together with the Markov inequality, this proves the upper bound in (6.11). For the lower bound, we use

$$\mathbb{E}\, q\big(\operatorname{Tr}\chi_{E-\ell}*\theta_\eta(H)\big) \leqslant \mathbb{P}\big(\operatorname{Tr}\chi_{E-\ell}*\theta_\eta(H) \leqslant 2/9\big) \leqslant \mathbb{P}\big(\mathcal{N}(E,\infty) \leqslant 2/9 + N^{-\varepsilon}\big) = \mathbb{P}\big(\mathcal{N}(E,\infty) = 0\big),$$

where we used the upper bound from (6.10) and that $\mathcal{N}$ is an integer. This completes the proof of the Corollary 6.2. ☐

6.1. *Green Function Comparison Theorem.* Recall the matrices $X^{\mathbf{v}} = [x_{ij}^{\mathbf{v}}]$, $X^{\mathbf{w}} = [x_{ij}^{\mathbf{w}}]$, $H^{\mathbf{v}} = (X^{\mathbf{v}})^\dagger X^{\mathbf{v}}$, $H^{\mathbf{w}} = (X^{\mathbf{w}})^\dagger X^{\mathbf{w}}$ and their respective Green functions $G^{\mathbf{v}}, G^{\mathbf{w}}$ from Section 5. Define $m^{\mathbf{v}} = \frac{1}{N}\operatorname{Tr} G^{\mathbf{v}}(z)$ and $m^{\mathbf{w}} = \frac{1}{N}\operatorname{Tr} G^{\mathbf{w}}(z)$.

Also notice from (6.5) that $\theta_\eta(H) = \frac{1}{\pi}\Im\, m(i\eta)$. Corollary 6.2 bounds the probability of $\mathcal{N}(E,\infty) = 0$ in terms of the expectations of two functionals of Green functions. In this subsection, we show that the difference between the expectations of these functionals w.r.t. two ensembles $X^{\mathbf{v}}$ and $X^{\mathbf{w}}$ is negligible assuming their second moments match. The precise statement is the following Green function comparison theorem on the edges. All statements are formulated for the upper spectral edge λ_+, but identical arguments hold for the lower spectral edge λ_- as well.

Theorem 6.3 (Green function comparison theorem on the edge). *Let $F:\mathbb{R}\to\mathbb{R}$ be a function whose derivatives satisfy*

$$\max_x |F^{(\alpha)}(x)|\,(|x|+1)^{-C_1} \leqslant C_1, \qquad \alpha = 1,\ 2,\ 3,\ 4 \tag{6.13}$$

with some constant $C_1 > 0$. Then there exists $\varepsilon_0 > 0$ depending only on C_1 such that for any $\varepsilon < \varepsilon_0$ and for any real numbers E, E_1 and E_2 satisfying

$$|E - \lambda_+| \leqslant N^{-2/3+\varepsilon}, \qquad |E_1 - \lambda_+| \leqslant N^{-2/3+\varepsilon}, \qquad |E_2 - \lambda_+| \leqslant N^{-2/3+\varepsilon},$$

and $\eta = N^{-2/3-\varepsilon}$, we have

$$\left| \mathbb{E}^{\mathbf{v}} F\left(N\eta \,\Im\, m^{\mathbf{v}}(z)\right) - \mathbb{E}^{\mathbf{w}} F\left(N\eta \,\Im\, m^{\mathbf{w}}(z)\right) \right| \leqslant CN^{-1/6+C\varepsilon}, \qquad z = E + i\eta, \tag{6.14}$$

and

$$\left| \mathbb{E}^{\mathbf{v}} F\left(N \int_{E_1}^{E_2} \mathrm{d}y \,\Im\, m^{\mathbf{v}}(y+i\eta) \right) - \mathbb{E}^{\mathbf{w}} F\left(N \int_{E_1}^{E_2} \mathrm{d}y \,\Im\, m^{\mathbf{w}}(y+i\eta) \right) \right| \leqslant CN^{-1/6+C\varepsilon} \tag{6.15}$$

for some constant C and large enough N.

Theorem 6.3 holds in a much greater generality. We state the following extension which can be used to prove (1.29), the generalization of Theorem 1.7. The class of functions F in the following theorem can be enlarged to allow some polynomially increasing functions similar to (6.13). But for the application to prove (1.29), the following form is sufficient. The proof of Theorem 6.4 is similar to that of Theorem 6.3 and will be omitted.

THEOREM 6.4. *Suppose that the assumptions of Theorem 1.7 hold. Fix any $k \in \mathbb{N}_+$ and let $F : \mathbb{R}^k \to \mathbb{R}$ be a bounded smooth function with bounded derivatives. Then for any sufficiently small ε there exists a $\delta > 0$ such that for any sequence of real numbers $E_k < \ldots < E_1 < E_0$ with $|E_j - \lambda_+| \leqslant N^{-2/3+\varepsilon}$, $j = 0, 1, \ldots, k$, we have*

$$\left| \mathbb{E} F\left(N \int_{E_1}^{E_0} \mathrm{d}y \,\Im\, m(y+i\eta), \ldots, N \int_{E_k}^{E_0} \mathrm{d}y \,\Im\, m(y+i\eta) \right) - \mathbb{E} F\left(m^{\mathbf{v}} \to m^{\mathbf{w}}\right) \right| \leqslant N^{-\delta}, \tag{6.16}$$

where in the second term the arguments of F are changed from $m^{\mathbf{v}}$ to to $m^{\mathbf{w}}$ and all other parameters remain unchanged.

Assuming that Theorem 6.3 holds, we now prove Theorem 1.7.

Proof of Theorem 1.7. As we discussed in (6.1) and (6.2), we can assume that (6.3) holds for the parameter s. We define $E := 2 + sN^{-2/3}$ that satisfies (6.9). We define E_ζ as in (6.4) with the C_ζ such that (6.1) and (6.2) hold. With the left side of (6.11), for any sufficiently small $\varepsilon > 0$, we have

$$\mathbb{E}^{\mathbf{w}}\, q\left(\operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H)\right) \leqslant \mathbb{P}^{\mathbf{w}}(\mathcal{N}(E,\infty)=0)$$

with the choice

$$\ell := \frac{1}{2}N^{-2/3-\varepsilon}, \qquad \eta := N^{-2/3-9\varepsilon}.$$

By definition:

$$\operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H) = N\frac{1}{\pi}\int_{E-\ell}^{E_\zeta} \Im\, m(y+i\eta)dy$$

The bound (6.15) applying to the case $E_1 = E-\ell$ and $E_2 = E_\zeta$ shows that there exist $\delta > 0$, for sufficiently small $\varepsilon > 0$, such that

$$\mathbb{E}^{\mathbf{v}}\, q\left(\operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H)\right) \leqslant \mathbb{E}^{\mathbf{w}}\, q\left(\operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H)\right) + N^{-\delta} \tag{6.17}$$

(note that 9ε plays the role of the ε in the Green function comparison theorem). Then applying the right side of (6.11) in Lemma 6.2 to the l.h.s of (6.17), we have

$$\mathbb{P}^{\mathbf{v}}(\mathcal{N}(E-2\ell,\infty)=0) \leqslant \mathbb{E}^{\mathbf{v}}\, q\left(\operatorname{Tr}\chi_{E-\ell} * \theta_\eta(H)\right) + C\exp\left[-c\varphi^{O(1)}\right].$$

Combining these inequalities, we have

$$\mathbb{P}^{\mathbf{v}}(\mathcal{N}(E-2\ell,\infty)=0) \leqslant \mathbb{P}^{\mathbf{w}}(\mathcal{N}(E,\infty)=0) + 2N^{-\delta} \tag{6.18}$$

for sufficiently small $\varepsilon > 0$ and sufficiently large N. Recalling that $E = 2 + sN^{-2/3}$, this proves the first inequality of (1.28) and, by switching the role of $\mathbf{v}, \mathbf{w}$, the second inequality of (1.28) as well. This completes the proof of Theorem 1.7. □

Proof of Theorem 6.3. The proof is similar to that of Lemma 5.3. We need to compare the matrices $H^{\mathbf{v}}$ and $H^{\mathbf{w}}$. Instead of replacing the matrix elements one by one (NM times) and comparing their successive differences, here we estimate the successive difference of matrices which differ by a column. Indeed for $1 \leqslant \gamma \leqslant N$, denote by X_γ the random matrix whose j-th column is the same as that of $X^{\mathbf{v}}$ if $j \leqslant \gamma$ and that of $X^{\mathbf{w}}$ otherwise; in particular $X_0 = X^{\mathbf{v}}$ and $X_N = X^{\mathbf{w}}$. As before, we define

$$H_\gamma = X_\gamma^\dagger X_\gamma\ .$$

We will compare $H_{\gamma-1}$ with H_γ using the following lemma. For simplicity, we denote

$$\widetilde{m}^{(i)}(z) = m^{(i)}(z) - (Nz)^{-1}.$$

Lemma 6.5. *For any random matrix X whose entries satisfy (1.1) and (1.2), if $|E-\lambda_+|\leqslant N^{-2/3+\varepsilon}$ and $N^{-2/3}\gg\eta\geqslant N^{-2/3-\varepsilon}$ for some $\varepsilon>0$, then we have*

$$\mathbb{E}\,F\left(N\eta\,\Im\,m(z)\right)-\mathbb{E}\,F\left(N\eta\,\Im\,\widetilde{m}^{(i)}(z)\right)=A(X^{(i)},m_1,m_2)+N^{-7/6+C\varepsilon} \tag{6.19}$$

where the functional $A(X^{(i)},m_1,m_2)$ only depends on the distribution of $X^{(i)}$ and the first two moments m_1,m_2 of $\sqrt{M}x_{ij}=\sqrt{M}(X)_{ij}$.

Notice that Lemma 6.5 implies that

$$\mathbb{E}\,F\left(\eta\,\Im\,\mathrm{Tr}\,\frac{1}{H_{\gamma-1}-z}\right)-\mathbb{E}\,F\left(\eta\,\Im\,\mathrm{Tr}\,\frac{1}{H_{\gamma}-z}\right)=A(X_\gamma^{(\gamma)},m_1,m_2)+N^{-7/6+C\varepsilon} \tag{6.20}$$

where $A(X_\gamma^{(\gamma)},m_1,m_2)$ only depends on the distribution of $X_\gamma^{(\gamma)}$ and m_1,m_2. As done in Theorem 5.3, the proof of Theorem 6.3 now can be completed via the telescoping argument.

Thus to finish the proof of Theorem 6.3, all that needs to be shown is Lemma 6.5 which is proven below:

Proof of Lemma 6.5. Without loss of generality, we assume that $i=1$ and ε is small enough. First, we claim some bounds about $G^{(1)}$ and $\mathcal{G}^{(1)}$. For any $\zeta>0$,

$$\left|\left(\mathbf{x}_1(\mathcal{G}^{(1)})^2\mathbf{x}_1\right)\right|\leqslant N^{1/3+C\varepsilon} \tag{6.21}$$

$$\left|[\mathcal{G}^{(1)}]_{ij}\right|\leqslant N^{C\varepsilon},\quad \left|\left([\mathcal{G}^{(1)}]^2\right)_{ij}\right|\leqslant N^{1/3+C\varepsilon}, \tag{6.22}$$

with ζ-high probability for some $C>0$, where i could be equal to j. We postpone the proof of these bounds to the end. For Lemma 6.5, using (2.5) and (2.7), we have

$$\begin{aligned}\mathrm{Tr}\,G-\mathrm{Tr}\,G^{(1)}+z^{-1}&=\left(G_{11}+z^{-1}\right)+\frac{\left(\mathbf{x}_1\cdot X^{(1)}G^{(1)}G^{(1)}X^{(1)\dagger}\cdot\mathbf{x}_1\right)}{-z-z(\mathbf{x}_1,\mathcal{G}^{(1)}(z)\,\mathbf{x}_1)}\\&=z\,G_{11}\left(\mathbf{x}_1(\mathcal{G}^{(1)})^2(z)\,\mathbf{x}_1\right)\ .\end{aligned} \tag{6.23}$$

Define the quantity B to be

$$B=-z\,m_W\left[\left(\mathbf{x}_1,\mathcal{G}^{(1)}(z)\mathbf{x}_1\right)-\left(\frac{-1}{z\,m_W(z)}-1\right)\right]\ . \tag{6.24}$$

By (2.5),

$$B=-z\,m_W\left[\left(\frac{-1}{z\,G_{11}(z)}-1\right)-\left(\frac{-1}{z\,m_W(z)}-1\right)\right]=\frac{m_W-G_{11}}{G_{11}}$$

From (1.18) , we obtain that

$$|B| \leqslant N^{-1/3+2\varepsilon} \ll 1 \,, \tag{6.25}$$

with with ζ-high probability with any $\zeta > 0$. Therefore, we have the identity

$$G_{11} = \frac{m_W}{B+1} = m_W \sum_{k\geqslant 0} (-B)^k \,. \tag{6.26}$$

Define y with the l.h.s of (6.23),

$$y := \eta \left(\operatorname{Tr} G - \operatorname{Tr} G^{(1)} + z^{-1} \right)$$

so that using (6.23) and (6.26) we obtain

$$y = \eta z \, G_{11} \left(\mathbf{x}_1 (\mathcal{G}^{(1)})^2 \mathbf{x}_1 \right) = \sum_{k=1}^{\infty} y_k, \quad y_k := \eta z \, m_W (-B)^{k-1} \left(\mathbf{x}_1 (\mathcal{G}^{(1)})^2 \mathbf{x}_1 \right) \,.$$

Since z and m_W are $O(1)$, together with (6.21) and (6.25),

$$|y_k| \leqslant O(N^{-k/3+C\varepsilon}) \quad \text{and} \quad |y| \leqslant O(N^{-1/3+C\varepsilon}) \tag{6.27}$$

holds with ζ-high probability. Consequently the expansion

$$F\left(N\eta \,\Im\, m(z)\right) - F\left(N\eta \,\Im\, \widetilde{m}^{(1)}(z)\right) = \sum_{k=1}^{3} F^{(k)}\left(N\eta \,\Im\, \widetilde{m}^{(1)}(z)\right) (\Im\, y)^k + O(N^{-4/3+C\varepsilon}) \tag{6.28}$$

holds with ζ-high probability. First using (6.27) we obtain that

$$F^{(3)}\left(N\eta \,\Im\, \widetilde{m}^{(1)}(z)\right) (\Im\, y)^3 = F^{(3)}\left(N\eta \,\Im\, \widetilde{m}^{(1)}(z)\right) (\Im\, y_1)^3 + O(N^{-4/3+C\varepsilon}) \tag{6.29}$$

holds with ζ-high probability. Moreover, we have

$$\mathbb{E}_1 (\Im\, y_1)^3 = \mathbb{E}_1 (\eta z \, m_W)^3 \left(\mathbf{x}_1 (\mathcal{G}^{(1)})^2 \mathbf{x}_1 \right)^3 = (\eta z \, m_W)^3 \sum_{k_1, \ldots, k_6} \mathbb{E}_1 (\prod_{i=1}^{6} x_{1k_i}) \prod_{i=1}^{3} [(\mathcal{G}^{(1)})^2]_{k_{2i-1},\, k_{2i}},$$

where $\mathbb{E}_1$ is the expectation value with respect to the first column of X. Recall m_k is the k-th moment of $\sqrt{M} x_{ij}$. We know if there is a k_i not equal to any other x_j then

$$\mathbb{E}_1 (\prod_{i=1}^{6} x_{1k_i}) = 0 = m_1 \,,$$

and if each k_i appears exactly twice, then

$$\mathbb{E}_1(\prod_{i=1}^{6} x_{1k_i}) = m_2^3 \,.$$

Therefore, we have

$$\mathbb{E}_1(\Im\, y_1)^3 = \widetilde{A}_3(X^{(1)}, m_1, m_2) + (\eta z\, m_W)^3 \sum_{(1),(2)} \mathbb{E}_1(\prod_{i=1}^{6} x_{1k_i})[\mathcal{G}^{(1)})^2]_{k_1k_2}[\mathcal{G}^{(1)})^2]_{k_3k_4}[\mathcal{G}^{(1)})^2]_{k_5k_6}$$

where we sum up the k_i's such that (1) no k_i appearing only once and (2) at least one k_i appearing three times and the functional $\widetilde{A}_3(X^{(1)}, m_1, m_2)$ only depends on $X^{(1)}$, m_1 and m_2. With this condition, we obtain that there are at most two elements in the set $\{k_1, k_2, \ldots, k_6\}$, i.e., $\sum_{(1),(2)} 1 \leqslant CN^2$. Then using (6.22), the bounds on m_k's, we have

$$\mathbb{E}_1(\Im\, y_1)^3 = \widetilde{A}_3(X^{(1)}, m_1, m_2) + O(N^{-2+C\varepsilon}) \tag{6.30}$$

By definition, it is easy to prove $|N\eta\, \Im\, \widetilde{m}^{(1)}| \leqslant N^{C\varepsilon}$ with ζ-high probability. Then using (6.29) and the fact that $\widetilde{m}^{(1)}$ only depends on $X^{(1)}$, we have

$$\mathbb{E}F^{(3)}\left(N\eta\, \Im\, \widetilde{m}^{(1)}(z)\right)(\Im\, y)^3 = A_3(X^{(1)}, m_1, m_2) + O(N^{-4/3+C\varepsilon})\,. \tag{6.31}$$

where $A_3(X^{(1)}, m_1, m_2)$ only depends on the distribution of $X^{(1)}$, m_1 and m_2.

Now we estimate the term with $F^{(2)}$ in (6.28). As in (6.29), we have

$$F^{(2)}\left(N\eta\, \Im\, \widetilde{m}^{(1)}(z)\right)(\Im\, y)^2 = F^{(2)}\left(N\eta\, \Im\, \widetilde{m}^{(1)}(z)\right)\left[(\Im\, y_1)^2 + 2(\Im\, y_1)(\Im\, y_2)\right] + O(N^{-4/3+C\varepsilon})\,. \tag{6.32}$$

By definition,

$$\mathbb{E}_1(\Im\, y_1)^2 + 2(\Im\, y_1)(\Im\, y_2) = C_1(z)\eta^2\left(\mathbf{x}_1(\mathcal{G}^{(1)})\mathbf{x}_1\right)\left(\mathbf{x}_1(\mathcal{G}^{(1)})^2\mathbf{x}_1\right)^2 + C_2(z)\eta^2\left(\mathbf{x}_1(\mathcal{G}^{(1)})^2\mathbf{x}_1\right)^2$$

where $C_1(z)$, $C_2(z) = O(1)$ are constant depending on z and $m_W(z)$. Using the bounds on $\mathcal{G}^{(1)}$ in (6.22), as in (6.30), we have

$$\mathbb{E}_1(\Im\, y_1)^2 + (\Im\, y_1)(\Im\, y_2) = \widetilde{A}_2(X^{(1)}, m_1, m_2) + O(N^{-5/3+C\varepsilon})$$

where $\widetilde{A}_2(X^{(1)}, m_1, m_2)$ only depends on $X^{(1)}$, m_1 and m_2. Then with (6.34), as in (6.31),

$$\mathbb{E}F^{(2)}\left(N\eta\, \Im\, \widetilde{m}^{(1)}(z)\right)(\Im\, y)^2 = A_2(X^{(1)}, m_1, m_2) + O(N^{-4/3+C\varepsilon})\,. \tag{6.33}$$

for some functional A_2 which only depends on the distribution of $X^{(1)}$, m_1 and m_2.

Now we estimate the term with $F^{(1)}$ in (6.28). As in (6.29), we have

$$F^{(1)}\left(N\eta\,\Im\,\widetilde{m}^{(1)}(z)\right)(\Im\,y)^2 = F^{(1)}\left(N\eta\,\Im\,\widetilde{m}^{(1)}(z)\right)[\Im\,y_1 + \Im\,y_2 + \Im y_3] + O(N^{-4/3+C\varepsilon})\,. \tag{6.34}$$

Similar argument as (6.33) and (6.31) yields

$$\mathbb{E}F^{(1)}\left(N\eta\,\Im\,\widetilde{m}^{(1)}(z)\right)(\Im\,y) = A_1(X^{(1)}, m_1, m_2) + O(N^{-4/3+C\varepsilon})\,. \tag{6.35}$$

Inserting (6.35), (6.33) and (6.31) into (6.28), we obtain (6.19) and complete the proof of Lemma 6.5 and consequently finish the proof of Theorem 6.3.

At last, we prove (6.21) and (6.22). For (6.21), using the large deviation lemma (Lemma 1.4), we obtain that for any $\zeta > 0$,

$$\begin{aligned}(6.36)\qquad \left(\mathbf{x}_1(\mathcal{G}^{(1)})^2\mathbf{x}_1\right) &\leqslant \varphi^{C_\zeta}(N^{-1}\operatorname{Tr}|\mathcal{G}^{(1)}|^4)^{1/2} \leqslant \varphi^{C_\zeta}\left(\frac{1}{N}\sum_\alpha\frac{1}{|\lambda^{(1)}_\alpha - z|^4}\right)^{1/2}\\ &\leqslant \varphi^{C_\zeta}\left(\frac{1}{N\eta^2}\sum_\alpha\frac{1}{|\lambda^{(1)}_\alpha - z|^2}\right)^{1/2} = \varphi^{C_\zeta}\left(\frac{1}{N\eta^3}\,\Im\,m^{(1)}(z)\right)^{1/2}\end{aligned}$$

with ζ-high probability. Then with (1.17) and (2.34), we have (6.21). For (6.22), we note

$$\mathcal{G}^{(1)} = \frac{1}{X^{(1)}(X^{(1)})^\dagger - z}$$

Comparing with (1.3), we see that $(\mathcal{G}^{(1)}, (X^{(1)})^\dagger)$ play the role of (G, X). Since $\sqrt{\frac{M}{N-1}}(X^{(1)})^\dagger$ is just a normal $(N-1)\times M$ random matrix, whose entry has variance $(N-1)^{-1}$, the results in (1.18) also holds for $\mathcal{G}^{(1)}$ with slight changes. One can easily obtain that

$$\max_{ij}|[\mathcal{G}^{(1)}]_{ij}| \leqslant C, \quad \max|[\mathcal{G}^{(1)}]_{ij}| \leqslant CN^{-1/3+C\varepsilon}$$

with ζ-high probability for any $\zeta > 0$ proving (6.22) and we are done. □

7. Acknowledgements. The authors would like to thank Horng-Tzer Yau, Antti Knowles, Lazlo Erdös and Paul Bourgade for very useful discussions and help. NSP and JY gratefully acknowledge the NSF grants DMS-1107070 and DMS-1001655 respectively.

REFERENCES

[1] Ben Arous, G., Peche, S., Universality of local eigenvalue statistics for some sample covariance matrices. *Comm. Pure Appl. Math.*, **58** , No. 10, 1316-1357 (2005).

[2] Dieng, M., Tracy, C.A., Application of Random Matrix Theory to Multivariate Statistics, *Arxiv Preprint*, http://arxiv.org/abs/math/0603543 (2006).

[3] Dyson, F.J.: A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.* **3**, 1191-1198 (1962)

[4] Erdős, L., Knowles, A., Yau, H.-T and Yin, J. Spectral Statistics of Erdós-Rényi Graphs I: Local Semicircle Law, arxiv preprints. 1103.1919

[5] Erdős, L., Knowles, A., Yau, H.-T and Yin, J. Spectral Statistics of Erdós-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues, arxiv preprints. 1103.3869

[6] Erdős, L., Schlein, B., Yau, H.-T., Universality of random matrices and local relaxation flow. Preprint. arxiv.org/abs/0907.5605

[7] Erdős, L., Péché, G., Ramírez, J., Schlein, B., and Yau, H.-T., Bulk universality for Wigner matrices. *Commun. Pure Appl. Math.* **63**, No. 7, 895–925 (2010)

[8] Erdős, L., Schlein, B., Yau, H.-T., Yin, J., The local relaxation flow approach to universality of the local statistics for random matrices. Preprint arXiv:0911.3687

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

[10] Erdős, L., Yau, H.-T., Yin, J., Universality for generalized Wigner matrices with Bernoulli distribution. Preprint arXiv: 1003.3813.

[11] Erdős, L., Yau, H.-T., Yin, J., Rigidity of Eigenvalues of Generalized Wigner Matrices.

[12] Harding, M., Explaining the single factor bias of arbitrage pricing models in finite samples, *Economic Letters*, **99**, 85-88 (2008).

[13] Johnstone, I. M., On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, No. 2, 295-327 (2001).

[14] Johnstone, I.M., High Dimensional Statistical Inference and Random Matrices, *International Congress of Mathematicans*, Vol. **1**, 307-333, Eur. Math. Soc. Zurich (2007).

[15] Johnstone, I.M., Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence, *Ann. of Stat.*, **36**, No. 6, 2638-2716 (2008).

[16] Onatski, A.,Testing hypotheses about the number of factors in large factor models. *Econometrika*, To Appear, (2009).

[17] Patterson, N., Price, A.L., Reich, D., Population structure and eigen analysis, *PLoS Genetics*, No. **2 (12)** (2006):e190.

[18] Peche, S., Universality results for largest eigenvalues of some sample covariance matrix ensembles, *Probab. Theory Related Fields* 143, No. 3-4, (2009).

[19] Soshnikov, A., Universality at the edge of the distribution of the largest eigenvalue in certain classes of sample covariance matrices, *J. Stat. Phys.*, **108**, 1033-1056 (2002).

[20] Tao, T. and Vu, V., *Random covariance matrices: Universality of local statistics of eigenvalues*, ArXiv e-prints : 0912.0966.

[21] Wang, K., Random covariance matrices: Universality of local statistics of eigenvalues up to the edge, *ArXiv e-prints, 1104.4832* (2011).

Department of Statistics,
Harvard University
1 Oxford Street, Cambridge
02138, MA, USA
E-mail: pillai@stat.harvard.edu

Department of math,
University of Wisconsin -Madison
480 Lincoln Dr. Madison
53706 ,WI, USA
E-mail: jyin@math.wisc.edu