The Annals of Applied Probability
2014, Vol. 24, No. 3, 935–1001
DOI: 10.1214/13-AAP939

UNIVERSALITY OF COVARIANCE MATRICES

BY NATESH S. PILLAI[1] AND JUN YIN[2]

Harvard University and University of Wisconsin–Madison

In this paper we prove the universality of covariance matrices of the form $H_{N\times N} = X^{\dagger}X$ where X is an $M \times N$ 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 or sufficiently high number of moments. We will study the asymptotics in the regime $N/M = d_N \in (0, \infty)$, $\lim_{N\to\infty} d_N \neq 0, \infty$. Our main result is the edge universality of the sample covariance matrix at *both* edges of the spectrum. In the case $\lim_{N\to\infty} d_N = 1$, we only focus on the largest eigenvalue. Our proof is based on a novel version of the Green function comparison theorem for data matrices with dependent entries. En route to proving edge universality, we establish 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. Combining these results with existing techniques we also show bulk universality of covariance matrices. All our results hold for both real and complex valued entries.

1. Introduction.

In this paper we prove the universality of covariance matrices. Let $X = (x_{ij})$ be an $M \times N$ data matrix with independent centered real valued entries with variance M^{-1},

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

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

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

The covariance matrix corresponding to data matrix X is given by $H = X^{\dagger}X$. We will be working in the regime

$$d = d_N = N/M, \qquad \lim_{N\to\infty} d \neq 0, \infty.$$

Received September 2012; revised April 2013.
[1]Supported by the NSF Grant DMS-11-07070.
[2]Supported by the NSF Grant DMS-10-01655.
MSC2010 subject classifications. 15B52, 82B44.
Key words and phrases. Covariance matrix, Marcenko–Pastur law, universality, Tracy–Widom law, Dyson Brownian motion.

Thus without loss of generality, we will assume henceforth that for some small constant θ, for all $N \in \mathbb{N}$,

$$\theta < d_N < \theta^{-1}.$$

All our constants may depend on θ and ϑ, but we will not denote this dependence. In this paper we focus on the case where the matrix X has real valued entries which is a natural assumption for applications in statistics, economics, etc. However all of the results in this paper also hold for complex valued entries with the moment condition (1.1) replaced with its complex valued analogue,

$$x_{ij} = M^{-1/2} q_{ij}, \qquad \mathbb{E} q_{ij} = 0, \qquad \mathbb{E} q_{ij}^2 = 0, \qquad \mathbb{E}|q_{ij}|^2 = 1. \tag{1.3}$$

Furthermore, in some technical results in the present work, the independence of matrix entries are weakened (see Theorem 3.6), which are the key inputs of [3] and [33].

Covariance matrices are fundamental objects in modern multivariate statistics where the advance of technology has led to high-dimensional data. They have manifold applications in various applied fields; see [7, 22–24] for an extensive account on statistical applications, [21, 28] for applications in economics and [30] in population genetics, to name a few. In the regime we study in this paper where N, M are proportional to each other, the exact asymptotic distribution of the eigenvalues is not known, except for some cases under specific assumptions on the distributions of the entries of the covariance matrix, for example, when the entries are Gaussian. In this context, akin to the central limit theorem, the phenomenon 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, as observed by Wigner, 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.

A fundamental example is the well-studied Wishart matrix (the covariance matrix obtained from a data matrix X consisting of i.i.d. centered Gaussian random variables) for which one has closed form expressions for many objects of interest including the joint distribution of the eigenvalues. In this paper we prove the universality of covariance matrices (both at the bulk and at the edges) under the assumption that entries of the corresponding data matrix 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 has been made in proving the universality properties of i.i.d. matrix elements (*standard Wigner ensembles*). The most general results to date for the universality of Wigner ensembles are obtained in Theorems 7.3 and 7.4 of [10], in which bulk (edge) universality is proved for Wigner matrices under the assumption that entries have a uniformly bounded $4+\varepsilon$ $(12+\varepsilon)$ moment for some $\varepsilon > 0$, and then recently improved further by [15] and [29]. The key ideas for the universality of Wigner ensembles were developed through several important steps in [12, 13, 16–18]. The ideas we use in this paper are also adapted from the above cited papers. There are also related results in [36, 37]. However, the results regarding universality of *local statistics* for covariance matrices have been obtained only recently, which we survey below.

1.1. *Review of previous work.* First we review previous results for extreme eigenvalues. In [1, 2, 40], the authors showed the almost sure convergence of extreme eigenvalues. In [20], the authors derived the rate of convergence of the spectrum to the Marchenko–Pastur law. In [34], Soshnikov showed that for $d_N = 1 - O(N^{-1/3})$, if q_{ij} in (1.1) have a symmetric distribution and Gaussian decay, then the largest eigenvalues (appropriately rescaled) converge to the Tracy–Widom distribution. This condition on d_N was replaced with $\lim_{N\to\infty} d_N \in (0,\infty)$ by Péché [31]. Using similar assumptions as in [34] and [31], Feldheim and Sodin [19] showed that the smallest eigenvalues (appropriately rescaled) converge to the Tracy–Widom distribution for $\lim_{N\to\infty} d_N \neq 1$. More recently, for $\lim_{N\to\infty} d_N \neq 1$, Wang [39] proved the Tracy–Widom law for the limiting distribution of the extreme eigenvalues under the assumption that q_{ij} in (1.1) have vanishing third moment and sufficiently high number of moments. For "square" matrices, that is, when $N = M$ and thus $d_N = 1$, Tao and Vu [35] proved the universality of the smallest eigenvalues assuming the matrix entries have sufficiently high number of moments. The limiting distribution of the smallest eigenvalue for square matrices with standard Gaussian entries were computed by Edelman [9]. In our main result below, *we show universality of eigenvalues for "rectangular" data matrices at both edges of the spectrum, assuming only* (1.1) *and* (1.2).

Now we review results for the local statistics of the eigenvalues in the bulk of the spectrum. It was widely believed until recently that the distribution of the distance between adjacent eigenvalues is independent of the distribution of q_{ij} in (1.1). In [4] Arous and Péché showed this bulk universality when $d_N = 1 + O(N^{-5/48})$. Tao and Vu [38] proved that the asymptotic distribution for local statistics at the bulk corresponding to two covariance matrices are identical, if the entries in these two matrices have identical first four moments. On the other hand, in [32] and [14], Péché, Erdős, Schlein, Yau and the second author of this paper showed this bulk universality under some regularity conditions and decay assumptions on the distribution of the matrix entries. *We also show bulk universality but under weaker assumptions than those in* [32] *and* [14]; see Remark 1.7 for more details.

1.2. *Our key results.* Let $X^{\mathbf{v}} = [x^{\mathbf{v}}_{ij}]$ with independent entries satisfying (1.1) and (1.2), and let

$$\lambda^{\mathbf{v}}_1 \ge \lambda^{\mathbf{v}}_2 \cdots \lambda^{\mathbf{v}}_{\min\{M,N\}} \ge 0$$

denote the nontrivial singular values of the data matrix $X^{\mathbf{v}}$. Let $\mathbb{P}^{\mathbf{v}}$ denote the probability measure according to which the entries of $X^{\mathbf{v}}$ are distributed. Let $X^{\mathbf{w}}$, $\{\lambda^{\mathbf{w}}_k\}_{k\le\min\{M,N\}}$ and $\mathbb{P}^{\mathbf{w}}$ be defined analogously. The following is our main result:

THEOREM 1.1 (Universality of extreme eigenvalues). *For* $\lim_{N\to\infty} d_N \in (0,\infty)$, *there is an* $\varepsilon > 0$ *and* $\delta > 0$ *such that for any real number* s (*which may depend on* N),

$$\begin{aligned} &\mathbb{P}^{\mathbf{v}}\big(N^{2/3}(\lambda_1^{\mathbf{v}} - \lambda_+) \le s - N^{-\varepsilon}\big) - N^{-\delta} \\ &\quad \le \mathbb{P}^{\mathbf{w}}\big(N^{2/3}(\lambda_1^{\mathbf{w}} - \lambda_+) \le s\big) \\ &\quad \le \mathbb{P}^{\mathbf{v}}\big(N^{2/3}(\lambda_1^{\mathbf{v}} - \lambda_+) \le s + N^{-\varepsilon}\big) + N^{-\delta} \end{aligned} \tag{1.4}$$

for $N \ge N_0$ *sufficiently large, where* N_0 *is independent of* s. *An analogous result holds for the smallest eigenvalues* $\lambda^{\mathbf{v},\mathbf{w}}_{\min\{M,N\}}$, *when* $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$.

In [31, 34] and [19], Soshnikov, Péché, Feldheim and Sodin proved that for the covariance matrices whose entries have a symmetric probability density function (which includes the Wishart matrix), the largest and smallest k eigenvalues after appropriate centering and rescaling converge in distribution to the Tracy–Widom law.[3] We have the following immediate corollary of Theorem 1.1:

COROLLARY 1.2. *Let* X *with independent entries satisfying* (1.1) *and* (1.2), *and let* $\lim_{N\to\infty} d_N \in (0,\infty)$. *For any fixed* $k > 0$, *we have*

$$\biggl(\frac{M\lambda_1 - (\sqrt{N}+\sqrt{M})^2}{(\sqrt{N}+\sqrt{M})((1/\sqrt{N})+(1/\sqrt{M}))^{1/3}}, \dots, \\ \frac{M\lambda_k - (\sqrt{N}+\sqrt{M})^2}{(\sqrt{N}+\sqrt{M})((1/\sqrt{N})+(1/\sqrt{M}))^{1/3}}\biggr) \longrightarrow \mathrm{TW}_1,$$

where TW_1 *denotes the Tracy–Widom distribution. An analogous statement holds for the smallest eigenvalues, when* $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$.

REMARK 1.3. Clearly, our result covers the case where the matrix entries have Gaussian divisible distribution (see [39], Section 2) and the case where the support of the distribution of the matrix entries consists of only two points. Using these two cases and the results of [39], the sub-exponential-decay assumption in Corollary 1.2 can be replaced with the existence of sufficiently high number of moments. For details, see the discussion below the Theorem 2.2 of [39]. However we believe that all of our results can be proved under a uniform bound on pth moments of the matrix elements (say $p = 4$ or 5), using the methods in [10] and [29]; we will pursue this elsewhere.

REMARK 1.4. Theorem 1.1 can be extended to obtain universality of finite correlation functions of extreme eigenvalues. For example, we have the following

[3] Here we use the term Tracy–Widom law as in [34].

extension of (1.4): for any fixed k,

$$
\begin{aligned}
&\mathbb{P}^{\mathbf{v}}\big(N^{2/3}(\lambda_1^{\mathbf{v}}-\lambda_+)\le s_1-N^{-\varepsilon},\dots,N^{2/3}(\lambda_k^{\mathbf{v}}-\lambda_+)\le s_k-N^{-\varepsilon}\big)-N^{-\delta}\\
&\quad\le \mathbb{P}^{\mathbf{w}}\big(N^{2/3}(\lambda_1^{\mathbf{w}}-\lambda_+)\le s_1,\dots,N^{2/3}(\lambda_k^{\mathbf{w}}-\lambda_+)\le s_k\big)\\
&\quad\le \mathbb{P}^{\mathbf{v}}\big(N^{2/3}(\lambda_1^{\mathbf{v}}-\lambda_+)\le s_1+N^{-\varepsilon},\dots,\\
&\qquad\qquad N^{2/3}(\lambda_k^{\mathbf{v}}-\lambda_+)\le s_k+N^{-\varepsilon}\big)+N^{-\delta}
\end{aligned}
\tag{1.5}
$$

for all sufficiently large N. The proof of (1.5) is similar to that of (1.4), and we will not provide details, except stating the general form of the Green function comparison theorem (Theorem 4.4) needed in this case. We remark that edge universality is usually formulated in terms of joint distributions of edge eigenvalues in the form (1.5) with fixed parameters $s_1, s_2, \dots$, etc. Our result holds uniformly in these parameters, that is, they may depend on N. However, the interesting regime is $|s_j| \le O((\log N)^{\log\log N})$; otherwise, the rigidity estimate obtained in (3.6) will give stronger control than (1.5).

The first step toward proving Theorem 1.1 is to obtain a strong *local Marcenko–Pastur law*, a precise estimate of the local eigenvalue density in the optimal scale $N^{-1+o(1)}$. We state and prove this in Theorem 3.1. This theorem is our key technical tool for proving rigidity of eigenvalues (see Theorem 3.3) and universality. En route to this, we also obtain precise bounds on the matrix elements of the corresponding Green function. All of our results regarding the strong Marcenko–Pastur law do not require independence of the entries of the data matrix, but need only weak dependence as will be explained in Section 3. An important technical ingredient required for the estimates for our strong Marcenko–Pastur law and the rigidity of eigenvalues is an abstract decoupling lemma (Lemma 7.3) for weakly dependent random variables, proved in Section 7.

Using the strong Marcenko–Pastur law and the existing results (such as [16] and Theorem 2.1 in [14]), we also show bulk universality holds for covariance matrices in almost optimal scale:

THEOREM 1.5 (Universality of eigenvalues in bulk). *Let $X^{\mathbf{v}}$, $X^{\mathbf{w}}$ be as defined before. Assume that $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$. Let $E \in [\lambda_- + r, \lambda_+ - r]$ with some $r > 0$. Then for any $\varepsilon > 0$, $N^{-1+\varepsilon} < b < r/2$, any fixed integer $n \ge 1$ and for any compactly supported continuous test function $O : \mathbb{R}^n \to \mathbb{R}$, we have*

$$
\begin{aligned}
\lim_{N\to\infty}\int_{E-b}^{E+b}\frac{\mathrm{d}E'}{2b}\int_{\mathbb{R}^n} &O(\alpha_1,\dots,\alpha_n)\big(p^{(n)}_{\mathbf{v}N}-p^{(n)}_{\mathbf{w},N}\big)\\
&\times\left(E'+\frac{\alpha_1}{N\varrho_c(E)},\dots,E'+\frac{\alpha_n}{N\varrho_c(E)}\right)\prod_i\frac{\mathrm{d}\alpha_i}{\varrho_c(E)}=0,
\end{aligned}
\tag{1.6}
$$

where $p^{(n)}_{\mathbf{v},N}$ and $p^{(n)}_{\mathbf{w},N}$ 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.

REMARK 1.6. As in Remark 1.3, using the four moment theorem in [38], the sub-exponential-decay assumption for the matrix entries can be replaced with the existence of a sufficiently high number of moments.

REMARK 1.7. Compared to the results obtained in [14, 32], our Theorem 1.5 is an improvement on two fronts: (i) in [14, 32], for (1.6), the authors required that

$$\sum_{i=1}^{M_k} \bigl|\partial_x^i \log u_0(x)\bigr| \le C_k\bigl(1+|x|\bigr)^{C_k}$$

for some M_k and C_k, where u_0 is the probability density function of the matrix entries; see formulas (1.3)–(1.5) in [32] and formula (3.6) in [14]. (ii) We show that the bulk university holds in almost optimal scale: $b = N^{-1+\varepsilon}$. In the main theorem of [14], bulk universality was shown for $b \sim O(1)$.[4] We also note that in [32], the integral in (1.6) is not required. On the other hand, the proof in [32] does not work for covariance matrices with real valued entries.

REMARK 1.8. Our result heavily relies on the Theorem 2.1 of [14], but we are able to show universality up to this optimal scale, mainly because of our stronger results on the strong local Marcenko–Pastur law and the rigidity result for eigenvalues obtained in Theorems 3.1 and 3.3, respectively.

REMARK 1.9. Tao and Vu [38] derived bulk universality without the integral in (1.6), but they required that the matrix entries of the two covariance matrices have identical first four moments.

1.3. *Main ideas*. The approach we take in this paper to prove universality is the one developed in a recent series of papers [10–14, 16–18]; however, there are some important differences which we highlight below. Our proof of the above result proceeds via the Green function comparison theorem as in the case of Wigner matrices; however, *unlike Wigner matrices*, *the elements within the same column of a covariance matrix are not independent*. In order to address this key difficulty, we introduce new ideas and *establish a novel version of the Green function comparison theorem*. In particular, in Theorem 4.5 (see Section 6) we give sufficient criteria for proving edge universality for matrix ensembles of the form $Y^\dagger Y$ for a generic data matrix Y with dependent entries (e.g., correlation matrices). *This enables us to show the edge universality for covariance matrices when* $\lim_{N\to\infty} d_N \in (0,\infty)$, *under the assumption that the first two moments of the matrix entries are equal to that of the standard Gaussian*. Our method is also useful for establishing universality for a huge class of matrix ensembles with dependent entries. For example, in a recent paper [3], Bao, Pan and Zhou used our method to show universality

[4] For two quantities a, b we write $a \sim b$ to denote $cb \le a \le Cb$ for some $c, C > 0$.

for a class of correlation matrices. For more general edge universality results for correlation matrices, see a later paper [33], which is also based on our Green function comparison theorem. As mentioned above, for our strong Marcenko–Pastur law, we use an abstract decoupling lemma (Lemma 7.3) for weakly dependent random variables. This lemma is novel and is applicable in other settings such as non-Hermitian ensembles [5].

For proving bulk universality of eigenvalues, we follow the general approach for the universality of Gaussian divisible ensembles [10, 11, 16, 18, 25, 26] by embedding 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 [8]. An important 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, first introduced in [13, 14], eliminates entirely the usage of explicit formulas. We will also follow this route and use the strong local Marcenko–Pastur law to show that the time for the Dyson Brownian motion (corresponding to the covariance matrix) to reach local equilibrium is about $O(N^{-1})$. Once we prove this result, all that remains to be done is to show that the local statistics at $t = O(N^{-1})$ coincide with those of the initial matrix, that is, $t = 0$. To achieve this, we again use the Green function comparison method. Roughly speaking, the Green function comparison method exploits the fact that the equilibrium time is very "small" $[O(N^{-1})]$, and therefore the first few moments of the matrix entries at time $t = N^{-1}$ will be nearly identical to those at $t = 0$.

1.4. *Comments on other limiting regimes of d_N.* The assumption $\lim_{N\to\infty} d_N \in (0,\infty)$ is mostly for simplicity, and we believe that with some more effort, most of our results can be extended to the case $\lim_{N\to\infty} d_N = \{0,\infty\}$. This will be pursued in our future works.

However, we believe that universality at the soft edge for $\lim_{N\to\infty} d_N = 1$ will be much harder. There is a singularity of the eigenvalue density at $x = 0$. More precisely, the typical distance between adjacent eigenvalues near $x = 0$ is $O(N^{-2})$. For studying the smallest eigenvalue one needs to overcome several obstacles: (1) The usual moment method which estimates $\mathbb{E}(X^\dagger X)^k$ with large $k \in \mathbb{N}$ does not work in obtaining bounds for the smallest eigenvalue. (2) For the "square case" ($N = M$), in [35] the authors proceeded via analyzing X^{-1} directly; this strategy seems out of reach for the nonsquare case. (3) In fact, as in [5, 6], one can prove that the $m(z)$ does satisfy the local Marchenko–Pastur law in the case $\lim_{N\to\infty} d_N = 1$ up to the scale $\eta \gg (N|m_c|)^{-1}$. Note $\eta = (N\Im m_c)^{-1}$ is the scale of individual eigenvalue. At the soft edge (i.e., for largest eigenvalues), it can be shown that $\Im m_c \le |m_c|$, and thus we have a strong estimate on $m(z)$ in the scale which is small enough for estimating the distribution of single eigenvalue. But at the hard edge $\Im m_c \sim |m_c|$, so our method used for estimating m_c at the soft edge cannot be directly applied to the hard edge. It is proved in [5, 6] that the density of eigenvalues satisfy the Marchenko–Pastur law. (Only the case $d_N = 1$ is proved in [5, 6], but the result can be easily extended to the case $\lim_{N\to\infty} d_N = 1$.) For the distribution of the smallest eigenvalues, the only universality result we know is in [35], as mentioned above.

Finally we note that the authors in [29] recently showed a necessary and sufficient condition on the edge universality of Wigner matrices. Based on this, we conjecture that for the edge universality of covariance matrices whose entries are i.i.d., the necessary and sufficient condition on the distribution of the matrix entries is given by $\lim_{s\to\infty} s^4\mathbb{P}(|q_{12}| \geq s) = 0$.

1.5. *Organization of the paper.* In Section 2 we set notation and give some basic definitions. In Section 3 we give statements of the strong version of the Marcenko–Pastur law, rigidity and delocalization of eigenvectors. In Sections 4 and 5, we prove, respectively, the edge and bulk universality results. In Sections 6–8 we give proofs of the strong Marcenko–Pastur law and rigidity of eigenvalues. In Section 7, we state and prove an abstract decoupling lemma for weakly dependent random variables which is used to prove the strong Marcenko–Pastur law.

2. Preliminaries.

Define

$$\begin{aligned} H &:= X^\dagger X, \qquad G(z) := (H - z)^{-1} = \big(X^\dagger X - z\big)^{-1}, \\ m(z) &:= \frac{1}{N}\operatorname{Tr} G(z), \qquad \mathcal{G}(z) := \big(XX^\dagger - z\big)^{-1}. \end{aligned} \tag{2.1}$$

Since the nonzero eigenvalues of $XX^\dagger$ and $X^\dagger X$ are identical and $XX^\dagger$ has $M - N$ more (or $N - M$ less) zero eigenvalues,

$$\operatorname{Tr} G(z) - \operatorname{Tr} \mathcal{G}(z) = \frac{M - N}{z}. \tag{2.2}$$

We will often need to consider minors of X defined below:

DEFINITION 2.1 (Minors). For $\mathbb{T} \subset \{1, \dots, N\}$ 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 $X^{(\mathbb{T})}$,

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

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

$$m^{(\mathbb{T})}(z) := \frac{1}{N}\sum_{i\notin\mathbb{T}} G_{ii}^{(\mathbb{T})}(z). \tag{2.3}$$

We denote the ith column of X by $\mathbf{x}_i$, which is an $M \times 1$ vector. Recall λ_+, λ_- from (2.8). For $z = E + i\eta$, set

$$\kappa := \min\bigl(|\lambda_+ - E|, |E - \lambda_-|\bigr). \tag{2.4}$$

Throughout the paper we will use the letters C, C_ζ, c to denote generic positive constants whose precise value may change from one occurrence to the next but independent of everything else.

Define the Green function of $X^\dagger X$ by

$$G_{ij}(z) = \left(\frac{1}{X^\dagger X - z}\right)_{ij}, \qquad z = E + i\eta, \qquad E \in \mathbb{R}, \qquad \eta > 0. \tag{2.5}$$

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

$$m(z) := \frac{1}{N} \sum_j G_{jj}(z) = \frac{1}{N} \operatorname{Tr} \frac{1}{X^\dagger X - z}. \tag{2.6}$$

We will be working in the regime

$$d := d_N := N/M, \qquad \lim_{N \to \infty} d \neq 0, \infty. \tag{2.7}$$

For our results at the hard-edge (smallest eigenvalues) and for bulk universality results, we will further require that $\lim_{N \to \infty} d_N \neq 1$. Define

$$\lambda_\pm := (1 \pm \sqrt{d})^2. \tag{2.8}$$

The Marchenko–Pastur law [27] (henceforth abbreviated by MP) is given by

$$\varrho_c(x) = \frac{1}{2\pi d} \sqrt{\frac{[(\lambda_+ - x)(x - \lambda_-)]_+}{x^2}}. \tag{2.9}$$

We define $m_c(z)$, $z \in \mathbb{C}$, as the Stieltjes transform of ϱ_c, that is,

$$m_c(z) = \int_{\mathbb{R}} \frac{\varrho_c(x)}{(x - z)} \, \mathrm{d}x. \tag{2.10}$$

The function m_c depends on d and has the closed form expression

$$m_c(z) = \frac{1 - d - z + i\sqrt{(z - \lambda_-)(\lambda_+ - z)}}{2dz}, \tag{2.11}$$

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

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

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

$$\mathfrak{n}(E) := \frac{1}{N} \#\{\lambda_j \geq E\}. \tag{2.12}$$

Let

$$n_c(E) := \int_E^\infty \varrho_c(x)\,\mathrm{d}x \tag{2.13}$$

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

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

$$X = \sum_{\alpha=1}^{M} \sqrt{\lambda_\alpha}\mathbf{u}_\alpha \mathbf{v}_\alpha^\dagger = \sum_{\alpha=1}^{N} \sqrt{\lambda_\alpha}\mathbf{u}_\alpha \mathbf{v}_\alpha^\dagger, \tag{2.14}$$

where $\lambda_1 \ge \lambda_2 \cdots \lambda_{\max\{M,N\}} \ge 0$, $\lambda_\alpha = 0$ for $\min\{N, M\} + 1 \le \alpha \le \max\{N, M\}$, and we let $\mathbf{v}_\alpha = 0$ if $\alpha > N$ and $\mathbf{u}_\alpha = 0$ for $\alpha > M$. We also define the classical location of the eigenvalues with ϱ_c as follows:

$$\int_{\gamma_j}^{\lambda_+} \varrho_c(x)\,\mathrm{d}x = \int_{\gamma_j}^{+\infty} \varrho_c(x)\,\mathrm{d}x = j/N. \tag{2.15}$$

Define the parameter

$$\varphi := (\log N)^{\log\log N}. \tag{2.16}$$

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

$$\mathbf{S}(\zeta) := \{z \in \mathbb{C} : \mathbf{1}_{d>1}(\lambda_-/5) \le E \le 5\lambda_+, \varphi^\zeta N^{-1} \le \eta \le 10(1+d)\}. \tag{2.17}$$

Note that $m_c \sim 1$ in $\mathbf{S}(0)$. Also the cases $d > 1$ and $d < 1$ are not symmetric in the above definition. Actually the proof of universality in the case $d > 1$ is much harder, since it has many zero eigenvalues. This issue can be easily avoided if matrix entries are independent since $X^\dagger X$ and $XX^\dagger$ have the same nonzero eigenvalues. Since, in the strong Marcenko–Pastur law established next section, *we do not assume independence unlike previous works*, the proof is more difficult.

DEFINITION 2.2 (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

$$\mathbb{P}(\Omega^c) \le N^C \exp(-\varphi^\zeta) \tag{2.18}$$

for large enough N.

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

LEMMA 2.3 (Resolvent identities).

$$G_{ii}(z) = \frac{1}{-z - z\langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\mathbf{x}_i\rangle}, \qquad \textit{i.e.}, \langle \mathbf{x}_i, \mathcal{G}^{(i)}(z)\mathbf{x}_i\rangle = \frac{-1}{zG_{ii}(z)} - 1, \tag{2.19}$$

$$G_{ij}(z) = zG_{ii}(z)G_{jj}^{(i)}(z)\langle \mathbf{x}_i, \mathcal{G}^{(ij)}(z)\mathbf{x}_j\rangle, \qquad i \ne j, \tag{2.20}$$

$$G_{ij}(z) = G_{ij}^{(k)}(z) + \frac{G_{ik}(z)G_{kj}(z)}{G_{kk}(z)}, \qquad i, j \ne k. \tag{2.21}$$

PROOF. The proof is straightforward and needs only elementary linear algebra; see Lemma 3.2 of [18]. □

3. Strong Marchenko–Pastur law.

Our goal in this section is to estimate the following quantities:

$$\Lambda_d := \max_k |G_{kk} - m_c|, \qquad \Lambda_o := \max_{k \neq \ell} |G_{k\ell}|, \qquad \Lambda := |m - m_c|, \tag{3.1}$$

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.

For simplicity of exposition, henceforth in this section we focus on the $\lim_N d_N \neq 1$ case. The proof of the distribution of the largest eigenvalue in the case $\lim_{N\to\infty} d_N = 1$ is a simple extension of our proof of the case $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$. Therefore, we will give only a brief discussion at the end of Section 4.

The following is the main result of this section and our main technical tool for establishing universality. It holds for both real and complex valued entries. The proof of the results in this section is given in Sections 6–8.

THEOREM 3.1 (Strong local Marchenko–Pastur law). *Let $X = [x_{ij}]$ with entries x_{ij} satisfying (1.1) and (1.2), and let $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$. 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*

$$\bigcap_{z \in \mathbf{S}(C_\zeta)} \left\{ \Lambda(z) \leq \varphi^{C_\zeta} \frac{1}{N\eta} \right\}. \tag{3.2}$$

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

$$\bigcap_{z \in \mathbf{S}(C_\zeta)} \left\{ \Lambda_o(z) + \Lambda_d \leq \varphi^{C_\zeta} \left(\sqrt{\frac{\Im m_c(z)}{N\eta}} + \frac{1}{N\eta} \right) \right\}. \tag{3.3}$$

(iii) *The smallest nonzero and largest eigenvalues of $X^\dagger X$ satisfy*

$$\lambda_- - N^{-2/3}\varphi^{C_\zeta} \leq \min_{j \leq \min\{M,N\}} \lambda_j \leq \max_j \lambda_j \leq \lambda_+ + N^{-2/3}\varphi^{C_\zeta}. \tag{3.4}$$

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

$$\max_{\alpha : \lambda_\alpha \neq 0} \|\mathbf{v}_\alpha\|_\infty \leq \varphi^{C_\zeta} N^{-1/2}. \tag{3.5}$$

REMARK 3.2. To our knowledge, there are two weaker versions of the above theorem previously established in [14, 20]. In [14] the error term obtained in (3.2) is of order $(N\eta)^{-1/2}/(\kappa+(N\eta)^{-1/2})^{1/2}$ [see (2.4)] and similar comments apply for the results in [20], whereas we need the above stronger estimates for our work, especially for edge universality.

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

THEOREM 3.3 (Rigidity of the eigenvalues of covariance matrix). *Recall* γ_j *in* (2.15). *Let* $X=[x_{ij}]$ *with entries* x_{ij} *satisfying* (1.1) *and* (1.2) *and* $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$. *For any* $1\le j\le N$, *let*

$$\widetilde{j}=\min\{\min\{N,M\}+1-j,j\}.$$

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

$$|\lambda_j-\gamma_j|\le \varphi^{C_\zeta} N^{-2/3}\widetilde{j}^{-1/3} \tag{3.6}$$

and

$$\left|\mathfrak{n}(E)-n_c(E)\right|\le \varphi^{C_\zeta}N^{-1} \tag{3.7}$$

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

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 3.1 and 3.3 can be replaced with a large deviation criteria as will be explained below.

Let us first recall the following large deviation lemma for independent random variables; see [17], Appendix B for a proof.

LEMMA 3.4 (Large deviation lemma). *Suppose* a_i *are 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*

$$\left|\sum_{i=1}^M a_iA_i\right|\le(\log M)^{\rho\zeta\log\log M}\sigma\|A\|, \tag{3.8}$$

$$\left|\sum_{i=1}^M \overline{a_i}B_{ii}a_i-\sum_{i=1}^M\sigma^2B_{ii}\right|\le(\log M)^{\rho\zeta\log\log M}\sigma^2\left(\sum_{i=1}^M|B_{ii}|^2\right)^{1/2}, \tag{3.9}$$

$$\left|\sum_{i\ne j}\overline{a_i}B_{ij}a_j\right|\le(\log M)^{\rho\zeta\log\log M}\sigma^2\left(\sum_{i\ne j}|B_{ij}|^2\right)^{1/2} \tag{3.10}$$

hold with ζ*-high probability.*

REMARK 3.5. When $M \sim N$, equation (3.8) yields that for any $\zeta > 0$, $|\sum_{i=1}^{M} a_i A_i| \le \varphi^{C_\zeta} \sigma \|A\|$ for some $C_\zeta > 0$ with ζ-high probability. Here φ is as defined in (2.16).

Next we extend Theorems 3.1 and 3.3 by relaxing the independence assumption.

THEOREM 3.6. *Let $X = [x_{ij}]$ be a random matrix with $\mathbb{E}(x_{ij}^2) = 1/M$ and $\lim_{N\to\infty} d_N \in (0, \infty) \setminus \{1\}$. Assume that the column vectors of the matrix X are mutually independent. Furthermore, suppose that for any fixed $j \le N$, the random variables defined by $a_i = x_{ij}$, $1 \le i \le M$, satisfy the large deviation bounds* (3.8), (3.9) *and* (3.10), *for any $A_i \in \mathbb{C}$ and $B_{ij} \in \mathbb{C}$ and some $\zeta > 0$. Then the conclusions of Theorems* 3.1 *and* 3.3 *hold for the random matrix X.*

Thus Theorem 3.6 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, and set

$$x_{ij} = \frac{h_{ij}}{\sqrt{\sum_{i=1}^{M} h_{ij}^2}}, \qquad 1 \le i \le M, 1 \le j \le N. \tag{3.11}$$

Thus the entries of the column vector $(x_{1j}, x_{2j}, \dots, x_{Mj})$ are not independent, but exchangeable. Clearly $\mathbb{E}(x_{ij}^2) = \frac{1}{M}$. The random variables x_{ij} given by (3.11) are called self normalized sums and arise in various statistical applications. For instance, the matrix $X = [x_{ij}]$ constructed above is called the correlation matrix (see [22, 33]) and is often preferred in applications such as principal component analysis (PCA) due to the scale invariance of the correlation matrix.

PROOF OF THEOREM 3.6. In the proofs of Theorems 3.1 and 3.3, we use only the large deviation properties of $a_i = x_{ij}$ and the fact that $\mathbb{E}(x_{ij}^2) = 1/M$, instead of independence and sub-exponential decay. Therefore the proofs of Theorems 3.1 and 3.3 in fact yield Theorem 3.6. $\square$

4. Universality of eigenvalues at edge. In this section we give the proof of edge universality stated in Theorem 1.1. For simplicity, we focus on the case $\lim_{N\to\infty} d_N \in (0, \infty) \setminus \{1\}$ first and return to the $\lim_{N\to\infty} d_N = 1$ at the end of this section. The proof is loosely based on Theorem 2.4 of [18] which is an analogous result for Wigner matrices, but in our case *there is a key difference*: the entries within the same column of the matrix $H = X^\dagger X$ are *dependent*. To address this difficulty, we give a novel argument involving the Green function comparison. In the following we consider the largest eigenvalue λ_1, but the same argument applies to the smallest nonzero eigenvalue as well. Also for the rest of this section, let us fix a constant $\zeta > 0$.

For any $E_1 \le E_2$ let

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

denote the number of eigenvalues of the covariance matrix $X^\dagger X$ in $[E_1, E_2]$ where X is a random matrix whose entries satisfy (1.1) and (1.2). By Theorems 3.1 and 3.3 (rigidity of eigenvalues), there exists a positive constant C_ζ such that

$$|\lambda_1 - \lambda_+| \le \varphi^{C_\zeta} N^{-2/3}, \tag{4.1}$$

$$\mathcal{N}\bigl(\lambda_+ - 2\varphi^{C_\zeta} N^{-2/3}, \lambda_+ + 2\varphi^{C_\zeta} N^{-2/3}\bigr) \le \varphi^{2C_\zeta} \tag{4.2}$$

hold with ζ-high probability. Using these estimates, we can assume that the parameter s in (1.4) satisfies

$$-\varphi^{C_\zeta} \le s \le \varphi^{C_\zeta}. \tag{4.3}$$

Set

$$E_\zeta := \lambda_+ + 2\varphi^{C_\zeta} N^{-2/3} \tag{4.4}$$

and for any $E \le E_\zeta$ 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

$$\theta_\eta(x) := \frac{\eta}{\pi(x^2 + \eta^2)} = \frac{1}{\pi} \Im \frac{1}{x - i\eta} \tag{4.5}$$

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. Notice that for any $\ell > 0$,

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

Let us fix $\varepsilon > 0$ and set

$$\eta_1 = N^{-2/3 - 9\varepsilon}. \tag{4.6}$$

LEMMA 4.1. *For any $\varepsilon > 0$, set $\ell_1 := N^{-2/3-3\varepsilon}$. Then for any E satisfying*

$$|E - \lambda_+| \le \tfrac{3}{2} \varphi^{C_\zeta} N^{-2/3}, \tag{4.7}$$

where the constant C_ζ is as in (4.1)–(4.4), the bound

$$\bigl|\operatorname{Tr} \chi_E(H) - \operatorname{Tr} \chi_E * \theta_{\eta_1}(H)\bigr| \le C\bigl(N^{-2\varepsilon} + \mathcal{N}(E - \ell_1, E + \ell_1)\bigr) \tag{4.8}$$

holds with ζ-high probability.

PROOF. From inequalities (4.1), (4.2) above, and (6.13) and (the first line of) (6.17) of [18] we obtain

$$\begin{aligned} &\bigl|\operatorname{Tr} \chi_E(H) - \operatorname{Tr} \chi_E * \theta_{\eta_1}(H)\bigr| \\ &\quad \le C\bigl(\mathcal{N}(E - \ell_1, E + \ell_1) + N^{-5\varepsilon}\bigr) \\ &\qquad + CN\eta_1(E_\zeta - E) \int_{\mathbb{R}} \frac{1}{y^2 + \ell_1^2} \Im m(E - y + i\ell_1)\, \mathrm{d}y. \end{aligned} \tag{4.9}$$

By definition, $\int_{\mathbb{R}} \Im m(E - y + i\ell_1)\,\mathrm{d}y = O(1)$. For any fixed small enough $c > 0$,

$$\int_{|y|\geq\varepsilon} \frac{1}{y^2 + \ell_1^2} \Im m(E - y + i\ell_1)\,\mathrm{d}y = O(c^{-2}).$$

On the interval $|y| \leq c$ we use (3.2), that is,

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

and the elementary estimate $\Im m_c(E - y + i\ell_1) \leq C\sqrt{\ell_1 + |E - y - \lambda_+|}$. Using the definitions of ℓ_1 and η_1 it can be shown that (see inequality (6.18) of [18])

$$N\eta_1(E_\zeta - E) \int_{\mathbb{R}} \frac{1}{y^2 + \ell_1^2} \Im m(E - y + i\ell_1)\,\mathrm{d}y \leq N^{-2\varepsilon}.$$

Now the lemma follows from (4.9). □

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

$$q(x) = 1 \qquad \text{if } |x| \leq 1/9,$$
$$q(x) = 0 \qquad \text{if } |x| \geq 2/9$$

and we assume that $q(x)$ is decreasing for $x \geq 0$. Then we have the following corollary for Lemma 4.1 (which is the counterpart of Corollary 6.2 in [18]):

COROLLARY 4.2. *Let ℓ_1 be as in Lemma 4.1, and set $\ell := \frac{1}{2}\ell_1 N^{2\varepsilon} = \frac{1}{2}N^{-2/3-\varepsilon}$. Then for all E such that*

$$|E - \lambda_+| \leq \varphi^{C_\zeta} N^{-2/3}, \tag{4.10}$$

where the constant C_ζ is as in (4.1)–(4.4), the inequality

$$\operatorname{Tr} \chi_{E+\ell} * \theta_{\eta_1}(H) - N^{-\varepsilon} \leq \mathcal{N}(E, \infty) \leq \operatorname{Tr} \chi_{E-\ell} * \theta_{\eta_1}(H) + N^{-\varepsilon} \tag{4.11}$$

holds with ζ-high probability. Furthermore, there exists $N_0 \in \mathbb{N}$ independent of E such that for all $N \geq N_0$,

$$\begin{aligned} &\mathbb{E}q(\operatorname{Tr} \chi_{E-\ell} * \theta_{\eta_1}(H)) \\ &\qquad \leq \mathbb{P}(\mathcal{N}(E, \infty) = 0) \leq \mathbb{E}q(\operatorname{Tr} \chi_{E+\ell} * \theta_{\eta_1}(H)) + Ce^{-\varphi^{C_\zeta}}. \end{aligned} \tag{4.12}$$

PROOF. For any E satisfying (4.10) we have $E_\zeta - E \gg \ell$ thus $|E - \lambda_+ - \ell|N^{2/3} \leq \frac{3}{2}\varphi^{C_\zeta}$ [see (4.7)]; therefore (4.8) holds for E replaced with $y \in [E - \ell, E]$

as well. We thus obtain

$$
\begin{aligned}
\operatorname{Tr}\chi_E(H) &\le \ell^{-1}\int_{E-\ell}^{E} \mathrm{d}y \operatorname{Tr}\chi_y(H) \\
&\le \ell^{-1}\int_{E-\ell}^{E} \mathrm{d}y \operatorname{Tr}\chi_y * \theta_{\eta_1}(H) \\
&\quad + C\ell^{-1}\int_{E-\ell}^{E} \mathrm{d}y\big[N^{-2\varepsilon} + \mathcal{N}(y-\ell_1, y+\ell_1)\big] \\
&\le \operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H) + CN^{-2\varepsilon} + C\frac{\ell_1}{\ell}\mathcal{N}(E-2\ell, E+\ell)
\end{aligned}
$$

holds with ζ-high probability. From (3.7), (4.10), $\ell_1/\ell = 2N^{-2\varepsilon}$ and $\ell \le N^{-2/3}$, we gather that

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

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

$$
\mathcal{N}(E, E_\zeta) = \operatorname{Tr}\chi_E(H) \le \operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H) + N^{-\varepsilon}.
$$

Using (4.1) we can replace $\mathcal{N}(E, E_\zeta)$ by $\mathcal{N}(E, \infty)$ with a change of probability of at most $O(e^{-\varphi^{C_\zeta}})$. This proves the upper bound of (4.11), and the lower bound can be proved similarly.

When event (4.11) holds, the condition $\mathcal{N}(E, \infty) = 0$ implies that $\operatorname{Tr}\chi_{E+\ell} * \theta_{\eta_1}(H) \le 1/9$. Thus we have

$$
\mathbb{P}\big(\mathcal{N}(E,\infty) = 0\big) \le \mathbb{P}\big(\operatorname{Tr}\chi_{E+\ell} * \theta_{\eta_1}(H) \le 1/9\big) + Ce^{-\varphi^{C_\zeta}}. \tag{4.13}
$$

Together with the Markov inequality, this proves the upper bound in (4.12). For the lower bound, we use

$$
\begin{aligned}
\mathbb{E}q\big(\operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H)\big) &\le \mathbb{P}\big(\operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H) \le 2/9\big) \\
&\le \mathbb{P}\big(\mathcal{N}(E,\infty) \le 2/9 + N^{-\varepsilon}\big) = \mathbb{P}\big(\mathcal{N}(E,\infty) = 0\big),
\end{aligned}
$$

where we used the upper bound from (4.11) and the fact that $\mathcal{N}(E, \infty)$ is an integer. This completes the proof of Corollary 4.2. □

4.1. *Green function comparison theorem.* Let $X^{\mathbf{v}} = [x^{\mathbf{v}}_{ij}]$, with the entries $x^{\mathbf{v}}_{ij}$ satisfying (1.1) and (1.2), $H^{\mathbf{v}} = X^{\mathbf{v}\dagger}X^{\mathbf{v}}$, 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 matrices $X^{\mathbf{w}}$, $H^{\mathbf{w}}$ and the Green function $G^{\mathbf{w}}(z)$ analogously. Define $m^{\mathbf{v}}(z) = \frac{1}{N}\operatorname{Tr} G^{\mathbf{v}}(z)$ and $m^{\mathbf{w}}(z) =$

$\frac{1}{N}\operatorname{Tr} G^{\mathbf{w}}(z)$. The operators $\mathbb{E}^{\mathbf{v}}, \mathbb{E}^{\mathbf{w}}$ denote the expectations under the distributions of $X^{\mathbf{v}}$ and $X^{\mathbf{w}}$, respectively.

Also notice from (4.5) that $\theta_\eta(H) = \frac{1}{\pi}\Im m(i\eta)$. Corollary 4.2 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 with respect to the 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 4.3 (Green function comparison theorem on the edge). *Let $F:\mathbb{R}\to\mathbb{R}$ be a function whose derivatives satisfy*

$$\max_x \big|F^{(\alpha)}(x)\big|\big(|x|+1\big)^{-C_1} \le C_1, \qquad \alpha = 1,2,3,4 \tag{4.14}$$

with some constant $C_1 > 0$. Then there exists $\varepsilon_0 > 0$, $N_0 \in \mathbb{N}$ depending only on C_1 such that for any $\varepsilon < \varepsilon_0$ and $N \ge N_0$ and for any real numbers E, E_1 and E_2 satisfying

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

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

$$\big|\mathbb{E}^{\mathbf{v}} F\big(N\eta\Im m^{\mathbf{v}}(z)\big) - \mathbb{E}^{\mathbf{w}} F\big(N\eta\Im m^{\mathbf{w}}(z)\big)\big| \le CN^{-1/6+C\varepsilon}, \qquad z = E+i\eta \tag{4.15}$$

and

$$\left|\mathbb{E}^{\mathbf{v}} F\left(N\int_{E_1}^{E_2} \mathrm{d}y\, \Im m^{\mathbf{v}}(y+i\eta)\right) - \mathbb{E}^{\mathbf{w}} F\left(N\int_{E_1}^{E_2} \mathrm{d}y\, \Im m^{\mathbf{w}}(y+i\eta)\right)\right| \le CN^{-1/6+C\varepsilon}. \tag{4.16}$$

Theorem 4.3 holds in much greater generality. We state the following extension which can be used to prove (1.5), the generalization of Theorem 1.1. The class of functions F in the following theorem can be enlarged to allow some polynomially increasing functions similar to (4.14). But for our application of the above theorem to prove (1.5), the following form is sufficient.

THEOREM 4.4. *Suppose that the assumptions of Theorem 1.1 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 there exists $\varepsilon_0 > 0$, $N_0 \in \mathbb{N}$ depending only on C_1 such that for any $\varepsilon < \varepsilon_0$ and $N \ge N_0$, there exists $\delta > 0$ such that for any sequence of real numbers $E_k < \cdots < E_1 < E_0$ with $|E_j - \lambda_+| \le N^{-2/3+\varepsilon}$, $j = 0,1,\dots,k$, and $\eta = N^{-2/3-\varepsilon}$*

we have

$$\left|\mathbb{E}^{\mathbf{v}} F\left(N \int_{E_1}^{E_0} \mathrm{d}y\, \Im m(y+i\eta), \ldots, N \int_{E_k}^{E_0} \mathrm{d}y\, \Im m(y+i\eta)\right) - \mathbb{E}^{\mathbf{w}} F(m^{\mathbf{v}} \to m^{\mathbf{w}})\right| \le N^{-\delta}, \tag{4.17}$$

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

PROOF. The proof of Theorem 4.4 is similar to that of Theorem 4.3 and will be omitted. □

Before proceeding further, let us state the following theorem which gives sufficient criteria for proving edge universality for matrix ensembles of the form $Y^\dagger Y$ for various types of data matrices Y. Let $Y_{M\times N} = [y_{ij}]$, $Z_{M\times N} = [z_{ij}]$ be two matrix ensembles, and set $H^Y = Y^\dagger Y$, $H^Z = Z^\dagger Z$. Define the corresponding Green functions $G^Y = (H^Y - z)^{-1}$, $G^Z = (H^Z - z)^{-1}$ and denote their respective empirical Stieltjes transforms by m^Y, m^Z.

THEOREM 4.5. *Assume that the matrices Y, Z satisfy the conclusions stated in items* (i), (ii) *and* (iii) *of Theorem* 3.1. *Furthermore, assume that m^Y and m^Z satisfy the conclusions of Theorems* 4.3 *and* 4.4. *Then the asymptotic eigenvalue distribution of the matrices H^Y, H^Z at the edge are identical; that is, the conclusions of Theorem* 1.1 *are satisfied with $X^{\mathbf{v}} = Y$ and $X^{\mathbf{w}} = Z$.*

REMARK 4.6. Thus our results can be used to show edge universality for cases far beyond covariance matrices. In [33] we use Theorem 4.5 to prove the edge universality of correlation matrices.

PROOF OF THEOREM 4.5. An inspection of the proofs will reveal that, for the arguments used in our application of the Green function comparison method to go through, all we need are the strong MP law and the rigidity of eigenvalues [items (i), (ii) and (iii) of Theorem 3.1] and Theorems 4.3 and 4.4. □

Recall that all discussion so far in this section has been under the assumption that $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$. Now we first prove Theorem 1.1 when $\lim_{N\to\infty} d_N \in (0,\infty)\setminus\{1\}$, assuming that Theorem 4.3 holds and then give the proof of Theorem 4.3. Finally we return to prove Theorem 1.1 for $\lim_{N\to\infty} d_N = 1$ at the end of this section.

PROOF OF THEOREM 1.1 FOR THE CASE $\lim_{N\to\infty} d_N = (0,\infty)\setminus\{1\}$. Define E_ζ as in (4.4) with a constant C_ζ such that (4.1) and (4.2) hold. Therefore we can assume that (4.3) holds for the parameter s. Let $E := \lambda_+ + sN^{-2/3}$ so that $|E-\lambda_+| \le \varphi^{C_\zeta} N^{-2/3}$. Using (4.12), for any sufficiently small $\varepsilon > 0$, we have

$$\mathbb{E}^{\mathbf{w}} q\big(\operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H)\big) \le \mathbb{P}^{\mathbf{w}}\big(\mathcal{N}(E,\infty) = 0\big)$$

with

$$\ell := \tfrac{1}{2} N^{-2/3-\varepsilon}, \qquad \eta_1 := N^{-2/3-9\varepsilon}.$$

Recall that by definition

$$\operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H) = N\frac{1}{\pi}\int_{E-\ell}^{E_\zeta} \Im m(y + i\eta_1)\,\mathrm{d}y.$$

Bound (4.16) applied to the case $E_1 = E - \ell$ and $E_2 = E_\zeta$ shows that there exists $\delta > 0$, such that

$$\mathbb{E}^{\mathbf{v}} q\big(\operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H)\big) \le \mathbb{E}^{\mathbf{w}} q\big(\operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H)\big) + N^{-\delta}. \tag{4.18}$$

Then applying the right-hand side of (4.12) in Lemma 4.2 to the left-hand side of (4.18), we have

$$\mathbb{P}^{\mathbf{v}}\big(\mathcal{N}(E-2\ell,\infty) = 0\big) \le \mathbb{E}^{\mathbf{v}} q\big(\operatorname{Tr}\chi_{E-\ell} * \theta_{\eta_1}(H)\big) + C\exp\big(-c\varphi^{O(1)}\big).$$

Combining these inequalities, we have

$$\mathbb{P}^{\mathbf{v}}\big(\mathcal{N}(E-2\ell,\infty) = 0\big) \le \mathbb{P}^{\mathbf{w}}\big(\mathcal{N}(E,\infty) = 0\big) + 2N^{-\delta} \tag{4.19}$$

for sufficiently small $\varepsilon > 0$ and sufficiently large N. Recalling that $E = \lambda_+ + sN^{-2/3}$, this proves the first inequality of (1.4) and, by switching the roles of $\mathbf{v}$, $\mathbf{w}$, the second inequality of (1.4) as well. This completes the proof of Theorem 1.1. □

PROOF OF THEOREM 4.3. 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, the *key new idea* here is to estimate the successive difference of matrices which differ by a column. Indeed for $1 \le \gamma \le N$, denote by X_γ the random matrix whose jth column is the same as that of $X^{\mathbf{v}}$ if $j < \gamma$ 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 4.7. *For any random matrix X whose entries satisfy* (1.1) *and* (1.2), *if $|E-\lambda_+| \le N^{-2/3+\varepsilon}$ and $N^{-2/3} \gg \eta \ge N^{-2/3-\varepsilon}$ for some $\varepsilon > 0$, then we have*

$$\mathbb{E}F\big(N\eta\Im m(z)\big) - \mathbb{E}F\big(N\eta\Im\widetilde{m}^{(i)}(z)\big) = A\big(X^{(i)}, m_1, m_2\big) + N^{-7/6+C\varepsilon}, \tag{4.20}$$

where the functional $A(X^{(i)}, m_1, m_2)$ depends only on the distribution of $X^{(i)}$ and the first two moments m_1, m_2 of $\sqrt{M}x_{ji} = \sqrt{M}(X)_{ji}$ $(1 \le j \le M)$.

Notice that $X_\gamma^{(\gamma)}$ is equal to $X_{\gamma-1}^{(\gamma)}$. We also have that the first two moments of the entries of $X^{\mathbf{v}}$ and $X^{\mathbf{w}}$ are identical. Thus Lemma 4.7 implies that

$$\mathbb{E}F\left(\eta\Im\operatorname{Tr}\frac{1}{H_{\gamma-1}-z}\right)-\mathbb{E}F\left(\eta\Im\operatorname{Tr}\frac{1}{H_\gamma-z}\right)=O\big(N^{-7/6+C\varepsilon}\big). \tag{4.21}$$

Now the proof of Theorem 4.3 now can be completed via a simple telescoping argument.Thus to finish the proof of Theorem 4.3, all that needs to be shown is Lemma 4.7 which is proven below. □

PROOF OF LEMMA 4.7. Fix $\zeta>0$, $\varepsilon>0$ and, without loss of generality, assume that $i=1$. Recall that $N^{-2/3}\gg\eta\ge N^{-2/3-\varepsilon}$ and $|E-\lambda_+|\le N^{-2/3+\varepsilon}$. First, we claim the following bounds for $G^{(1)}$ and $\mathcal{G}^{(1)}$:

$$\big|\langle\mathbf{x}_1,(\mathcal{G}^{(1)}(z))^2\mathbf{x}_1\rangle\big|\le N^{1/3+C\varepsilon},\qquad z=E+i\eta \tag{4.22}$$

$$\begin{aligned}&\big|[\mathcal{G}^{(1)}(z)]_{ij}\big|\le N^{C\varepsilon},\\&\big|[[\mathcal{G}^{(1)}(z)]^2]_{ij}\big|\le N^{1/3+C\varepsilon},\qquad z=E+i\eta\end{aligned} \tag{4.23}$$

with ζ-high probability for some $C>0$. In the above, $\mathbf{x}_1$ denotes the first column of the matrix X. In (4.23) we allow $i=j$. The proof of these bounds is postponed to the end.

Now using (2.19) and (2.21), we have

$$\begin{aligned}\operatorname{Tr}G-\operatorname{Tr}G^{(1)}+z^{-1}&=\big(G_{11}+z^{-1}\big)+\frac{\langle\mathbf{x}_1,X^{(1)}G^{(1)}G^{(1)}X^{(1)\dagger}\mathbf{x}_1\rangle}{-z-z(\mathbf{x}_1,\mathcal{G}^{(1)}(z)\mathbf{x}_1)}\\&=zG_{11}\langle\mathbf{x}_1,(\mathcal{G}^{(1)})^2(z)\mathbf{x}_1\rangle.\end{aligned} \tag{4.24}$$

Define the quantity B to be

$$B=-zm_c\left[\langle\mathbf{x}_1,\mathcal{G}^{(1)}(z)\mathbf{x}_1\rangle-\left(\frac{-1}{zm_c(z)}-1\right)\right]. \tag{4.25}$$

By (2.19),

$$B=-zm_c\left[\left(\frac{-1}{zG_{11}(z)}-1\right)-\left(\frac{-1}{zm_c(z)}-1\right)\right]=\frac{m_c-G_{11}}{G_{11}}.$$

From (3.3), we obtain that

$$|B|\le N^{-1/3+2\varepsilon}\ll 1, \tag{4.26}$$

with ζ-high probability. Therefore, we have the identity

$$G_{11}=\frac{m_c}{B+1}=m_c\sum_{k\ge0}(-B)^k. \tag{4.27}$$

Define y with the left-hand side of (4.24),

$$y:=\eta\big(\operatorname{Tr}G-\operatorname{Tr}G^{(1)}+z^{-1}\big) \tag{4.28}$$

so that we have

$$N\eta\Im m(z) = N\eta\Im\widetilde{m}^{(1)}(z) + y. \tag{4.29}$$

Using (4.24) and (4.27) we obtain

$$y = \eta z G_{11}\langle \mathbf{x}_1, (\mathcal{G}^{(1)})^2\mathbf{x}_1\rangle = \sum_{k=1}^{\infty} y_k, \qquad y_k := \eta z m_c(-B)^{k-1}\langle \mathbf{x}_1, (\mathcal{G}^{(1)})^2\mathbf{x}_1\rangle.$$

Since z and m_c are $O(1)$, together with (4.22) and (4.26) we see that the bounds

$$|y_k| \le O(N^{-k/3+C\varepsilon}) \quad \text{and} \quad |y| \le O(N^{-1/3+C\varepsilon}) \tag{4.30}$$

hold with ζ-high probability. Consequently, using (4.29), the expansion

$$\begin{aligned} &F(N\eta\Im m(z)) - F(N\eta\Im\widetilde{m}^{(1)}(z)) \\ &\quad = \sum_{k=1}^{3} \frac{1}{k!} F^{(k)}(N\eta\Im\widetilde{m}^{(1)}(z))(\Im y)^k + O(N^{-4/3+C\varepsilon}) \end{aligned} \tag{4.31}$$

holds with ζ-high probability.

Now we estimate each of the three terms ($k = 1, 2, 3$) on the right-hand side of (4.31) individually. First, using (4.30) we obtain that

$$F^{(3)}(N\eta\Im\widetilde{m}^{(1)}(z))(\Im y)^3 = F^{(3)}(N\eta\Im\widetilde{m}^{(1)}(z))(\Im y_1)^3 + O(N^{-4/3+C\varepsilon}) \tag{4.32}$$

holds with ζ-high probability. Moreover, we have

$$\begin{aligned} \mathbb{E}_1(\Im y_1)^3 &= \mathbb{E}_1(\eta z m_c)^3 \langle \mathbf{x}_1, (\mathcal{G}^{(1)})^2\mathbf{x}_1\rangle^3 \\ &= (\eta z m_c)^3 \sum_{k_1,\dots,k_6=1}^{M} \mathbb{E}_1\left(\prod_{i=1}^{6} x_{k_i 1}\right) \prod_{i=1}^{3} [(\mathcal{G}^{(1)})^2]_{k_{2i-1},k_{2i}}, \end{aligned} \tag{4.33}$$

where $\mathbb{E}_1$ is the expectation value with respect to $\mathbf{x}_1$, the first column of X. Recall that m_k denotes the kth moment of $\sqrt{M}x_{j1}$. If there is an index k_i which is different from all the others in the product $\prod_{i=1}^{6} x_{k_i 1}$, then

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

and if each k_i appears exactly twice, then

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

Isolating the above two cases from the sum (4.33), we have

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

where $\mathcal{A}$ denotes the set of indices $k_i \in \{1, 2, \dots, M\}$ such that (1) no k_i appears exactly once in the product $\prod_{i=1}^{6} x_{k_i 1}$ and (2) there is an index k_i which appears at least three times. Clearly, the functional $\widetilde{A}_3(X^{(1)}, m_1, m_2)$ depends only on $X^{(1)}$, m_1 and m_2. Furthermore, it readily follows that

$$\#\mathcal{A} \le CN^2.$$

Then using (4.23) and the bounds on m_k's, it follows that

$$\mathbb{E}_1(\Im y_1)^3 = \widetilde{A}_3\big(X^{(1)}, m_1, m_2\big) + O\big(N^{-2+C\varepsilon}\big). \tag{4.34}$$

It is easy to prove that $|N\eta\Im\widetilde{m}^{(1)}| \le N^{C\varepsilon}$ with ζ-high probability. Using (4.32) and the fact that $\widetilde{m}^{(1)}$ depends only on $X^{(1)}$, we have

$$\mathbb{E}F^{(3)}\big(N\eta\Im\widetilde{m}^{(1)}(z)\big)(\Im y)^3 = A_3\big(X^{(1)}, m_1, m_2\big) + O\big(N^{-4/3+C\varepsilon}\big), \tag{4.35}$$

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

Now we estimate the term with $F^{(2)}$ in (4.31). As in (4.32), we have

$$\begin{aligned} &F^{(2)}\big(N\eta\Im\widetilde{m}^{(1)}(z)\big)(\Im y)^2 \\ &\quad = F^{(2)}\big(N\eta\Im\widetilde{m}^{(1)}(z)\big)\big[(\Im y_1)^2 + 2(\Im y_1)(\Im y_2)\big] + O\big(N^{-4/3+C\varepsilon}\big). \end{aligned} \tag{4.36}$$

By definition,

$$\begin{aligned} &\mathbb{E}_1(\Im y_1)^2 + 2(\Im y_1)(\Im y_2) \\ &\quad = C_1(z)\eta^2\langle \mathbf{x}_1, (\mathcal{G}^{(1)})\mathbf{x}_1\rangle\langle \mathbf{x}_1(\mathcal{G}^{(1)})^2\mathbf{x}_1\rangle^2 + C_2(z)\eta^2\langle \mathbf{x}_1(\mathcal{G}^{(1)})^2\mathbf{x}_1\rangle^2, \end{aligned}$$

where $C_1(z)$, $C_2(z) = O(1)$ are constants which depend only on z and $m_c(z)$. Using the bounds on $\mathcal{G}^{(1)}$ in (4.23), as in (4.34), we have

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

where $\widetilde{A}_2(X^{(1)}, m_1, m_2)$ depends only on the distribution of $X^{(1)}$, m_1 and m_2. Then with (4.36), as in (4.35), we conclude that

$$\mathbb{E}F^{(2)}\big(N\eta\Im\widetilde{m}^{(1)}(z)\big)(\Im y)^2 = A_2\big(X^{(1)}, m_1, m_2\big) + O\big(N^{-4/3+C\varepsilon}\big) \tag{4.37}$$

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

Finally we estimate the term $F^{(1)}$ in (4.31). As in (4.32), we have

$$\begin{aligned} &F^{(1)}\big(N\eta\Im\widetilde{m}^{(1)}(z)\big)(\Im y)^2 \\ &\quad = F^{(1)}\big(N\eta\Im\widetilde{m}^{(1)}(z)\big)[\Im y_1 + \Im y_2 + \Im y_3] + O\big(N^{-4/3+C\varepsilon}\big). \end{aligned} \tag{4.38}$$

A similar argument as in (4.37) and (4.35) yields

$$\mathbb{E}F^{(1)}\big(N\eta\Im\widetilde{m}^{(1)}(z)\big)(\Im y) = A_1\big(X^{(1)}, m_1, m_2\big) + O\big(N^{-4/3+C\varepsilon}\big). \tag{4.39}$$

Inserting (4.39), (4.37) and (4.35) into (4.31), we obtain (4.20). Now to complete the proof of Lemma 4.7 we need to prove (4.22) and (4.23).

For (4.22), using the large deviation lemma (Lemma 3.4), we obtain that for any $\zeta > 0$,

$$\begin{aligned}
\big|(\mathbf{x}_1(\mathcal{G}^{(1)})^2\mathbf{x}_1)\big| &\le \varphi^{C_\zeta}(N^{-1}\operatorname{Tr}|\mathcal{G}^{(1)}|^4)^{1/2} \\
&\le \varphi^{C_\zeta}\left(\frac{1}{N}\sum_\alpha \frac{1}{|\lambda_\alpha^{(1)}-z|^4}\right)^{1/2} \\
&\le \varphi^{C_\zeta}\left(\frac{1}{N\eta^2}\sum_\alpha \frac{1}{|\lambda_\alpha^{(1)}-z|^2}\right)^{1/2} \\
&= \varphi^{C_\zeta}\left(\frac{1}{N\eta^3}\Im m^{(1)}(z)\right)^{1/2}
\end{aligned} \tag{4.40}$$

with ζ-high probability. Then using (3.2) we have (4.22). For (4.23), we note that

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

Comparing with (2.5), we see that the pair $\{\mathcal{G}^{(1)}, (X^{(1)})^\dagger\}$ plays the role of $\{G, X\}$. Since $\sqrt{\frac{M}{N-1}}(X^{(1)})^\dagger$ is just an $(N-1)\times M$ random data matrix, whose entries have variance $(N-1)^{-1}$, the results in (3.3) also hold for $\mathcal{G}^{(1)}$ with slight changes. One can easily obtain that

$$\max_{ij}\big|[\mathcal{G}^{(1)}]_{ij}\big| \le C, \qquad \max_{i\ne j}\big|[\mathcal{G}^{(1)}]_{ij}\big| \le CN^{-1/3+C\varepsilon}$$

with ζ-high probability showing (4.23) and finishing the proof of Lemma 4.7 and consequently we have proved Theorem 4.3. $\square$

PROOF OF THEOREM 1.1 FOR THE CASE $\lim_{N\to\infty} d_N = 1$. Note that this proof holds only for the largest eigenvalues. Without loss of generality, set $1/2 \le d_N \le 2$. First, in the proof of estimates in (3.2) and (3.3) of $m(z)$, we never used the assumption $\lim_{N\to\infty} d_N \ne 1$ directly. We only needed the property of $m_c(z)$ listed in Lemma 6.5. One can easily check that if $\Re z \ge \varepsilon$ for some constant ε independent of N, then $m_c(z)$ also satisfies the properties in Lemma 6.5, even if $\lim_{N\to\infty} d_N = 1$. Therefore for any fixed $\varepsilon > 0$, expressions (3.2) and (3.3) still hold with ζ-high probability if we replace $\bigcap_{z\in\mathbf{S}(C_\zeta)}$ with $\bigcap_{z\in\mathbf{S}(C_\zeta)\text{ and }\Re z\ge\varepsilon}$.

Next, in step 1 in the proof of (3.4), using the estimate of $m(z)$ from (3.2) and (3.3), and properties on $m_c(z)$ in Lemma 6.5, we obtain that for any $\zeta > 0$, there exists some $D_\zeta > 0$ such that

$$\max\{\lambda_j : \lambda_j \le 5\lambda_+\} \le \lambda_+ + N^{-2/3}\varphi^{4D_\zeta} \tag{4.41}$$

holds with ζ-high probability. In the proof, we used only the estimates of $m(z)$ from (3.2) and (3.3) for $z \in \mathbf{S}(C_\zeta)$ and $\Re z \in [\lambda_+, 5\lambda_+]$. Now, using our modified version of (3.2) and (3.3) (obtained by replacing $\bigcap_{z\in\mathbf{S}(C_\zeta)}$ with $\bigcap_{z\in\mathbf{S}(C_\zeta) \text{ and } \Re z\geq\varepsilon}$), (4.41) can be easily extended to the case $\lim_{N\to\infty} d_N = 1$.

Now, we claim that when $\lim_{N\to\infty} d_N = 1$, $\lambda_1 \leq 5\lambda_+$ holds with ζ-high probability. The $M \times N$ data matrix can be considered as a minor of a matrix $\widetilde{X}$, which (1) is an $M \times \widetilde{N}$ matrix with $\lim_{\widetilde{N}\to\infty} \widetilde{N}/M \geq 1 + c$ for some fixed $c > 0$, (2) satisfies the condition of Theorem 1.5. Let λ_1, $\widetilde{\lambda}_1$ be the largest eigenvalue of $X^\dagger X$ and $\widetilde{X}^\dagger \widetilde{X}$. By definition and Theorem 1.5, for small enough c we have

$$\lambda_1 \leq \widetilde{\lambda}_1 \leq 5\lambda_+.$$

Combining the above two statements we obtain that for any $\zeta > 0$, there exists some $D_\zeta > 0$ such that with ζ-high probability

$$\lambda_1 \leq \lambda_+ + N^{-2/3}\varphi^{4D_\zeta}. \tag{4.42}$$

Likewise, step 2 [formula (8.6)] in the proof of (3.4) can also be extended to

$$\bigl|\bigl(\mathfrak{n}(E_1) - \mathfrak{n}(E_2)\bigr) - \bigl(n_c(E_1) - n_c(E_2)\bigr)\bigr| \leq \frac{C(\log N)\varphi^{C_\zeta}}{N}, \qquad E_1, E_2 \in [\lambda_+/2, \lambda_+]$$

since the proof relies only on the estimates of $m(z)$ for $z \in \mathbf{S}(C_\zeta)$ and $\Re z \in [E_1, E_2]$ given by our modified version of (3.2) and (3.3). Together with (4.42), we obtain that for any fixed $\varepsilon > 0$, the rigidity result (3.6) holds for $j \leq (1-\varepsilon)N$, and (3.7) holds for $E \geq \varepsilon$.

Therefore, we conclude that (4.1) and (4.2) hold with ζ-high probability for the case $\lim_{N\to\infty} d_N = 1$. Now to obtain Theorem 1.1 when $\lim_{N\to\infty} d_N = 1$, one needs only to repeat the argument in this section. We note that in the proof of (4.9), we used (3.2), but only for z's such that $\Re z$ is very close to λ_+, which is covered by our modified version of (3.2). Similarly for Corollary 4.2, we used (3.7) but only for E's which are very close to λ_+. Therefore, we obtain Theorem 1.1 in the case $\lim_{N\to\infty} d_N = 1$. $\square$

5. Universality of eigenvalues in bulk.

In this section, our goal is to prove Theorem 1.5. This follows from our key technical result in Section 3 and the usual arguments using the ergodicity of the Dyson Brownian motion mentioned in the Introduction. Throughout this section we assume that $\lim_{N\to\infty} d_N \in (0,\infty) \setminus \{1\}$. Again, we note that 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 $X_0 = X = [x_{ij}]$ satisfies (1.1) and (1.2). For any fixed $t \geq 0$, the distribution of X_t coincides with that of

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

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

$$\mu = \mu_N(\mathrm{d}\mathbf{w}) = \frac{e^{-\mathcal{H}_W^\beta(\mathbf{w})}}{Z_\beta} \,\mathrm{d}\mathbf{w},$$

$$\begin{aligned} \mathcal{H}_W^\beta(\mathbf{w}) = \beta \Bigg[&\sum_{i=1}^N \frac{w_i^2}{2d} - \frac{1}{N} \sum_{i<j} \log|w_j^2 - w_i^2| \\ &- \left(\frac{1}{d} - 1 + \frac{1 - \beta^{-1}}{N}\right) \sum_{i=1}^N \log|w_i| \Bigg] \end{aligned} \tag{5.3}$$

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 (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 a 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

$$\partial_t f_t = \mathcal{L}^W f_t, \tag{5.4}$$

where

$$\begin{aligned} \mathcal{L}^W = L_{\beta,N}^W = \sum_{i=1}^N \frac{1}{2N} \partial_i^2 + \sum_{i=1}^N \Bigg(&-\frac{\beta w_i}{2d} + \frac{\beta}{N} \sum_{j \neq i} \frac{w_i}{w_i^2 - w_j^2} \\ &+ \frac{1}{2}\left(\beta\left(\frac{1}{d} - 1\right) + \frac{\beta - 1}{N}\right) \frac{1}{w_j} \Bigg) \partial_i. \end{aligned} \tag{5.5}$$

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

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

With a slight abuse of notation, 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

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

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

$$\sup_{t \ge N^{-1+\delta+\varepsilon'}} \left| \int_{E-b}^{E+b} \frac{\mathrm{d}E'}{2b} \int_{\mathbb{R}^n} \mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_n \, O(\alpha_1, \ldots, \alpha_n) \frac{1}{\varrho_c(E)^n} (p_{t,N}^{(n)} - p_{\mu,N}^{(n)}) \right.$$
$$\left. \times \left(E' + \frac{\alpha_1}{N\varrho_c(E)}, \ldots, E' + \frac{\alpha_n}{N\varrho_c(E)} \right) \right| \tag{5.8}$$
$$\le C_n N^{2\varepsilon'} [b^{-1} N^{-1+\varepsilon'} + b^{-1/2} N^{-\delta/2}],$$

where $p_{t,N}^{(n)}$ and $p_{\mu,N}^{(n)}$, (5.6) and (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 almost no averaging in E.

PROOF OF THEOREM 5.1. The proof follows from the main result in [14] (Theorem 2.1) which states that the local ergodicity of Dyson Brownian motion (5.8) holds for $t \ge N^{-2\mathfrak{a}+\delta}$ for any $\delta > 0$ provided that there exists an $\mathfrak{a} > 0$ such that

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

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). Condition (5.9) is a simple consequence of (3.6) as long as $\mathfrak{a} < 1/2$.

Strictly speaking, there are four assumptions in the hypothesis of Theorem 2.1 in [14]. Assumptions I and II of Theorem 2.1 in [14] 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 [14] states that the local density of the singular values of X_t in the scale larger than N^{-1+c} for any $c > 0$, is bounded above by a constant. As in [14] this follows from the large deviation estimate (3.2) 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 [14] on the entropy $S_\mu(f_{t_0}) \le CN^m$ 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 [14] and using (3.6), we obtain (5.9) and the proof is finished. $\square$

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

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

where the matrix ξ_G has independent Gaussian random variables with mean 0 and variance 1, $t \sim N^{-\varepsilon}$, 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 1.5, 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 entries of the matrix $\widetilde{\xi}_t$ 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 (rescaled) initial condition $\sqrt{M}X$. 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 [16] 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 $1 \le k \le 3$,

$$\mathbb{E}\sqrt{M}x_{ij}^k = \mathbb{E}[\widetilde{\xi}_t]_{ij}^k, \qquad \left|\mathbb{E}(\sqrt{M}x_{ij})^4 - \mathbb{E}[\widetilde{\xi}_t]_{ij}^4\right| \le Ct \sim N^{-\varepsilon}.$$

Now to finish the proof of Theorem 1.5, it remains only 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 in above sense. To achieve this, as shown for the Wigner matrices [17] (see Sections 8.6–8.13 of [17]), 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 [17], the Green function comparison theorem.

Recall the matrices $X^{\mathbf{v}}$, $X^{\mathbf{w}}$, $H^{\mathbf{v}}$, $H^{\mathbf{w}}$ and the Green functions $G^{\mathbf{v}}$, $G^{\mathbf{w}}$ from Section 4.

THEOREM 5.3. *Assume that the first three moments of $x^{\mathbf{v}}_{ij}$ and $x^{\mathbf{w}}_{ij}$ are identical, that is,*

$$\mathbb{E}(x^{\mathbf{v}}_{ij})^u = \mathbb{E}(x^{\mathbf{w}}_{ij})^u, \qquad 0 \le u \le 3$$

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

$$\big|\mathbb{E}(\sqrt{M}x^{\mathbf{v}}_{ij})^4 - \mathbb{E}(\sqrt{M}x^{\mathbf{w}}_{ij})^4\big| \le N^{-\delta} \tag{5.11}$$

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

$$z^m_j = E^m_j \pm i\eta, \qquad j = 1, \dots, k_i, \qquad m = 1, \dots, n,$$

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

$$\max\Big\{\big|\partial^\alpha F(x_1, \dots, x_n)\big| : \max_j |x_j| \le N^{\varepsilon'}\Big\} \le N^{C_0\varepsilon'}, \tag{5.12}$$

$$\max\Big\{\big|\partial^\alpha F(x_1, \dots, x_n)\big| : \max_j |x_j| \le N^2\Big\} \le N^{C_0} \tag{5.13}$$

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 $N^{-1-\varepsilon} \le \eta \le N^{-1}$ and for any choices of the signs in the imaginary part of z^m_j,

$$\begin{aligned} &\left|\mathbb{E}F\left(\frac{1}{N^{k_1}}\operatorname{Tr}\left[\prod_{j=1}^{k_1} G^{\mathbf{v}}(z^1_j)\right], \dots, \frac{1}{N^{k_n}}\operatorname{Tr}\left[\prod_{j=1}^{k_n} G^{\mathbf{v}}(z^n_j)\right]\right) - \mathbb{E}F(G^{\mathbf{v}} \to G^{\mathbf{w}})\right| \\ &\quad \le C_1 N^{-1/2 + C_1\varepsilon} + C_1 N^{-\delta + C_1\varepsilon}, \end{aligned} \tag{5.14}$$

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 (5.14) as discussed in [17] (Sections 8.6–8.13). The only difference is that in [17], the equivalence is proved for Wigner matrices, but the arguments are easily adapted for covariance matrices. Thus to complete the proof of Theorem 1.5, all that remains is Theorem 5.3 which is proved below.

PROOF OF THEOREM 5.3. The proof is very similar to Lemma 2.3 of [17]. 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,

$$\phi : \{(i, j) : 1 \le i \le M, 1 \le j \le N\} \to \{1, \dots, MN\}$$

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

$$\begin{aligned} [X_\gamma]_{ij} &= [X^{\mathbf{v}}]_{ij}, \qquad \phi(i, j) > \gamma, \\ &= [X^{\mathbf{w}}]_{ij}, \qquad \phi(i, j) \le \gamma. \end{aligned}$$

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

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

First, using the delocalization result (3.5) and the rigidity of eigenvalues (3.6), it is easy to have the following estimate on the matrix elements of the resolvent:

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

with ζ-high probability for any $\zeta > 0$. 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 λ_α, $\mathbf{v}_\alpha$ are the eigenvalues and eigenvectors of H_0. By the delocalization result (3.5), we obtain

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

We write the above sum as

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

where I_k is the set of all α such that

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

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

Recall that $\mathbf{x}_i$ denotes the ith column of X. For $1 \le i \le N$, using straightforward algebra, it is easy to check that

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

From (2.19) we obtain

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

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

Furthermore, from (2.20) it follows that

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

Similarly

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

which implies that

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

Let x_i be the ith row of X. By symmetry, the above identities also hold if one switches $\{G, \mathbf{x}_i\}$ and $\{\mathcal{G}, x_i\}$.

Combining the above identities with (5.15), we obtain the bound

$$\begin{aligned}\max_\gamma \max_{k,l} \max_{\eta\ge N^{-1-\varepsilon}} \max_{\kappa\ge c} &\left|[G_\gamma(z)]_{kl}\right| + \left|[X_\gamma G_\gamma(z)]_{kl}\right| + \left|[G_\gamma X_\gamma^\dagger(z)]_{kl}\right| \\ &+ \left|[X_\gamma G_\gamma X_\gamma^\dagger(z)]_{kl}\right| \le N^{C\varepsilon},\end{aligned} \tag{5.22}$$

with ζ-high probability.

Consider the telescopic sum of differences of expectations

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

Let $E^{(ij)}$ denote the matrix whose matrix elements are zero everywhere except at the (i, j) position, where it is 1, that is, $E^{(ij)}_{k\ell} = \delta_{ik}\delta_{j\ell}$. Fix a $\gamma \ge 1$, 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

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

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 [17] (subsequently used in [10, 16, 18]) which states that the difference of smooth functionals of Green functions of two matrices which differ by a single entry can be bounded above as a function of its first four moments. □

LEMMA 5.4. *Let m_k be the kth moment of $\sqrt{M}x_{ij}^{\mathbf{v}}$, 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 depends only 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

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

Before giving the proof of Lemma 5.4, let us use it to conclude the foregoing 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 $\sqrt{M}x_{ij}^{\mathbf{w}}$ (by hypothesis, the first three moments of $x_{ij}^{\mathbf{w}}$ are identical to those of $x_{ij}^{\mathbf{v}}$). Since $|m_4'-m_4|\le N^{-\delta}$ by hypothesis, we have

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

Using the above estimate and summation over γ yields [see (5.23)]

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

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 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, from the resolvent expansion we obtain

$$R=S+S(V^\dagger X+X^\dagger V+V^\dagger V)S+\cdots+\left[S(V^\dagger X+X^\dagger V+V^\dagger V)\right]^9S + \left[S(V^\dagger X+X^\dagger V+V^\dagger V)\right]^{10}R.$$

Since the matrix V has only at most one nonzero entry, when computing the (k,ℓ) matrix element of the matrix identity above, each term is a finite sum involving matrix elements of S, XS, $SX^\dagger$, $XSX^\dagger$ or R (only for the last term) and $x_{ij}^{\mathbf{v}}$. Using the bound (5.22) for the S matrix elements, the subexponential decay for $x_{ij}^{\mathbf{v}}$ and the trivial bound $|R_{ij}|\le\eta^{-1}$, 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,

$$S = R - R(V^\dagger Q + Q^\dagger V + V^\dagger V)R + \cdots - \bigl[R(V^\dagger Q + Q^\dagger V + V^\dagger V)\bigr]^9 R + O(N^{-4}) \tag{5.26}$$

holds with extremely high probability. Thus we may write

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

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

$$\begin{gathered} |y_k| \le N^{C\varepsilon} N^{-k/2}, \\ \mathbb{E}_{\mathbf{v}} y_{k_1} y_{k_2} \cdots y_{k_n} = N^{-|K|/2} m_{|K|} z_K(Q), \\ \bigl|z_K(Q)\bigr| \le N^{C\varepsilon} \end{gathered} \tag{5.27}$$

for some $z_K(Q)$ depending only on the distribution Q, and the last inequality holds 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

$$\begin{aligned} &\mathbb{E}F\biggl(\frac{1}{N}\operatorname{Tr}\frac{1}{H_{\gamma-1} - z}\biggr) \\ &\quad = \mathbb{E}\sum_{n=0}^{4}\frac{1}{n!}F^{(n)}\biggl(\frac{1}{N}\operatorname{Tr} R\biggr)\biggl(\sum_{k\le 20} y_k\biggr)^n + O(N^{-5/2+C\varepsilon}). \end{aligned} \tag{5.28}$$

From (5.27) we obtain

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

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

$$B = \mathbb{E}\sum_{n=0}^{4}\frac{1}{n!}F^{(n)}\biggl(\frac{1}{N}\operatorname{Tr} R\biggr)\biggl(\sum_{k_1,\dots,k_n : |K|\ge 5, k_i\le 20} N^{-|K|/2} m_{|K|} z_K(Q)\biggr),$$

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

In the above $K = \sum_i k_i$. Now it remains only to prove

$$|B| \le O(N^{-5/2+C\varepsilon}), \qquad \widetilde{A}(Q) \le O(N^{-2+C\varepsilon}).$$

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 $F^{(n)}(\frac{1}{N}\operatorname{Tr} R)$ ($n \le 4$) are bounded by $N^{C\varepsilon}$ with ζ-high probability. Similarly z_K ($k_i \le 20$) 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 the definition of z_K, we know that $F^{(n)}(\frac{1}{N}\operatorname{Tr} R)$ and z_K are bounded by N^C for some $C > 0$ in Ξ_g^c. Since Ξ_g^c has a very small probability by (5.22), we have

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

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

6. A priori bound for the strong local Marcenko–Pastur law.

Our goal in this section is to prove the following weaker form of Theorem 3.1, and in Section 8 we will use this a priori bound to obtain the stronger form as claimed in Theorem 3.1. Throughout this section, we will assume that $\lim_{N\to\infty} d_N \in (0, \infty) \setminus \{1\}$.

THEOREM 6.1. *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 event holds with ζ-high probability*:

$$\bigcap_{z\in\mathbf{S}(C_\zeta)} \left\{\Lambda_d(z) + \Lambda_o(z) \le \varphi^{C_\zeta} \frac{1}{(N\eta)^{1/4}}\right\}. \tag{6.1}$$

6.1. *A roadmap for the reader.* For conveying the key ideas of the computations involved in this section, we first give a brief outline of the proof of Theorem 6.1. For the reader's convenience, we also indicate the corresponding theorems/lemmas in which the estimates mentioned below are proved.

The proof of Theorem 6.1 proceeds via "self-consistent equations" explained below. Let us fix $\zeta > 0$. By definition it follows that

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

where

$$Z_i := z\langle \mathbf{x}_i, \mathcal{G}^{(i)}\mathbf{x}_i\rangle - \frac{z}{M}\operatorname{Tr}\mathcal{G}^{(i)}. \tag{6.2}$$

We will first establish Theorem 6.1 for $\Im z = \eta \sim 1$. For $\eta \sim 1$, the empirical Stieltjes transform satisfies

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

with ζ-high probability (see Lemma 6.10) where

$$\Psi := \sqrt{\frac{\Im m_c + \Lambda}{N\eta}}. \tag{6.3}$$

REMARK 6.2. Notice that when $m_c + \Lambda \le O(1)$, we have

$$\Psi \le O(N\eta)^{-1/2}. \tag{6.4}$$

Consequently, we deduce that for $\eta \sim 1$, the function $m(z)$ satisfies the "self-consistent" equation

$$m(z) = \frac{1}{1-z-d-zdm(z)} + O\big(\varphi^{C_\zeta}\Psi\big) \tag{6.5}$$

with ζ-high probability. Notice that the above equation satisfied by $m(z)$ is nearly identical to the fixed point equation satisfied by the Stieltjes transform of the MP-law, namely

$$m_c(z) + \frac{1}{z-(1-d)+zdm_c(z)} = 0 \tag{6.6}$$

with $\Im m_c > 0$ when $\Im z > 0$. From (6.5) and (6.6), we immediately deduce that (Lemma 6.10) for $\eta \sim 1$, with ζ-high probability,

$$|m - m_c| = \Lambda(z) \le \varphi^{C_\zeta}\frac{1}{(N\eta)^{1/4}}. \tag{6.7}$$

We now use (6.7) to establish Theorem 6.1 for $\eta \sim 1$. To this end, we identify the following "bad sets" (improbable events). For $z \in \mathbf{S}(0)$, define

$$\Omega(z,K) := \Big\{\max\Big\{\Lambda_o(z), \max_i\big|G_{ii}(z) - m(z)\big|, \max_i |Z_i|\Big\} \ge K\Psi(z)\Big\}. \tag{6.8}$$

Then the event (Lemma 6.9)

$$\bigcap_{z\in\mathbf{S}(0),\eta\sim 1} \Omega\big(z,\varphi^{C_\zeta}\big)^c \tag{6.9}$$

holds with ζ-high probability. Here A^c denotes the complement of the set A. The estimate (6.9) coupled with (6.7) immediately establishes Theorem 6.1 for $\eta \sim 1$.

Before proceeding, we notice the following important point. When η is not assumed to be ~ 1, a statement analogous to (6.9) holds with a different assumption. Set

$$\mathbf{B}(z) := \{\Lambda_o(z) + \Lambda_d(z) > (\log N)^{-1}\}, \tag{6.10}$$

$$\Gamma(z, K) := \Omega(z, K)^c \cup \mathbf{B}(z). \tag{6.11}$$

In Lemma 6.8 we show that

$$\bigcap_{z \in \mathbf{S}(C_\zeta)} \Gamma(z, \varphi^{C_\zeta}) \tag{6.12}$$

holds with ζ-high probability. It can also be shown that for $\eta \sim 1$, the event $\mathbf{B}^c(z)$ holds with ζ-high probability.

For proving the result for all $z \in \mathbf{S}(C_\zeta)$ (i.e., for all $\eta \geq \varphi^\zeta N^{-1}$) we proceed as follows. For a function $u(z)$, define its "deviance" to be

$$\mathcal{D}(u)(z) := (u^{-1}(z) + zdu(z)) - ({m_c}^{-1}(z) + zdm_c(z)). \tag{6.13}$$

Clearly, $\mathcal{D}(m_c) = 0$. The plan is to show that $|\mathcal{D}(m)| \approx 0$ and, therefore, $|m_c - m| \approx 0$.

More precisely, suppose that for two numbers L, K satisfying $\varphi^L \geq K^2(\log N)^4$ and for some $A \subset \bigcap_{z\in\mathbf{S}(L)} \Gamma(z, K) \bigcap_{\eta\sim 1} \mathbf{B}^c(z)$ (i.e., A is not in the bad sets of z such that $\Im z \sim 1$) one has the bound

$$|\mathcal{D}(m)(z)| \leq \delta(z) + \infty 1_{\mathbf{B}(z)} \qquad \forall z \in \mathbf{S}(L), \tag{6.14}$$

where $\delta : \mathbb{C} \mapsto \mathbb{R}_+$ is a continuous function, decreasing in $\Im z$ and $|\delta(z)| \leq (\log N)^{-8}$. Then, via a continuity argument, we show in Lemma 6.12 that from (6.14) one indeed has the following stronger conclusion:

$$\Lambda \leq C(\log N)\frac{\delta(z)}{\sqrt{\kappa + \eta + \delta}} \qquad \forall z \in \mathbf{S}(L) \tag{6.15}$$

and $A \subset \bigcap_{z\in\mathbf{S}(L)} \mathbf{B}^c(z)$ [i.e., A is contained in the bad sets of z for all $z \in \mathbf{S}(L)$]. This estimate with a brief additional argument will yield that for large enough C and $z \in \mathbf{S}(\varphi^C)$, we have $\Lambda = o(1)$ and $\Omega(z, \varphi^{C_\zeta})^c$ holds with ζ-high probability. These two conclusions immediately yield Theorem 6.1.

6.2. *Preliminary estimates.* We start with the following elementary lemma whose proof is standard:

LEMMA 6.3. *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*:

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

LEMMA 6.4. *For any z not in the spectrum of $X^\dagger X$, we have*

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

PROOF. Indeed from the SVD decomposition given in (2.14), we have

$$\begin{aligned} X(X^\dagger X - z)^{-1} X^\dagger &= \sum_\alpha \frac{\lambda_\alpha}{\lambda_\alpha - z} \mathbf{u}_\alpha \mathbf{u}_\alpha^\dagger \\ &= \sum_\alpha \left(1 + \frac{z}{\lambda_\alpha - z}\right) \mathbf{u}_\alpha \mathbf{u}_\alpha^\dagger = I + z(XX^\dagger - z)^{-1} \end{aligned}$$

and the lemma is proved. □

We record the following properties of m_c without proof.

LEMMA 6.5 (Properties of m_c). *For $z = E + i\eta \in \mathbf{S}(0)$ we have the following bounds*:

$$|m_c(z)| \sim 1, \qquad |1 - m_c^2(z)| \sim \sqrt{\kappa + \eta}, \tag{6.16}$$

$$\Im m_c(z) \sim \begin{cases} \dfrac{\eta}{\sqrt{\kappa + \eta}}, & \text{if } \kappa \ge \eta \text{ and } |E| \notin [\lambda_-, \lambda_+], \\ \sqrt{\kappa + \eta}, & \text{if } \kappa \le \eta \text{ or } |E| \in [\lambda_-, \lambda_+]. \end{cases} \tag{6.17}$$

Furthermore

$$\frac{\Im m_c(z)}{N\eta} \ge O\left(\frac{1}{N}\right) \quad \text{and} \quad \partial_\eta \frac{\Im m_c(z)}{\eta} \le 0. \tag{6.18}$$

Recall $\mathbf{B}(z)$ from (6.10).

LEMMA 6.6 (Rough bounds of $\Lambda_o^{(\mathbb{T})}$ and $\Lambda_d^{(\mathbb{T})}$). *Fix $\mathbb{T} \subset \{1, 2, \dots, N\}$ such that $|\mathbb{T}| = O(1)$. For $z \in \mathbf{S}(0)$, there exists a constant $C = C_{|\mathbb{T}|}$ such that the following estimates hold in $\mathbf{B}^c(z)$*:

$$\max_{k \notin \mathbb{T}} |G_{kk}^{(\mathbb{T})} - G_{kk}| \le C\Lambda_o^2, \tag{6.19}$$

$$\frac{1}{C} \le |G_{kk}^{(\mathbb{T})}| \le C, \tag{6.20}$$

$$\Lambda_o^{(\mathbb{T})} \le C\Lambda_o. \tag{6.21}$$

PROOF. For $\mathbb{T} = \varnothing$, (6.19) and (6.21) follow from definition, and (6.20) follows from the definition of $\mathbf{B}(z)$ and (6.16). For nonempty $\mathbb{T}$, one can prove the lemma using an induction on $|\mathbb{T}|$. For example, for $|\mathbb{T}| = 1$, using (2.21) we can show that

$$|G_{kk}(z) - G_{kk}^{(\mathbb{T})}(z)| \le C\Lambda_o^2, \tag{6.22}$$

which implies bound (6.19). A similar argument will yield (6.20) and (6.21). □

On the other hand, when $\eta \sim 1$, a bound similar to (6.20) holds without the assumption of $\mathbf{B}^c$.

LEMMA 6.7 (Rough bounds for G_{kk} for $\eta \sim 1$). *Fix $\mathbb{T} \subset \{1, 2, \dots, N\}$ such that $|\mathbb{T}| = O(1)$. For any $z \in \mathbf{S}(0)$ and $\eta \sim 1$, we have the bound*

$$\max_i \bigl|G_{ii}^{(\mathbb{T})}(z)\bigr| \le C$$

for some $C > 0$ and $1 \le i \le N$.

PROOF. Let us show the result first for $|\mathbb{T}| = \varnothing$. By definition,

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

where in the second inequality we have used $|\lambda_\alpha - z| \ge \Im z = \eta$. The claim for a general $\mathbb{T}$ follows similarly. □

Recall from (6.8) and (6.11), the event

$$\Gamma(z, \varphi^{C_\zeta}) = \Omega(z, \varphi^{C_\zeta})^c \cup \mathbf{B}(z).$$

Define the events

$$\begin{aligned} &\Omega_o(z, K) := \{\Lambda_0 \ge K\Psi(z)\}, \\ (6.23)\qquad &\Omega_d(z, K) := \Bigl\{\max_i \bigl|G_{ii}(z) - m(z)\bigr| \ge K\Psi(z)\Bigr\}, \\ &\Omega_Z(z, K) := \Bigl\{\max_i |Z_i| \ge K\Psi(z)\Bigr\}. \end{aligned}$$

Note: $\Omega_d(z, K)$ is defined with m, not m_c. Set

$$\Omega(z, K) = \Omega_o(z, K) \cup \Omega_d(z, K) \cup \Omega_Z(z, K).$$

LEMMA 6.8. *For any $\zeta > 0$ there exists a constant C_ζ such that*

$$\bigcap_{z \in \mathbf{S}(C_\zeta)} \Gamma(z, \varphi^{C_\zeta}) \tag{6.24}$$

holds with ζ-high probability.

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

$$\Gamma(z, \varphi^{C_\zeta}) \tag{6.25}$$

holds with ζ-high probability. It is clear that (6.24) follows from (6.25) and the fact that

$$|\partial_z G_{ij}| \le N^C, \qquad \eta > N^{-1}. \tag{6.26}$$

Note $\Gamma(z, K) = (\Omega_o^c \cup \mathbf{B}) \cap (\Omega_d^c \cup \mathbf{B}) \cap (\Omega_Z^c \cup \mathbf{B})$. First we shall prove that the $\Omega_o^c \cup \mathbf{B}$ holds with ζ-high probability. Using formula (2.20) and the fact that $|G|^2 = G^* G$, we infer that there exists a constant C_ζ such that with ζ-high probability,

$$\begin{aligned}
\Lambda_o(z) &\le C|z| \max_{i\neq j} |\langle \mathbf{x}_i, \mathcal{G}^{(ij)} \mathbf{x}_j \rangle| \le \varphi^{C_\zeta} \frac{|z|}{N} \biggl(\sum_{k,l} |\mathcal{G}_{kl}^{(ij)}|^2\biggr)^{1/2} \\
&\le \varphi^{C_\zeta} \frac{|z|}{N} (\mathrm{Tr} |\mathcal{G}^{(ij)}|^2)^{1/2} \\
&\le \varphi^{C_\zeta} |z| \sqrt{\frac{\Im \operatorname{Tr} \mathcal{G}^{(ij)}}{N^2 \eta}} \qquad \text{in } \mathbf{B}^c(z),
\end{aligned} \tag{6.27}$$

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

$$\operatorname{Tr} G^{(\mathbb{T})}(z) - \operatorname{Tr} \mathcal{G}^{(\mathbb{T})}(z) = \frac{M - N + |\mathbb{T}|}{z}, \tag{6.28}$$

formula (6.19) and $\Im(z^{-1}) = \eta |z|^{-2}$, we deduce that with ζ-high probability

$$\Lambda_o(z) \le \varphi^{C_\zeta} \sqrt{\frac{\Im m_c + \Lambda + \Lambda_o^2}{N\eta} + \frac{1}{N}} \qquad \text{in } \mathbf{B}^c(z).$$

For the above choice of C_ζ, for $z \in \mathbf{S}(3C_\zeta)$, with $\Im m_c \le O(1)$, the bound

$$\Lambda_o(z) \le \varphi^{C_\zeta} \sqrt{\frac{\Im m_c + \Lambda}{N\eta} + \frac{1}{N}} + o(\Lambda_o) \qquad \text{in } \mathbf{B}^c(z) \tag{6.29}$$

holds with ζ-high probability. From (6.29) and (6.18) it follows that $\Omega_o^c \cup \mathbf{B}$ holds with ζ-high probability.

A similar argument using the large deviation lemma will give

$$|Z_i| = |z| \left| \langle \mathbf{x}_i, \mathcal{G}^{(i)} \mathbf{x}_i \rangle - \frac{1}{M} \operatorname{Tr} \mathcal{G}^{(i)} \right| \le |z| \varphi^{C_\zeta} \sqrt{\frac{\Im \operatorname{Tr} \mathcal{G}^{(i)}}{N^2 \eta}} \le \varphi^{C_\zeta} \Psi \qquad \text{in } \mathbf{B}^c(z) \tag{6.30}$$

holds with ζ-high probability implying that

$$\max_i |Z_i| \le \varphi^{C_\zeta} \Psi$$

and therefore $\Omega_Z^c \cup \mathbf{B}$ holds with ζ-high probability.

Finally notice that $\max_i |G_{ii} - m| \le \max_{i \ne j} |G_{ii} - G_{jj}|$. From (2.19) we obtain that

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

holds with ζ-high probability, where the last inequality follows from (6.30), (2.2), (6.19) and (6.20). Thus we have shown that $\Omega_d^c \cup \mathbf{B}$ holds with ζ-high probability, and the lemma is proved. □

On the other hand, in the case of $\eta \sim 1$, a result similar to Lemma 6.8 holds without the assumption of $\mathbf{B}^c$.

LEMMA 6.9. *For any $\zeta > 0$, there exists a constant C_ζ such that the event*

$$\bigcap_{z \in \mathbf{S}(0), \eta \sim 1} \Omega\big(z, \varphi^{C_\zeta}\big)^c \tag{6.31}$$

holds with ζ-high probability.

PROOF. From (6.26) we see that we need only to prove (6.31) for fixed z. First we note in this case, that is, $\eta \sim 1$, we have $\Im m_c \sim 1$ and from Lemma 6.7 we have $\Lambda = O(1)$ and therefore

$$\Psi \sim N^{-1/2}. \tag{6.32}$$

As in (6.27) and Lemma 6.7 we obtain that

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

with ζ-high probability. The estimate for Z_i can be proved as in (6.30) using Lemma 6.7. The estimate for Ω_d [see (6.23)] can also be proved similarly using the identity

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

which follows from Cauchy's interlacing theorem of eigenvalues, that is,

$$\big|m - m^{(i)}\big| \le (N\eta)^{-1} \tag{6.33}$$

and the proof is finished. □

6.3. *Self-consistent equations.* In Section 2, we have bounded Λ_o and $\max_i(G_{ii} - m)$ in terms of m_c, η and Λ in $\mathbf{B}^c$ (we do not need the event $\mathbf{B}^c$ when $\eta \sim 1$). In this subsection, we will give the desired bound for Λ and show that the event $\mathbf{B}^c$ holds with ζ-high probability.

First we give the bound for Λ in the case of $\eta \sim 1$.

LEMMA 6.10. *For any $\zeta > 0$, there exists a constant C_ζ such that*

$$\bigcap_{z \in \mathbf{S}(0), \eta = 10(1+d)} \Lambda(z) \le \varphi^{C_\zeta} N^{-1/4} \tag{6.34}$$

holds with ζ-high probability.

PROOF. By the definition of Z_i given in formulas (6.2) and (2.19),

$$\big(G_{ii}(z)\big)^{-1} = -z - z\frac{1}{M}\operatorname{Tr}\mathcal{G}^{(i)} - Z_i. \tag{6.35}$$

Using (6.28) and (6.33), we obtain that if $\eta \sim 1$,

$$\left| z\frac{1}{M}\operatorname{Tr}\mathcal{G}^{(i)} - zdm(z) + 1 - d \right| \le CN^{-1}. \tag{6.36}$$

Together with $|Z_i| \le \varphi^{C_\zeta}\Psi$ [see (6.31)], estimate (6.36) implies that

$$m(z) = \frac{1}{N}\sum_i \frac{1}{1 - z - d - zdm(z) + Y_i}, \qquad \max_i |Y_i| \le \varphi^{C_\zeta}\Psi \le O\big(\varphi^{C_\zeta}N^{-1/2}\big).$$

It thus follows that $|m(z)| \sim 1$ for $\eta \sim 1$ with ζ-high probability. Then using the fact that $\sum_i (G_{ii} - m) = 0$ we obtain that

$$\sum_i \big(G_{ii}(z)\big)^{-1} = m^{-1}(z) + O\Big(\max_i |G_{ii} - m|\Big)^2.$$

Recall $\mathcal{D}$ in (6.13). Using (6.35), (6.32) and the bound $|Z_i| + |G_{ii} - m| \le \varphi^{C_\zeta}\Psi$ [see (6.31)], and we have

$$\mathcal{D}(m) = \delta(z), \qquad \big|\delta(z)\big| \le \varphi^C\Psi \le O\big(\varphi^C N^{-1/2}\big).$$

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

$$\begin{aligned} m_{1,2} &= \frac{\delta(z) + 1 - d - z \pm i\sqrt{(z - \lambda_{-,\delta})(\lambda_{+,\delta} - z)}}{2dz}, \\ \lambda_{\pm,\delta} &= 1 + d \pm 2\sqrt{d - \delta(z)} - \delta(z), \qquad |\lambda_{\pm,\delta} - \lambda_\pm| = O(\delta). \end{aligned} \tag{6.37}$$

Therefore, we obtain $m = m_1$ or m_2. It is easy to see that $|m_1 - m_2| \ge O(1)$, since $\eta \sim 1$. Since $m(z)$ is continuous with respect to E (for fixed η), $m = m_1$ (say) for $E = 0$ implies that $m = m_1$ for all $E = O(1)$. Using this fact and $\Im m > 0$, we obtain that $m(z) = \frac{\delta(z)+1-d-z+i\sqrt{(z-\lambda_{-,\delta})(\lambda_{+,\delta}-z)}}{2dz}$, and thus we obtain (6.34) and the proof of the lemma is complete. $\square$

Now combining (6.31) with (6.34), we have proved that for any $\zeta > 0$, there exists a constant C_ζ such that, for $\eta = 10(1+d)$, formula (6.1) holds with ζ-high probability. It immediately follows that the event

$$\bigcap_{z \in \mathbf{S}(0), \eta = 10(1+d)} \mathbf{B}^c(z) \tag{6.38}$$

holds with ζ-high probability for any $\zeta > 0$.

Now we prove (6.1) for general $\eta > 0$. Recall the *deviance* function from (6.13), Z_i from (6.2) and set

$$[Z] = \frac{1}{N} \sum_{i=1}^{N} Z_i. \tag{6.39}$$

Recall the set $\mathbf{B}(z)$ from (6.10) and $\Gamma(z, K)$ from Lemma 6.8.

LEMMA 6.11. *Fix $1 \le K \le (\log N)^{-1}(N\eta)^{1/2}$. Then, on the set $\Gamma(z, K)$, we have the bound*

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

PROOF. Using (2.19), (6.19), (6.28) and the definition of m_c, on the set $\Gamma(z, K)$, we obtain a more precise version of (6.35),

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

where $\Omega := \Omega(z, K)$. Then

$$G_{ii}^{-1} - m^{-1} = \mathcal{D}(m) - Z_i + O(K^2\Psi^2) + O(N^{-1}) \qquad \text{in } \mathbf{B}^c \cap \Omega^c \tag{6.40}$$

and averaging over i yields

$$\frac{1}{N}\sum_{i=1}^{N}(G_{ii}^{-1} - m^{-1}) = \mathcal{D}(m) - [Z] + O(K^2\Psi^2) + O(N^{-1}) \qquad \text{in } \mathbf{B}^c \cap \Omega^c.$$

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

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

Together with (6.20) and (6.8), it follows that

$$\frac{1}{N}\sum_{i=1}^{N}(G_{ii}^{-1} - m^{-1}) \le C(K\Psi)^2 \qquad \text{in } \mathbf{B}^c \cap \Omega^c. \tag{6.41}$$

Now the lemma follows from (6.40) and (6.41). □

LEMMA 6.12. *Let $K, L > 0$ be two numbers such that $\varphi^L \ge K^2(\log N)^4$, and let A be an event given by*

$$A \subset \bigcap_{z \in S(L)} \Gamma(z, K) \cap \bigcap_{z \in S(L), \eta = 10(1+d)} \mathbf{B}^c(z). \tag{6.42}$$

Suppose that, in A, we have the bound

$$|\mathcal{D}(m)(z)| \le \delta(z) + \infty 1_{\mathbf{B}(z)} \qquad \forall z \in \mathbf{S}(L),$$

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

$$|m(z) - m_c(z)| = \Lambda(z) \le C(\log N)\frac{\delta(z)}{\sqrt{\kappa + \eta + \delta}} \qquad \forall z \in \mathbf{S}(L) \tag{6.43}$$

holds in A and

$$A \subset \bigcap_{z \in \mathbf{S}(L)} \mathbf{B}^c(z). \tag{6.44}$$

REMARK 6.13. Formula (6.42) says that if $\Im z = 10(1 + d)$, then $A \subset \Omega(z, K)^c$; that is, A is not in the bad sets of such z, and (6.44) implies that A is not in the bad sets of all $z \in \mathbf{S}(L)$. The difficulty in the proof is that our hypothesis yields the bound $\mathcal{D}(m) \le \delta(z)$ only in the set $\mathbf{B}^c$, but we need to prove (6.43) for both $\mathbf{B}$ and $\mathbf{B}^c$.

PROOF OF LEMMA 6.12. Let us first fix E and define the set

$$I_E = \left\{\eta : \Lambda_o(E + i\widehat{\eta}) + \Lambda_d(E + i\widehat{\eta}) \le \frac{1}{\log N}\ \forall \widehat{\eta} \ge \eta, E + i\widehat{\eta} \in \mathbf{S}(L)\right\}.$$

We first prove (6.43) for all $z = E + i\eta$ with $\eta \in I_E$. Define

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

Since δ is a continuous decreasing function of η by assumption, $\delta(E + i\eta) \le (\log N)^{-1}(\kappa + \eta_1)$ for $\eta \ge \eta_1$. Let m_1 and m_2 be the two solutions of the equation $\mathcal{D}(m) = \delta(z)$ as given in (6.37). Note by assumption we do have $|D(m)| \le \delta(z)$

for $z = E + \eta i$ and $\eta \in I_E$, since we are in $\mathbf{B}^c(z)$. Then it can be easily verified that

$$|m_1 - m_2| \geq C\sqrt{\kappa + \eta}, \qquad \eta \geq \eta_1 \tag{6.45}$$
$$\leq C(\log N)\sqrt{\delta(z)}, \qquad \eta \leq \eta_1.$$

The difficulty here is that we do not know which of the two solutions m_1, m_2 is equal to m. However for $\eta = O(1)$, we claim that $m = m_1$. For $\eta = O(1)$, $|m - m_c| = \Lambda \leq \Lambda_d \ll 1$. Also, a direct calculation using (6.37) gives

$$|m_1 - m_c| = C\frac{\delta(z)}{\sqrt{\kappa + \eta}} \ll \frac{1}{\log N}. \tag{6.46}$$

Since $|m_1 - m_2| \geq C\sqrt{\kappa + \eta}$ for $\eta = O(1)$ [see (6.45)], 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 $m_1 \neq m_2$ for $\eta > \eta_1$, it follows that $m = m_1$ for $\eta \geq \eta_1$. Thus for $\eta \geq \eta_1$,

$$\left|m(z) - m_c(z)\right| = \left|m_1(z) - m_c(z)\right| \leq C\frac{\delta(z)}{\sqrt{\kappa + \eta}} \leq C\frac{\delta(z)}{\sqrt{\kappa + \eta + \delta}},$$

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

For $\eta \leq \eta_1$, we take advantage of the fact that the difference $|m_1 - m_2|$ is the same order as the middle term of (6.46). Indeed, for $\eta \leq \eta_1$, if $m = m_2$ (say), then using (6.45),

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

verifying (6.43) for $\eta \in I_E$.

From the above computations for $\eta \sim 1$, we know $I_E \neq \varnothing$. Now we prove that I_E is exactly the desired region, that is, $[\varphi^L N^{-1}, 10(1 + d)]$, and this will verify (6.44). We argue by contradiction. Indeed, assume that $I_E \neq [\varphi^L N^{-1}, 10(1 + d)]$. Let $\eta_0 = \inf I_E$. Then the continuity assumption yields that

$$\Lambda_o(z_0) + \Lambda_d(z_0) = (\log N)^{-1}, \qquad z_0 = E + i\eta_0 \tag{6.47}$$

and thus $\Lambda(z_0) \leq \Lambda_d(z_0) \leq (\log N)^{-1}$. On the other hand, from the calculations done above we deduce that (6.43) holds for $\eta \in I_E$ and thus

$$\Lambda(z_0) \leq (\log N)^{-3}. \tag{6.48}$$

By definition,

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

and therefore

$$\Lambda_o(z_0) + \max_k \left|G_{kk}(z_0) - m(z_0)\right| \leq CK\Psi(z_0).$$

From the assumption $\varphi^L \geq K^2(\log N)^4$, we have $\Psi(z_0) \leq \sqrt{\frac{\Im m_c}{N\eta} + \frac{\Lambda(z_0)}{N\eta}} \ll K^{-1}(\log N)^{-2}$ 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 (6.48) we deduce that

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

which contradicts (6.47), and therefore (6.44) is verified. This completes the proof of the lemma. □

Now we complete the proof of Theorem 6.1.

PROOF OF THEOREM 6.1. From (6.30), Lemmas 6.8 and 6.11, it follows that for any $\zeta > 0$, there exist constants C_ζ, D_ζ and $\widetilde{C}_\zeta$ that

$$\bigl|\mathcal{D}(m)(z)\bigr| \leq \varphi^{\widetilde{C}_\zeta}\Psi + \infty 1_{\mathbf{B}(z)} \qquad \forall z \in \mathbf{S}(C_\zeta)$$

holds on the event A_ζ given by

$$A_\zeta = \bigcap_{z \in \mathbf{S}(C_\zeta)} \Gamma\bigl(z, \varphi^{D_\zeta}\bigr). \tag{6.49}$$

Choosing a larger C_ζ, applying Lemma 6.12 with

$$A = A_\zeta \cap \bigcap_{z \in S(0), \eta = 10(1+d)} \mathbf{B}^c(z)$$

and $\delta(z) = \varphi^{C_\zeta}(N\eta)^{-1/2}$, we obtain that

$$\Lambda(z) \leq \varphi^{C_\zeta}(N\eta)^{-1/4} \qquad \forall z \in \mathbf{S}(C_\zeta) \tag{6.50}$$

holds in A. Furthermore, (6.44) implies that

$$A \subset \bigcap_{z \in \mathbf{S}(C_\zeta)} \mathbf{B}^c(z). \tag{6.51}$$

This observation gives that $\Lambda(z) \leq \Lambda_d(z) = o(1)$ in A and $\Psi \leq C(N\eta)^{-1/2}$ in A. Now since both A_ζ and $\bigcap_{z \in S(0), \eta = 10(1+d)} \mathbf{B}^c(z)$ hold with ζ-high probability [proved, resp., in Lemma 6.8 and (6.38)] it follows that the event A holds with ζ-high probability. Now from the observation (6.51) we see that $\Omega(z, \varphi^{C_\zeta})^c$ holds with ζ-high probability. Together with $\Psi \leq C(N\eta)^{-1/2}$ in A, we obtain (6.1). This completes the proof of Theorem 6.1. □

7. Strong bound on $[Z]$. For proving Theorems 3.1 and 3.3, the key input is the following lemma which gives a much stronger bound on $[Z]$. Throughout this section, we will assume that $\lim_{N\to\infty} d_N \in (0, \infty) \setminus \{1\}$. The following is the main result of this section:

LEMMA 7.1. *Let $K, L > 0$ be such that $\varphi^L \ge K^2(\log N)^4$. Suppose for some event*

$$\Xi \subset \bigcap_{z \in \mathbf{S}(L)} \big(\Gamma(z, K) \cap B^c(z)\big),$$

we have

$$\Lambda(z) \le \widetilde{\Lambda}(z) \qquad \forall z \in \mathbf{S}(L),$$

where $\widetilde{\Lambda}(z)$ is some deterministic number and $\mathbb{P}(\Xi^c) \le e^{-p(\log N)^2}$ with

$$1 \ll p \ll \big(\log(NK)\big)^{-1} \varphi^{L/2}. \tag{7.1}$$

Then there exists Ξ' such that $\mathbb{P}(\Xi') \ge 1 - \frac{1}{2}e^{-p}$, and for any $z \in \mathbf{S}(L)$,

$$\big|[Z]\big| \le Cp^5 K^2 \widetilde{\Psi}^2, \qquad \widetilde{\Psi} := \sqrt{\frac{\Im m_c + \widetilde{\Lambda}}{N\eta}} \qquad \textit{in } \Xi'. \tag{7.2}$$

REMARK 7.2. In the application of the above lemma in Section 8, we will set p_N and $K = O(\varphi^{O(1)})$. This lemma is analogous to Lemma 5.2 in [16] [with $p = O(1)$], Corollary 4.2 in [18] and Lemma 4.1 in [10], which are used in the contexts of Wigner matrices and sparse matrices. The basic idea is to utilize the fact that the entries of Green's function are weakly correlated. But in our work, we give a simple, general lemma (Lemma 7.3) on the cancellation of weakly coupled random variables, which may not have the special structure of Green function, and is thus useful in more general contexts. For instance, our lemma is used for proving universality in non-Hermitian matrices in [5].

7.1. *Abstract decoupling lemma.* First, we are going to introduce the following abstract decoupling lemma[5] which is similar to Theorem 5.6 of [11] and Lemma 4.1 of [18]. However, our lemma as stated here is more general and focuses on weakly coupled random variables and thus is independent of the structure of the matrix ensemble. Due to this generality, it has been useful in other contexts; for instance in [5] where the authors used it in the context of local circular law.

Let $\mathcal{I}$ be a finite set which may depend on N and

$$\mathcal{I}_i \subset \mathcal{I}, \qquad 1 \le i \le N.$$

Let $\{x_\alpha, \alpha \in \mathcal{I}\}$ be a collection of independent random variables and $\mathcal{Z}_1, \dots, \mathcal{Z}_N$ be random variables which are functions of $\{x_\alpha, \alpha \in \mathcal{I}\}$. Let $\mathbb{E}_i$ denote the expectation value operator with respect to $\{x_\alpha, \alpha \in \mathcal{I}_i\}$. Define the commuting projection operators

$$Q_i = 1 - \mathbb{E}_i, \qquad P_i = \mathbb{E}_i, \qquad P_i^2 = P_i,$$
$$Q_i^2 = Q_i, \qquad [Q_i, P_j] = [P_i, P_j] = [Q_i, Q_j] = 0$$

[5]This lemma is joint work with Prof. H. T. Yau, and we thank him for kindly allowing us to include it here.

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

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

We use the notation

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

LEMMA 7.3 (Abstract decoupling lemma). *Let Ξ be an event and p an even integer, which may depend on N. Suppose the 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 $\mathcal{X} < 1$ and $\mathcal{Y}$ such that for any set $A \subset \{1, 2, \dots, N\}$ with $i \in A$ and $|A| \le p$, $Q_A \mathcal{Z}_i$ in Ξ can be written as the sum of two new random variables*

$$\mathbf{1}(\Xi)(Q_A \mathcal{Z}_i) = \mathcal{Z}_{i,A} + \mathbf{1}(\Xi) Q_A \mathbf{1}(\Xi^c) \widetilde{\mathcal{Z}}_{i,A} \tag{7.3}$$

and

$$|\mathcal{Z}_{i,A}| \le \mathcal{Y}\bigl(C_0 \mathcal{X} |A|\bigr)^{|A|}, \qquad |\widetilde{\mathcal{Z}}_{i,A}| \le \mathcal{Y} N^{C_0 |A|}. \tag{7.4}$$

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

$$\max_i |\mathcal{Z}_i| \le \mathcal{Y} N^{C_0}. \tag{7.5}$$

(iii) (Ξ is a high probability event).

$$\mathbb{P}\bigl[\Xi^c\bigr] \le e^{-c_0 (\log N)^{3/2} p}. \tag{7.6}$$

Then, *under assumptions* (i), (ii) *and* (iii) *above*, *we have*

$$\mathbb{E}[\mathcal{Q}\mathcal{Z}]^p \le (Cp)^{4p} \bigl[\mathcal{X}^2 + N^{-1}\bigr]^p \mathcal{Y}^p \tag{7.7}$$

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

The intuition behind Lemma 7.3 is the following. If Z_i are totally independent, that is, $Q_A Z_i = 0$ if $\exists j \in A$ and $i \ne j$, we see that $\sum Z_i$ is less than $\sum |Z_i|$ by a factor $N^{-1/2}$. In this case Z_i depends only on $\{x_\alpha, \alpha \in \mathcal{I}_i\}$. For the general case considered in Theorem 7.3, Z_i also weakly depends on sets $\{x_\alpha, \alpha \in \mathcal{I}_j\}$ for $i \ne j$. Here $Q_j Z_i$ can be considered as the set $\{x_\alpha, \alpha \in \mathcal{I}_j\}$ "acting" on X_i, and $Q_k Q_j Z_i$ the action of $\{x_\alpha, \alpha \in \mathcal{I}_k\}$ on the action of $\{x_\alpha, \alpha \in \mathcal{I}_j\}$ on X_i, so on and so forth. This lemma shows that if the "action" is hierarchical, then indeed $\sum Z_i$ is much less than $\sum |Z_i|$ in the sense of (7.7).

Before we give a proof of Lemma 7.3, we introduce a trivial but useful identity

$$\prod_{i=1}^{n} (x_i + y_i) = \sum_{s=1}^{n+1} \left[\left(\prod_{i=1}^{s-1} x_i \right) y_s \left(\prod_{i=s+1}^{n} (x_i + y_i) \right) \right] \tag{7.8}$$

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

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

For any $1 \le k \le n$, it follows from $\prod_{i=1}^{n}(x_i + y_i) = (x_k + y_k)\prod_{i\neq k}(x_i + y_i)$ and formula (7.8) that

$$\prod_{i=1}^{n}(x_i + y_i) = \sum_{s\neq k, s=1}^{n}(x_k + y_k)\left[\left(\prod_{i\neq k, i=1}^{s-1} x_i\right) y_s \left(\prod_{i\neq k, i=s+1}^{n}(x_i + y_i)\right)\right]. \tag{7.9}$$

PROOF OF LEMMA 7.3. First, by definition, we have

$$\mathbb{E}[\mathcal{Q}\mathcal{Z}]^p = \frac{1}{N^p}\sum_{j_1,\dots,j_p}\mathbb{E}\prod_{\alpha=1}^{p} \mathcal{Q}_{j_\alpha}\mathcal{Z}_{j_\alpha}.$$

For fixed $j_1, \dots, j_p$, let $T_\alpha = \mathcal{Q}_{j_\alpha}\mathcal{Z}_{j_\alpha}$. Now choosing $k = 1$, $x_i = P_{j_1}T_i$ and $y_i = \mathcal{Q}_{j_1}T_i$ in (7.9) (noting that $x_i + y_i = T_i$), we have

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

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

$$\prod_{\alpha=1}^{p} T_\alpha = \sum_{s=2}^{p+1}\prod_{\alpha} P_{A_{\alpha,s}}\mathcal{Q}_{B_{\alpha,s}}T_\alpha.$$

Generalizing, we replace s with s_1 to obtain

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

and

$$A_{\alpha,s_1} = \{j_1 : \alpha < s_1, \alpha \neq 1\}, \qquad B_{\alpha,s_1} = \{j_1 : s_1 = \alpha\}. \tag{7.10}$$

Iterating for $1 \le j_1, j_2, \dots, j_p \le N$, we have

$$\prod_{\alpha=1}^{p} T_\alpha = \sum_{s_1,s_2,\dots,s_p=1}^{p+1}\prod_i \mathbf{1}(s_i \neq i)\prod_{\alpha} P_{A_{\alpha,\mathbf{s}}}\mathcal{Q}_{B_{\alpha,\mathbf{s}}}T_\alpha,$$

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

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

Then it follows that

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

Now to prove (7.7), it remains only to show that for any $\{j_1,\dots,j_p\}$ and $\mathbf{s}=\{s_1,s_2,\dots,s_p\}$ such that $s_i\ne i$, we have

$$\left|\mathbb{E}\prod_\alpha P_{A_{\alpha,\mathbf{s}}}Q_{B_{\alpha,\mathbf{s}}}T_\alpha\right| \le (Cp)^{2p}\mathcal{Y}^p\mathcal{X}^{2t}, \qquad t:=\left|\{j_1,\dots,j_p\}\right|. \tag{7.11}$$

For simplicity, we denote $A_{\alpha,\mathbf{s}}$ and $B_{\alpha,\mathbf{s}}$ by A_α and B_α and denote the characteristic function $\mathbf{1}(\Xi)$ by Ξ. Thus we need to show that

$$\left|\mathbb{E}\prod_\alpha P_{A_\alpha}Q_{B_\alpha}T_\alpha\right| \le (Cp)^{2p}\mathcal{Y}^p\mathcal{X}^{2t}, \qquad t:=\left|\{j_1,\dots,j_p\}\right|. \tag{7.12}$$

Since $T_1=Q_{j_1}T_1$ and the operators P_{A_α} and Q_{B_α} commute, we have

$$\mathbb{E}\prod_\alpha P_{A_\alpha}Q_{B_\alpha}T_\alpha = \mathbb{E}(Q_{j_1}P_{A_1}Q_{B_1}T_1)\left(\prod_{\alpha=2}^{p}(P_{A_\alpha}Q_{B_\alpha})T_\alpha\right). \tag{7.13}$$

Hence we can assume that $j_1\notin\bigcap_{\alpha\ne1}A_\alpha$, and so $1<s_1\le p$ [see (7.10)], $j_1\in\bigcup_{\alpha\ne1}B_\alpha$. Similarly for j_i, we have $j_i\in\bigcup_{\alpha\ne i}B_\alpha$ where $i=2,\dots,p$. Recall that $j_\alpha\notin B_\alpha$. With these two constraints, B_α satisfies the inequality

$$p+t\ge\sum_\alpha\left|B_\alpha\cup\{j_\alpha\}\right|\ge 2t, \qquad t:=\left|\{j_1,\dots,j_p\}\right|. \tag{7.14}$$

Now it remains only to prove (7.12) under condition (7.14). First, we write

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

Using (7.8) with $x=P\Xi Q\mathcal{Z}$ and $y=P\Xi^c Q\mathcal{Z}$ ($x+y=PQ\mathcal{Z}$), we have

$$\begin{aligned}&\mathbb{E}\prod_{\alpha=1}^{p}(P_{A_\alpha}Q_{\widetilde{B}_\alpha}\mathcal{Z}_{j_\alpha})\\ &\quad=\sum_{s=1}^{p+1}\left(\mathbb{E}\prod_{i=1}^{s-1}(P_{A_i}(\Xi)Q_{\widetilde{B}_i}\mathcal{Z}_{j_i})(P_{A_s}(\Xi^c)Q_{\widetilde{B}_s}\mathcal{Z}_{j_s})\prod_{i=s+1}^{p}(P_{A_i}Q_{\widetilde{B}_i}\mathcal{Z}_{j_i})\right).\end{aligned} \tag{7.15}$$

First for $s\le p$, we use the following formula. For any bounded functions f and h,

$$\mathbb{E}\left|h(P\Xi^c Qf)\right|\le\|h\|_\infty\left\|(\Xi^c Qf)\right\|_2\le\sqrt{\mathbb{P}(\Xi^c)}\|f\|_\infty\|h\|_\infty. \tag{7.16}$$

Let

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

By (7.5) and $p \geq 1$, we have

$$|h| \leq \mathcal{Y}^{p-1} N^{Cp}, \qquad |f| \leq \mathcal{Y} N^{C}.$$

Then with (7.6), we have proved that [see (7.15)]

$$\sum_{s=1}^{p} \left(\mathbb{E} \prod_{i=1}^{s-1} (P_{A_i}(\Xi) Q_{\widetilde{B}_i} \mathcal{Z}_{j_i}) (P_{A_s}(\Xi^c) Q_{\widetilde{B}_s} \mathcal{Z}_{j_s}) \prod_{i=s+1}^{p} (P_{A_i} Q_{\widetilde{B}_i} \mathcal{Z}_{j_i}) \right) \\ \leq \mathcal{Y}^p N^{Cp} \exp[-c(\log N)^{3/2} p].$$

Thus the contribution from the above term can be neglected in proving (7.12). It remains only to bound the RHS of (7.15) in the case $s = p+1$; that is, we need to show that

$$\left| \mathbb{E} \prod_{\alpha=1}^{p} (P_{A_\alpha} \Xi Q_{\widetilde{B}_\alpha} \mathcal{Z}_{j_\alpha}) \right| \leq (Cp)^{2p} \mathcal{Y}^p \mathcal{X}^{2t}, \qquad t := |\{j_1, \dots, j_p\}| \tag{7.17}$$

under assumption (7.14). Using (7.3) and (7.8), with $x = P\Xi\mathcal{Z}$ and $y = P\Xi Q \Xi^c \widetilde{\mathcal{Z}}$ we can write the LHS of (7.17) as

$$\mathbb{E} \prod_{\alpha=1}^{p} (P_{A_\alpha} \Xi Q_{\widetilde{B}_\alpha} \mathcal{Z}_{j_\alpha}) \\ = \sum_{s=1}^{p+1} \left(\mathbb{E} \prod_{i=1}^{s-1} (P_{A_i}(\Xi) \mathcal{Z}_{j_i, \widetilde{B}_i}) (P_{A_s}(\Xi) Q_{\widetilde{B}_s} (\Xi^c) \widetilde{\mathcal{Z}}_{j_s, \widetilde{B}_s}) \times \prod_{i=s+1}^{p} (P_{A_i} \Xi Q_{\widetilde{B}_i} \mathcal{Z}_{j_i}) \right). \tag{7.18}$$

Now we repeat the argument for (7.15). For $s \leq p$, one can use the following formula which is similar to (7.16). For any bounded function f and h

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

Let

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

With the assumptions in (7.4) and (7.14), we know the sum over $1 \leq s \leq p$ of RHS of (7.18) is bounded above by

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

which can be neglected in proving (7.17). For the main term, with $s = p + 1$ on the RHS of (7.18), using (7.4) and (7.14), we have

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

and this completes the proof of Lemma 7.3. □

7.2. *A stronger bound on* $[Z]$. In this section we are going to apply Lemma 7.3 to prove a stronger bound on $[Z]$. We note that using (2.19) and (6.2), Z can be written as

$$Z_i = Q_i\left[\frac{-1}{G_{ii}}\right], \qquad Q_i := 1 - P_i, \qquad P_i := \mathbb{E}_{\mathbf{x}_i}. \tag{7.19}$$

LEMMA 7.4. *Let* $\mathcal{Z}_i = (G_{ii})^{-1}$, P_i *and* Q_i *defined as in* (7.19). *We assume that* $\eta = \Im z \ge N^{-C}$ *for some* $C > 0$. *Suppose there exists an even integer* p *and an event* Ξ, *such that* $\mathbb{P}(\Xi^c) \le e^{-p(\log N)^{3/2}}$, *and in* Ξ,

$$\begin{aligned} &\max_i |Q_i\mathcal{Z}_i| \le C\mathcal{Y}\mathcal{X}, \qquad \frac{\Lambda_o(z)}{\min_i |G_{ii}(z)|} \le C\mathcal{X} \ll 1, \\ &\min_i\left|G_{ii}(z)\right| \ge \mathcal{Y}^{-1}, \qquad p \le \frac{C}{(\log N)\mathcal{X}}, \end{aligned} \tag{7.20}$$

where $\mathcal{X} \ll 1$ *and* $\mathcal{Y}$ *are deterministic numbers. Then there exists* Ξ' *with* $\mathbb{P}((\Xi')^c) \le e^{-p}$ *and in* Ξ',

$$\left|\frac{1}{N}\sum_i Q_i\mathcal{Z}_i\right| \le Cp^5(\mathcal{X}^2 + N^{-1})\mathcal{Y}. \tag{7.21}$$

PROOF. We are going to apply Lemma 7.3. The claim given in (7.21) will follow from (7.7) and Markov's inequality. Using the hypothesis, one can easily verify (7.5) and (7.6) in the hypotheses of Lemma 7.3. It remains only to show that for $i \in A \subset \{1, 2, \dots, N\}$ and $|A| \le p$, there exist $\mathcal{Z}_{i,A}$ and $\widetilde{\mathcal{Z}}_{i,A}$ such that

$$\begin{aligned} &\mathbf{1}(\Xi)(Q_A\mathcal{Z}_i) = \mathcal{Z}_{i,A} + \mathbf{1}(\Xi)Q_A\left(\Xi^c\right)\widetilde{\mathcal{Z}}_{i,A}, \\ &\qquad\qquad \mathcal{Z}_{i,A} \le \mathcal{Y}\left(C\mathcal{X}|A|\right)^{|A|}, \widetilde{\mathcal{Z}}_{i,A} \le \mathcal{Y}N^{C|A|} \end{aligned} \tag{7.22}$$

for some $C > 0$. By assumption, formula (7.22) holds when $A = \{i\}$. Thus we assume that $|A| \ge 2$. As in Lemma 5.1 in [11], let $\mathcal{A} = \mathcal{A}(H) = \mathcal{A}(X^\dagger X)$ be a function of $X^\dagger X$, and define

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

for any $S, U \subset \{1, 2, \dots, N\}$. Then we have

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

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

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

In our case,

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

Now we choose

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

It is easy to prove the bound for $\widetilde{\mathcal{Z}}_{i,A}$ in (7.22) using its definition. For bounding $\mathcal{Z}_{i,A}$, it remains only to prove that, for $2 \le |A| \le p_N$,

$$\left|\mathbf{1}(\Xi)\left(\frac{1}{G_{ii}}\right)^{A/\{i\}, A/\{i\}}\right| \le \mathcal{Y}(C\mathcal{X}|A|)^{|A|}. \tag{7.23}$$

To prove this, we first show that for $|\mathbb{T}| \le p$,

$$\max_{i,j \notin \mathbb{T}} |G_{ij}^{(\mathbb{T})}| \le C \max_{i,j} |G_{ij}|, \qquad \min_{i \notin \mathbb{T}} |G_{ii}^{(\mathbb{T})}| \ge c \min_i |G_{ii}| \tag{7.24}$$

with the constants C, c independent of N, i, j. We start from $|\mathbb{T}| = 1$, that is, $\mathbb{T} = \{k\}$. First using (2.21) and the hypotheses of this lemma, we have

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

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

It follows that

$$\max_{i,j \ne k} |G_{ij}^{(k)}| \le (1 + O(\mathcal{X})) \max_{i,j} |G_{ij}|, \qquad \min_{i \ne k} |G_{ii}^{(k)}| \ge (1 - O(\mathcal{X})) \min_i |G_{ii}|.$$

Then using induction on $|\mathbb{T}|$ and the assumption $\mathcal{X}p \ll 1$, we obtain the desired result (7.24).

Now we return to prove (7.23) for the case $|A| = 2$. If $i \ne j$, using (2.21), (7.24) and (7.20), we have

$$\left(\frac{1}{G_{ii}}\right)^{j,j} = (G_{ii})^{-1} - (G_{ii}^{(j)})^{-1} = \frac{-G_{ij}G_{ji}}{G_{ii}G_{jj}G_{ii}^{(j)}} \le O(\mathcal{Y}\mathcal{X}^2).$$

The general case has been proved in Lemma 5.11 of [11] (also see below), which gives that

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

Together with (7.24) and (7.20), we obtain (7.23) for $|A| = 2$.

Finally we need to point out that the definition of $G_{ij}^{(V)}$ $(ij \notin V)$ in [11] is different from the definition in our paper, although they are equivalent. We have

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

and [11] has

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

where $H^{(V)}$ is the minor of H obtained by removing all ith rows and columns of H indexed by $i \in V$. But one can see that if $H = X^{\dagger}X$, then $H^{(V)} = (X^{(V)})^{\dagger}(X^{(V)})$. Thus we finish the proof of Lemma 7.4. □

Finally we give the proof of the main result of this section.

PROOF OF LEMMA 7.1. It is a special case of Lemma 7.4 with $\mathcal{X} = K\widetilde{\Psi}$ and $\mathcal{Y} = C$ for a constant C (possibly large, but independent of N). First, the bound $\max_i |Q_i \mathcal{Z}_i| \leq C\mathcal{Y}\mathcal{X}$ is proved in (6.30). By assumption, if $\Xi \subset \bigcap_{z\in\mathbf{S}(L)}(\Gamma(z, K) \cap \mathbf{B}^c(z))$, then

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

in Ξ. Thus we obtain

$$\frac{\Lambda_o(z)}{\min_i |G_{ii}(z)|} \leq C\mathcal{X} \ll 1, \qquad \min_i |G_{ii}(z)| \geq \mathcal{Y}^{-1}.$$

Furthermore formula (7.1) and $\eta \geq N^{-1}\varphi^L$ [since $z \in S(L)$] imply that $p \leq C((\log N)\mathcal{X})^{-1}$, and the proof of Theorem 7.1 is finished. □

8. Strong Marcenko–Pastur law and rigidity of eigenvalues. In this section, our goal is to prove Theorems 3.1 and 3.3. Throughout this section, we will assume that $\lim_{N\to\infty} d_N \in (0, \infty) \setminus \{1\}$.

Let us first give a brief sketch of the proof strategy for the main technical estimate (3.2). We will prove, by an induction on the exponent τ, that $\Lambda(z) \leq (N\eta)^{-\tau}$ holds modulo logarithmic factors with high probability. Notice that we have already proved this statement for $\tau = 1/4$ in Theorem 6.1. Lemma 6.12 asserts that if this statement is true for some τ, then it also holds for $\frac{1+\tau}{2}$ assuming a bound on $[Z]$. Now, an application of Lemma 7.1 will yield that the required bound for $[Z]$ holds with high probability. Repeating the induction step for $O(\log\log N)$ times, we will obtain that τ is essentially one, implying Theorem 3.1. However, we must keep track of the increasing logarithmic factors and the deteriorating probability estimates of the exceptional sets.

8.1. *Proof of Theorem* 3.1. We start by establishing (3.2) and (3.3).

PROOF OF (3.2) AND (3.3). Without loss of generality, we assume $\zeta \geq 1$. Using Lemma 6.8 and Theorem 6.1, for any $\zeta > 0$, there exists C_ζ such that

$$\Xi_1 \subset \bigcap_{z \in \mathbf{S}(C_\zeta)} \mathbf{B}^c(z) \cap \Gamma(z, C_\zeta) \tag{8.1}$$

holds with $(\zeta + 4)$-high probability. Then from Lemma 6.11, we see that for $z \in \mathbf{S}(3C_\zeta)$,

$$\left|\mathcal{D}(m)(z)\right| \leq \varphi^{2C_\zeta} \Psi^2 + \left|[Z]\right| \qquad \text{in } \Xi_1. \tag{8.2}$$

Let $\Lambda_1 = 1$, so that $\Lambda \leq \Lambda_1$ in Ξ_1. Therefore, we can apply Lemma 7.1 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 replace Ξ_1 by a subset of itself. It follows that

$$p_1 = C\varphi^{\zeta+4}/(\log N)^2.$$

We assume that $C_\zeta \geq 6\zeta$ and therefore (7.1) holds. Then (7.2) gives that, for $z \in \mathbf{S}(3C_\zeta)$, there exists Ξ_2 such that

$$\Xi_2 \subset \Xi_1, \qquad \mathbb{P}(\Xi_2) = 1 - e^{-p_1}$$

and

$$\left|[Z]\right| \leq \varphi^{2C_\zeta + 11\zeta} \Psi_1^2, \qquad \Psi_1 := \sqrt{\frac{\Im m_W + \Lambda_1}{N\eta}} \qquad \text{in } \Xi_2.$$

Since in $\Xi_2 \subset \Xi_1$, by (8.2), $\Lambda \leq \Lambda_1$ and thus $\Psi \leq \Psi_1$ in Ξ_2, and consequently

$$\left|\mathcal{D}(m)(z)\right| \leq \varphi^{2C_\zeta + 11} \frac{\Im m_W + \Lambda_1}{N\eta} \qquad \text{in } \Xi_2. \tag{8.3}$$

Then applying Lemma 6.12, (6.44) shows that, for $z \in \mathbf{S}(3C_\zeta)$,

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

Now the proof proceeds via iterating the above process. Indeed, by choosing

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

we deduce that there exists Ξ_3 such that

$$\Xi_3 \subset \Xi_2, \qquad \mathbb{P}(\Xi_3) = 1 - e^{-p_2}$$

and, for $z \in \mathbf{S}(3C_\zeta)$,

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

We iterate this process K times, $K := \log\log N/(\log 1.9)$. For $k \le K$, we infer that for some

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

where

$$p_k = -\log[1 - \mathbb{P}(\Xi_{k-1})]/(\log N)^2 = C\varphi^{\zeta+4}/(\log N)^{2k} \ge \varphi^{\zeta}$$

and, for $z \in \mathbf{S}(3C_\zeta)$,

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

Note that

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

Thus for $k = K$ and $z \in \mathbf{S}(3C_\zeta)$, the bound

$$\Lambda(z) \le \Lambda_{k+1}(z) \le \varphi^{2C_\zeta+12\zeta}(N\eta)^{-1+(1/2)^K} \le \varphi^{2C_\zeta+12\zeta+1}(N\eta)^{-1} \tag{8.4}$$

holds with ζ-high probability, and this completes the proof of (3.2). Furthermore, since $\Xi_{K+1} \subset \Xi_1$ with (8.1), we obtain (3.3). □

Next we assume (3.4) holds and prove (3.5) first.

PROOF OF (3.5). Using (3.3), we have for any i,

$$\max_{\lambda_-/5 \le E \le 5\lambda_+} \Im G_{ii}(E + i\varphi^{C_\zeta}N^{-1}) \le C. \tag{8.5}$$

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 $\eta = \varphi^{C_\zeta}N^{-1}$, using (8.5), we deduce that for any index α

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

which implies (3.5). Here formula (3.4) guarantees that $\lambda_-/5 \le E \le 5\lambda_+$. □

Now to establish Theorem 3.1, all that remains is the proof of (3.4) which we give below.

PROOF OF (3.4). The proof proceeds via taking the following four steps:

- *Step* 1. For any $\zeta > 0$, there exists some $D_\zeta > 0$ such that

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

  and

$$\min\{\lambda_j : \lambda_j \ge \mathbf{1}_{d>1}\lambda_-/5\} \ge \lambda_- - N^{-2/3}\varphi^{D_\zeta}$$

  hold with ζ-high probability.
- *Step* 2. Recall $\mathfrak{n}(E)$ and $n_c(E)$ from (2.12) and (2.13). We will show that

$$\big|\big(\mathfrak{n}(E_1) - \mathfrak{n}(E_2)\big) - \big(n_c(E_1) - n_c(E_2)\big)\big| \le \frac{C(\log N)\varphi^{C_\zeta}}{N}, \quad E_1, E_2 \in [\mathbf{1}_{d>1}\lambda_-/4, 4\lambda_+], \tag{8.6}$$

  which implies that

$$\#\{j : \lambda_j \notin [\mathbf{1}_{d>1}\lambda_-/5, 5\lambda_+]\} \le \varphi^{C_\zeta}. \tag{8.7}$$

  We note that though we need only (8.7) for (3.4), but (8.6) will be used later to prove Theorem 3.3.
- *Step* 3. Next, using the above two steps we will show that $\max_j \lambda_j \le 5\lambda_+$, with ζ-high probability. This step will imply (3.4) in the case $d < 1$.
- *Step* 4. Finally, we show that, for $d > 1$, that is, $N > M$, we have $\lambda_M \ge \lambda_-/5$, with ζ-high probability.

Step 1 of proof of (3.4). By repeating the iteration in the proof of (3.4) one more time, that is, replacing Λ_1 in (8.3) with Λ_{k+1} in (8.4), we obtain

$$\big|\mathcal{D}(m)(z)\big| \le \varphi^{C_\zeta}\frac{\Im m_c + (1/N\eta)}{N\eta}$$

for some large C_ζ. From (6.43) again, we obtain that for some $D_\zeta \ge 1$

$$\Lambda(z) \le \varphi^{D_\zeta}\frac{\delta}{\sqrt{\kappa + \eta + \delta}}, \qquad \delta := \left(\frac{\Im m_c}{N\eta} + \frac{1}{(N\eta)^2}\right). \tag{8.8}$$

For any E such that $E \ge \lambda_+ + N^{-2/3}\varphi^{4D_\zeta}$, and

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

(thus $\kappa \ge N^{-2/3}\varphi^{4D_\zeta}$), it is easy to check that

$$\kappa \gg \varphi^{D_\zeta}\eta, \qquad N\eta\sqrt{\kappa} \gg \varphi^{D_\zeta}, \qquad \frac{\sqrt{\kappa}}{N\eta^2} \gg 1. \tag{8.9}$$

Using (6.17) and (8.9), we have

$$\Im m_c(z) = C\frac{\eta}{\sqrt{\kappa}}, \tag{8.10}$$

which implies

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

Therefore, $\kappa \ge \delta$. Together with (8.8) and (8.9), we have

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

Combining (8.10) and the last inequality of (8.9) yields

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

and therefore we can conclude that

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

Note that if $\Im m(z) < (2N\eta)^{-1}$ (recall $z = E + i\eta$), then the number of the eigenvalues in the interval $[E - \eta, E + \eta]$ is zero, which is implied by the following observation:

$$\Im m(z) = \frac{1}{N}\sum_{\alpha}\frac{\eta}{(\lambda_\alpha - E)^2 + \eta^2} \ge \sum_{\alpha : |\lambda_\alpha - E| \le \eta}\frac{1}{2N\eta}. \tag{8.11}$$

Since $\Im m(z) \ll \frac{1}{N\eta}$ holds for any $E \ge \lambda_+ + N^{-2/3}\varphi^{4D_\zeta}$, we have proved that for any $\zeta > 0$, there exists some $D_\zeta > 0$ such that

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

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

Step 2 of proof of (3.4). The proof is similar to that of Theorem 2.2 in [18]. The strategy is to translate the information on the Stieltjes transform obtained in Theorem 3.1 to prove (8.6) on the location of the eigenvalues.

In the following lemma, A_1, A_2 represent two numbers with $|A_1 + A_2| \le O(1)$. For any $E_1, E_2 \in [A_1, A_2]$, and $\eta = N^{-1}$ we define

$$f(\lambda) := f_{E_1,E_2,\eta}(\lambda)$$

to be the characteristic function of $[E_1, E_2]$ smoothed on scale η, that is, $f \equiv 1$ on $[E_1 + \eta, E_2 - \eta]$, $f \equiv 0$ on $\mathbb{R} \setminus [E_1, E_2]$ and $|f'| \le C\eta^{-1}$, $|f''| \le C\eta^{-2}$.

LEMMA 8.1. *Let ϱ^Δ be a signed measure on the real line and m^Δ be the Stieltjes transform of ϱ^Δ. Suppose for some positive number U (which may depend on N) we have*

$$\left|m^\Delta(x + iy)\right| \le \frac{CU}{Ny} \qquad \textit{for } y < 1, x \in [A_1, A_2]. \tag{8.12}$$

Then

$$\left|\int_{\mathbb{R}} f_{E_1,E_2,\eta}(\lambda)\varrho^{\Delta}(\lambda)\,\mathrm{d}\lambda\right| \le \frac{CU|\log\eta|}{N}. \tag{8.13}$$

PROOF. For notational simplicity, we drop the Δ superscript in the proof. Let $\chi(y)$ be a smooth cutoff function with support in $[-1, 1]$, with $\chi(y) = 1$ for $|y| \le 1/2$ and with bounded derivatives. Using Helffer–Sjostrand functional calculus, we obtain

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

Since f and χ are real,

$$\begin{aligned}\left|\int f(\lambda)\varrho(\lambda)\,\mathrm{d}\lambda\right| &\le C\int_{\mathbb{R}^2}\big(|f(x)| + |y||f'(x)|\big)|\chi'(y)||m(x+iy)|\,\mathrm{d}x\,\mathrm{d}y \\ &\quad + C\left|\int_{|y|\le\eta}\int yf''(x)\chi(y)\Im m(x+iy)\,\mathrm{d}x\,\mathrm{d}y\right| \\ &\quad + C\left|\int_{|y|\ge\eta}\int_{\mathbb{R}} yf''(x)\chi(y)\Im m(x+iy)\,\mathrm{d}x\,\mathrm{d}y\right|.\end{aligned} \tag{8.14}$$

Using (8.12), the first term can be estimated as

$$\int_{\mathbb{R}^2}\big(|f(x)| + |y||f'(x)|\big)|\chi'(y)||m(x+iy)|\,\mathrm{d}x\,\mathrm{d}y \le CU. \tag{8.15}$$

For the second term on the RHS of (8.14), notice that from (8.12) it follows that, for any $0 < y \le 1$,

$$y|\Im m(x+iy)| \le CU. \tag{8.16}$$

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

$$\operatorname{supp} f'(x) \subset \{|x - E_1| \le \eta\} \cup \{|x - E_2| \le \eta\}, \tag{8.17}$$

we get

$$\left|\int_{|y|\le\eta}\int yf''(x)\chi(y)\Im m(x+iy)\,\mathrm{d}x\,\mathrm{d}y\right| \le CU.$$

Now we integrate the third term in (8.14) by parts first in x, then in y. Then we bound it in absolute value by

$$\begin{aligned}&C\int_{\mathbb{R}}\eta|f'(x)||\Re m(x+i\eta)|\,\mathrm{d}x + C\int_{\mathbb{R}^2} y|f'(x)\chi'(y)\Re m(x+iy)|\,\mathrm{d}x\,\mathrm{d}y \\ &\quad + \frac{C}{\eta}\int_{\eta\le y\le 1}\int_{\operatorname{supp} f'}|\Re m(x+iy)|\,\mathrm{d}x\,\mathrm{d}y.\end{aligned} \tag{8.18}$$

By using (8.12) and (8.17) in the first term, (8.15) in the second and (8.12) in the third, we have

$$(8.18) \le CU + CU\eta^{-1}\int_{\operatorname{supp} f'} \mathrm{d}x \int_{\eta \le y \le 1} \frac{1}{yN}\,\mathrm{d}y \le CU|\log\eta|.$$

This completes the proof of Lemma 8.1. □

We will apply Lemma 8.1 with $[A_1, A_2] \subset [\mathbf{1}_{d>1}\lambda_-/4, 4\lambda_+]$ and the signed measure ϱ^Δ equal to the difference of the empirical density and the MP law,

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

Now we prove that (8.6) holds. By Theorem 3.1, if $y \ge y_0 := \varphi^{C_\zeta}/N$, the assumptions of Lemma 8.1 hold for the difference $m^\Delta = m - m_c$ and $U = \varphi^{C_\zeta}$. For $y \le y_0$, set $z = x + iy$, $z_0 = x + iy_0$ and estimate

$$\begin{aligned} &\bigl|m(z) - m_c(z)\bigr| \\ &\quad \le \bigl|m(z_0) - m_c(z_0)\bigr| + \int_y^{y_0} \bigl|\partial_\eta\bigl(m(x+i\eta) - m_c(x+i\eta)\bigr)\bigr|\,\mathrm{d}\eta. \end{aligned} \tag{8.19}$$

Note that

$$\begin{aligned} \bigl|\partial_\eta m(x+i\eta)\bigr| &= \biggl|\frac{1}{N}\sum_j \partial_\eta G_{jj}(x+i\eta)\biggr| \\ &\le \frac{1}{N}\sum_{jk}\bigl|G_{jk}(x+i\eta)\bigr|^2 = \frac{1}{N\eta}\sum_j \Im G_{jj}(x+i\eta) = \frac{1}{\eta}\Im m(x+i\eta) \end{aligned}$$

and similarly

$$\bigl|\partial_\eta m_c(x+i\eta)\bigr| = \biggl|\int \frac{\varrho_c(s)}{(s-x-i\eta)^2}\,\mathrm{d}s\biggr| \le \int \frac{\varrho_c(s)}{|s-x-i\eta|^2}\,\mathrm{d}s = \frac{1}{\eta}\Im m_c(x+i\eta).$$

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 (8.19) can be bounded by

$$\begin{aligned} &\int_y^{y_0} \frac{\mathrm{d}\eta}{\eta}\bigl[\Im m(x+i\eta) + \Im m_W(x+i\eta)\bigr] \\ &\quad \le y_0\bigl[\Im m(z_0) + \Im m_W(z_0)\bigr]\int_y^{y_0}\frac{\mathrm{d}\eta}{\eta^2}. \end{aligned} \tag{8.20}$$

By definition, $\Im m_c(x+iy_0) \le |m_c(x+iy_0)| \le C$. By the choice of y_0 and Theorem 3.1, we have

$$\Im m(x+iy_0) \le \Im m_c(x+iy_0) + \frac{\varphi^{C_\zeta}}{Ny_0} \le C \tag{8.21}$$

with ζ-high probability for any $\zeta > 0$. Together with (8.20) and (8.19), this proves that (8.12) holds for $y \le y_0$ as well if U is increased to $U = C\varphi^{C_\zeta}$.

The application of Lemma 8.1 shows that, for any $\eta \ge 1/N$,

$$\left|\int_{\mathbb{R}} f_{E_1,E_2,\eta}(\lambda)\varrho(\lambda)\,\mathrm{d}\lambda - \int_{\mathbb{R}} f_{E_1,E_2,\eta}(\lambda)\varrho_c(\lambda)\,\mathrm{d}\lambda\right| \le \frac{C(\log N)\varphi^{C_\zeta}}{N}. \tag{8.22}$$

Using the fact $y \to y\Im m(x+iy)$ is monotone increasing for any $y > 0$, we now use (8.21) to deduce a crude upper bound on the empirical density. Indeed, for any interval $I := [x-\eta, x+\eta]$, with $\eta = 1/N$, we have

$$\mathfrak{n}(x+\eta) - \mathfrak{n}(x-\eta) \le C\eta\Im m(x+i\eta) \le Cy_0\Im m(x+iy_0) \le \frac{C\varphi^{C_\zeta}}{N}. \tag{8.23}$$

Formulas (8.22) and (8.23) yield (8.6) and we have achieved Step 2.

Step 3 of proof of (3.4): now we prove $\lambda_1 \le 5\lambda_+$ holds with ζ-high probability. Note that there is nothing special about the number 5 and below we show that some large K,

$$\lambda_1 \le K\lambda_+$$

with ζ-high probability. Let

$$z = E + i\eta, \qquad E \ge K\lambda_+, \qquad \eta = EN^{-2/3}. \tag{8.24}$$

With (8.6) and choosing $E_1 = \lambda_-$ and $E_2 = K\lambda_+$, we have proved that there are at least $\varphi^{O(1)}$ eigenvalues larger than $K\lambda_+$. Then by definition,

$$\Im m^{(\mathbb{T})} \le \frac{C\eta}{E^2} + \frac{\varphi^{C_\zeta}}{N\eta}, \qquad |\Re m^{(\mathbb{T})}| \le CE^{-1} + \frac{\varphi^{C_\zeta}}{N\eta} \le O(E^{-1}) \tag{8.25}$$

for any index set $\mathbb{T}$ with $|\mathbb{T}| = O(1)$. Now using the large deviation lemma, as in (6.27) and (6.30), we have

$$|Z_i| \le |E|\left(E^{-1}N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta}\right), \qquad \langle \mathbf{x}_i, \mathcal{G}^{(i,j)}\mathbf{x}_j\rangle \le E^{-1}N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta}. \tag{8.26}$$

First we estimate G_{ii}, with (2.19), (6.2) and (6.28),

$$|G_{ii}| = |1 - z - d - zdm^{(i)}(z) - Z_i|^{-1}$$

and

$$\tfrac{1}{2}E^{-1} \le |G_{ii}| \le 2E^{-1}, \tag{8.27}$$

where we used (8.25), (8.26), $\eta = EN^{-2/3}$ and the fact K is large enough. Similarly for G_{ij}, from (2.20) and (6.27) it follows that

$$|G_{ij}| \le E^{-1}\left(\frac{\varphi^{C_\zeta}}{N\eta} + E^{-1}N^{-1/2}\right). \tag{8.28}$$

Furthermore with (2.21) and (2.20),

$$\left|m^{(i)}-m\right|=\frac{1}{N}\left|\sum_j\frac{G_{ji}G_{ij}}{G_{ii}}\right|\le E^{-1}\left|\frac{\varphi^{C_\zeta}}{N\eta}+E^{-1}N^{-1/2}\right|^2.$$

Using these bounds,

$$G_{ii}=\frac{1}{1-z-d-zdm}+O\left(m^{(i)}-m\right)+\frac{Z_i}{(1-z-d-zdm)^2}+E^{-3}O\left(Z_i^2\right)$$

and

$$\begin{aligned} m&=\frac{1}{N}\sum_i G_{ii}\\ &=\frac{1}{1-z-d-zdm}+O\left(E^{-1}\right)\left(\frac{\varphi^{C_\zeta}}{N\eta}+E^{-1}N^{-1/2}\right)^2+O\left(E^{-2}[Z]\right). \end{aligned} \tag{8.29}$$

Since $|\Re(1-z-d-zdm)|\ge|\Im(1-z-d-zdm)|$,

$$\Im\frac{1}{1-z-d-zdm}\le CE^{-2}\eta+\frac{1}{2}\Im m(z). \tag{8.30}$$

Together with (8.29) and (8.26), with ζ-high probability,

$$\begin{aligned} \Im m(z)&\le CE^{-2}\eta+E^{-1}\left(\frac{\varphi^{C_\zeta}}{N\eta}+E^{-1}N^{-1/2}\right)\\ &=\left(\frac{N\eta^2}{E^2}+\frac{\eta N^{1/2}}{E^2}+\frac{\varphi^{C_\zeta}}{E}\right)\frac{1}{N\eta}. \end{aligned} \tag{8.31}$$

If $E\ge N^\varepsilon$ for some $\varepsilon>0$, with ζ-high probability, we have

$$\Im m\ll\frac{1}{N\eta}. \tag{8.32}$$

From the observation made in (8.11), it follows that there are no eigenvalues in the interval $[E-\eta,E+\eta]$ with ζ-high probability, or equivalently there are no eigenvalues larger than N^ε with ζ-high probability.

Now, it only remains to prove (8.32) for $K\lambda_+\le E\le N^\varepsilon$. Using the above result, $\max_j\lambda_j\le N^\varepsilon$, with ζ-high probability we have

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

Therefore, applying (7.21) and (7.19) with $\mathcal{X}=N^\varepsilon(N^{-1/2}+\frac{\varphi^{C_\zeta}}{N\eta})$, $\mathcal{Y}=N^{2\varepsilon}$ and $p=N^\varepsilon$ and by using (8.24), (8.28), (8.26), (8.27), we have

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

Inserting this in (8.29), with (8.25), (8.30), we obtain that the conclusion (8.32) with ζ-high probability for $K\lambda_+ \le E \le N^\varepsilon$. Again using (8.11), we deduce that there are no eigenvalues located in the interval $[K\lambda_+, N^\varepsilon]$ with ζ-high probability. Thus we have achieved Step 3.

Step 4 of proof of (3.4). Now we prove the last component of the proof for (3.4), that is, in the case of $d > 1$ and thus $N > M$, we have $\lambda_M \ge \lambda_-/5$. As remarked earlier, it remains only to prove that for some large K, the following bound holds with ζ-high probability,

$$\lambda_M \ge \lambda_-/K. \tag{8.33}$$

Recall $\mathcal{G} = (XX^\dagger - z)^{-1}$. Let

$$z = E + i\eta, \qquad 0 \le E \le \lambda_-/K, \qquad \eta = N^{-1/2-\varepsilon} \tag{8.34}$$

for some small enough $\varepsilon > 0$. Recall we have proved that among λ_i, $i \le M$, there are at least $\varphi^{O(1)}$ eigenvalues less than λ_-. Then for some $C, c \ge 0$

$$\Im \frac{1}{N} \operatorname{Tr} \mathcal{G}(z) \le C\eta + \frac{\varphi^{C_\zeta}}{N\eta}, \qquad c \le \Re \frac{1}{N} \operatorname{Tr} \mathcal{G}(z) \le C. \tag{8.35}$$

In the above, the term $\frac{\varphi^{C_\zeta}}{N\eta}$ is contributed by these $\varphi^{O(1)}$ eigenvalues. Using Cauchy's interlacing theorem of eigenvalues, it is easy to see that (8.35) also holds for $\mathcal{G}^{(\mathbb{T})}$ for $|\mathbb{T}| = O(1)$. Using the large deviation lemma, with ζ-high probability,

$$\begin{aligned} |Z_i| &\le |z|\left(N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta}\right) \le |z| N^{-1/2+2\varepsilon}, \\ \langle \mathbf{x}_i, \mathcal{G}^{(i,j)} \mathbf{x}_j \rangle &\le N^{-1/2} + \frac{\varphi^{C_\zeta}}{N\eta} \le N^{-1/2+2\varepsilon}. \end{aligned} \tag{8.36}$$

First using (2.19), we obtain,

$$G_{ii} = \left(-z - zd\frac{1}{N} \operatorname{Tr} \mathcal{G}^{(i)}(z) - Z_i\right)^{-1}. \tag{8.37}$$

Then using (8.35) we deduce that with ζ-high probability,

$$c|z|^{-1} \le |G_{ii}| \le C|z|^{-1}. \tag{8.38}$$

Similarly from (2.20), it follows that with ζ-high probability,

$$|G_{ij}| \le |z|^{-1} N^{-1/2+C\varepsilon}. \tag{8.39}$$

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 (8.37),

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

Using the bound [see (8.35)],

$$c|z| \le \left|-z - zd\frac{1}{N}\operatorname{Tr}\mathcal{G}(z)\right| \le C|z|,$$

equation (8.36) and $|m^{(i)} - m| \le (N\eta)^{-1}$, we take the average of G_{ii} and use Taylor expansion to obtain [similar to (8.29)]

$$\begin{aligned} m &= \frac{1}{1 - z - d - zdm(z)} + \delta, \\ \delta &:= |z|^{-1} O\left(\frac{1}{N}\sum_i (m^{(i)} - m) - (Nz)^{-1}\right) \\ &\quad + |z|^{-2} O([Z]) + |z|^{-1} O(N^{-1+C\varepsilon}) \end{aligned} \tag{8.40}$$

with ζ-high probability. Similarly, by estimating the difference $G_{ii} - G_{jj}$, we have

$$|G_{ii} - m| \le |z|^{-1} N^{-1/2+C\varepsilon} \tag{8.41}$$

with ζ-high probability. First for the term $m^{(i)} - m$ in (8.40), using (2.21), (8.38) and (8.41), we have

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

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

$$\frac{1}{N}\sum_i (m^{(i)} - m) = \frac{-1}{N^2}\frac{\operatorname{Tr}[G^2]}{m} + O(|z|N^{-5/2+C\varepsilon})\sum_i |(G^2)_{ii}|. \tag{8.42}$$

Since we have proved that there are at least $\varphi^{O(1)}$ nonzero eigenvalues less than $0.9\lambda_-$, then under (8.34), with ζ-high probability

$$\operatorname{Tr}[G^2] = \sum_\alpha \frac{1}{(\lambda_\alpha - z)^2} = \frac{N-M}{z^2} + O(\varphi^{C_\zeta})\eta^{-2} + O(N). \tag{8.43}$$

These three terms come from zero eigenvalues, small eigenvalues (which are less than $0.9\lambda_-$) and the eigenvalues in the interval $[\lambda_-, \lambda_+]$, respectively. We denote the three terms appearing on the RHS of (8.43) as T_0, T_s and T_n, respectively. Similarly, we have [note that here $z \le O(1)$ is small enough]

$$Nm = \operatorname{Tr}[G] = \frac{N-M}{-z} + O(\varphi^{C_\zeta})\eta^{-1} + O(N) = \frac{N-M}{-z}(1 + O(z)) \tag{8.44}$$

with ζ-high probability and

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

The last bound implies that

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

Together with (8.42), we have

$$\frac{1}{N}\sum_i (m^{(i)} - m) = \frac{-1}{N^2}\frac{\operatorname{Tr}(G^2)}{m} + O\big(|z|^{-1}N^{-3/2+C\varepsilon}\big). \tag{8.45}$$

Dividing (8.43) by Nm [see (8.44)], for $|z|$ small enough, we have

$$\frac{\operatorname{Tr}(G^2)}{Nm} = \frac{-1}{z} + O\big(zN^{2\varepsilon}\big) + O(1). \tag{8.46}$$

Recall δ from (8.40). Now combining (8.45) and (8.46) with (8.40), we obtain

$$\delta \le O\big(|z^{-2}|N^{-3/2+C\varepsilon} + |z^{-1}|N^{-1+C\varepsilon}\big) + |z|^{-2}O\big([Z]\big). \tag{8.47}$$

Now we apply Lemma 7.4 (with $\mathcal{X} = N^{-1/2+C\varepsilon}$, $\mathcal{Y} = C|z|$ and $p = N^\varepsilon$) to estimate $[Z]$. Using Lemma 7.4, (8.36), (8.38) and (8.39), we get

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

Combining the above with (8.47) gives

$$\delta \le O\big(|z^{-2}|N^{-3/2+C\varepsilon} + |z^{-1}|N^{-1+C\varepsilon}\big). \tag{8.48}$$

Using (8.40) and the definition of m_c,

$$m - m_c = \frac{1}{1 - z - d - zdm(z)} - \frac{1}{1 - z - d - zdm_c(z)} + \delta,$$

which implies that

$$\left(\frac{zd}{(1 - z - d - zdm(z))(1 - z - d - zdm_c(z))} - 1\right)(m - m_c) = \delta.$$

As above, we have $c|z| \le |1 - z - d - zdm(z)|, |1 - z - d - zdm_c(z)| \le C|z|$ for all $|z| \le \varepsilon_0$ for a constant ε_0 independent of N. Therefore, we have

$$|m - m_c| \le |z\delta|.$$

Using (8.48), we have

$$|m - m_c| \le O\big(|z^{-1}|N^{-3/2+C\varepsilon} + N^{-1+C\varepsilon}\big) \ll (N\eta)^{-1}.$$

Furthermore, it is easy to prove that

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

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

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

with ζ-high probability. As in (8.11), we have $\lambda_\alpha \notin [E-\eta, E+\eta]$ for $E \in [0, \lambda_-/K]$ with large enough $K = O(1)$ obtaining (8.33). This completes step 4 and we have thus proved (3.4). □

Thus we have verified (3.2), (3.3), (3.4) and (3.5) and have finished the proof of Theorem 3.1.

8.2. *Proof of Theorem* 3.3. We confirm formulas (3.7) and (3.6) separately.

PROOF OF (3.7). Recall (8.6) and the fact that there is no eigenvalue in $(0, \lambda_-/4] \cup [4\lambda_+, +\infty]$. We deduce that

$$\max_{E\in\mathbb{R}} \left|\mathfrak{n}(E) - n_c(E)\right| \le \frac{C(\log N)\varphi^{C_\zeta}}{N} \tag{8.49}$$

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

Now we give the proof of (3.6).

PROOF OF (3.6). The proof is very similar to the one for generalized Wigner matrix obtained in formula (2.25) of [18]. For the reader's sake, we reproduce that argument below. By symmetry, we assume that $1 \le j \le N/2$ and set $E = \gamma_j$, $E' = \lambda_j$. Also $t_N = (\log N)\varphi^{C_\zeta}$ for compactness of notation. From (8.49) we have

$$n_c(E) = \mathfrak{n}(E') = n_c(E') + O(t_N/N). \tag{8.50}$$

Clearly $E \ge \lambda_C := (\lambda_+ + 3\lambda_-)/4$, and using (8.49) we see that $E' \ge \lambda_C$ also holds with ζ-high probability. First, using (3.4) and

$$n_c(x) \sim (\lambda_+ - x)^{3/2} \qquad \text{for } \lambda_C \le x \le \lambda_+, \tag{8.51}$$

or equivalently,

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

we know that (3.6) holds (possibly with a larger constant) if

$$E, E' \ge \lambda_+ - t_N N^{-2/3}.$$

Hence, we can assume that one of E and E' is in the interval $[\lambda_C, \lambda_+ - t_N N^{-2/3}]$. With (8.51), this assumption implies that at least one of $n_c(E)$ and $n_c(E')$ is larger than $t_N^{3/2}/N$. Inserting this information into (8.50), we obtain that both $n_c(E)$ and $n_c(E')$ are positive and

$$n_c(E) = n_c(E')[1 + O(t_N^{-1/2})]$$

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

$$|n_c(E') - n_c(E)| \le C|n_c'(E)||E' - E|. \tag{8.52}$$

Since $n_c'(E) = \varrho_c(E) \sim \sqrt{\kappa}$ and $n_c(E) \sim \kappa^{3/2}$, moreover, by $E = \gamma_j$ we also have $n_c(E) = j/N$, we obtain from (8.50) and (8.52) that

$$|E' - E| \le \frac{C|n_c(E') - n_c(E)|}{n_c'(E)} \le \frac{Ct_N}{Nn_c'(E)} \le \frac{Ct_N}{N(n_c(E))^{1/3}} \le \frac{Ct_N}{N^{2/3} j^{1/3}},$$

which proves (3.6), again with a larger constant. □

We have proved (3.6) and (3.7) and the proof of Theorem 3.3 is complete.

Acknowledgments. The authors would like to thank Horng-Tzer Yau, Antti Knowles, Lazlo Erdös, Paul Bourgade and Alain Pajor for very useful discussions and help. The authors also thank the editorial panel for useful feedback which significantly improved the presentation. Special thanks to Steve Finch for a careful reading.

REFERENCES

[1] BAI, Z. D., SILVERSTEIN, J. W. and YIN, Y. Q. (1988). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166–168. MR0963829

[2] BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

[3] BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** 1–32. MR2988403

[4] BEN AROUS, G. and PÉCHÉ, S. (2005). Universality of local eigenvalue statistics for some sample covariance matrices. *Comm. Pure Appl. Math.* **58** 1316–1357. MR2162782

[5] BOURGADE, P., YAU, H.-T. and YIN, J. (2013). Local circular law for random matrices. *Probab. Theory Related Fields*. To appear. DOI:10.1007/s00440-013-0514-z.

[6] CACCIAPUOTI, C., MALTSEV, A. and SCHLEIN, B. (2013). Local Marchenko–Pastur law at the hard edge of sample covariance matrices. *J. Math. Phys.* **54** 043302, 13. MR3088802

[7] DIENG, M. and TRACY, C. A. (2011). Application of random matrix theory to multivariate statistics. In *Random Matrices, Random Processes and Integrable Systems* 443–507. Springer, New York. MR2858441

[8] DYSON, F. J. (1962). A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.* **3** 1191–1198. MR0148397

[9] EDELMAN, A. (1991). The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. *Linear Algebra Appl*. **159** 55–80. MR1133335

[10] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2012). Spectral statistics of Erdős–Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm*. *Math*. *Phys*. **314** 587–640. MR2964770

[11] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann*. *Probab*. **41** 2279–2375. MR3098073

[12] ERDŐS, L., PÉCHÉ, S., RAMÍREZ, J. A., SCHLEIN, B. and YAU, H.-T. (2010). Bulk universality for Wigner matrices. *Comm*. *Pure Appl*. *Math*. **63** 895–925. MR2662426

[13] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2011). Universality of random matrices and local relaxation flow. *Invent*. *Math*. **185** 75–119. MR2810797

[14] ERDŐS, L., SCHLEIN, B., YAU, H.-T. and YIN, J. (2012). The local relaxation flow approach to universality of the local statistics for random matrices. *Ann*. *Inst*. *Henri Poincaré Probab*. *Statist*. **48** 1–46. MR2919197

[15] ERDŐS, L. and YAU, H.-T. (2012). A comment on the Wigner–Dyson–Mehta bulk universality conjecture for Wigner matrices. *Electron*. *J*. *Probab*. **17** 1–5. MR2915664

[16] ERDŐS, L., YAU, H.-T. and YIN, J. (2011). Universality for generalized Wigner matrices with Bernoulli distribution. *J*. *Comb*. **2** 15–81. MR2847916

[17] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Bulk universality for generalized Wigner matrices. *Probab*. *Theory Related Fields* **154** 341–407. MR2981427

[18] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv*. *Math*. **229** 1435–1515. MR2871147

[19] FELDHEIM, O. N. and SODIN, S. (2010). A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom*. *Funct*. *Anal*. **20** 88–123. MR2647136

[20] GÖTZE, F. and TIKHOMIROV, A. (2004). Rate of convergence in probability to the Marchenko–Pastur law. *Bernoulli* **10** 503–548. MR2061442

[21] HARDING, M. C. (2008). Explaining the single factor bias of arbitrage pricing models in finite samples. *Econom*. *Lett*. **99** 85–88. MR2401243

[22] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann*. *Statist*. **29** 295–327. MR1863961

[23] JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians* 307–333. Eur. Math. Soc., Zürich. MR2334195

[24] JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann*. *Statist*. **36** 2638–2716. MR2485010

[25] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm*. *Pure Appl*. *Math*. **66** 1663–1749.

[26] KNOWLES, A. and YIN, J. (2013). Eigenvector distribution of Wigner matrices. *Probab*. *Theory Related Fields* **155** 543–582. MR3034787

[27] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb*. *Math*. **4** 457–483.

[28] ONATSKI, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070

[29] OON LEE, J. and YIN, J. (2012). A necessary and sufficient condition for edge universality of Wigner matrices. Preprint. Available at arXiv:1206.2251.

[30] PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population structure and eigenanalysis. *PLoS Genet*. **2** e190.

[31] PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab*. *Theory Related Fields* **143** 481–516. MR2475670

[32] PÉCHÉ, S. (2012). Universality in the bulk of the spectrum for complex sample covariance matrices. *Ann*. *Inst*. *Henri Poincaré Probab*. *Statist*. **48** 80–106. MR2919199

[33] PILLAI, N. S. and YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042

[34] SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444

[35] TAO, T. and VU, V. (2010). Random matrices: The distribution of the smallest singular values. *Geom. Funct. Anal.* **20** 260–297. MR2647142

[36] TAO, T. and VU, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449

[37] TAO, T. and VU, V. (2011). Random matrices: Universality of local eigenvalue statistics. *Acta Math.* **206** 127–204. MR2784665

[38] TAO, T. and VU, V. (2012). Random covariance matrices: Universality of local statistics of eigenvalues. *Ann. Probab.* **40** 1285–1315. MR2962092

[39] WANG, K. (2012). Random covariance matrices: Universality of local statistics of eigenvalues up to the edge. *Random Matrices: Theory Appl.* **1** 1150005, 24. MR2930383

[40] YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
1 OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138
USA
E-MAIL: pillai@stat.harvard.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN–MADISON
480 LINCOLN DR.
MADISON, WISCONSIN 53706
USA
E-MAIL: jyin@math.wisc.edu