The Local Marchenko-Pastur Law for Sparse Covariance Matrices

Ben Adlam

Department of Mathematics, Harvard University
Cambridge MA 02138, USA
adlam@college.harvard.edu

March 25, 2013

We review the resolvent method for proving local laws of random matrix ensembles by coving all the necessary prerequisites and outlining the general strategy. With this machinery we consider $N \times N$ sparse covariance matrices $H := X^*X$, where the moments of the entries of $X = (x_{ij})$ decay slowly at a rate inversely proportional to the sparseness parameter q. We calculate the higher moments of the entries of H to their leading order and reproduce a large deviation estimate for quadratic forms involving the random variables x_{ij}. After rescaling H, so its bulk eigenvalues are typically of order one, we prove the new result that, as long as q grows at least logarithmically with N, the density of the ensembles' eigenvalues is given by the Marchenko-Pastur law for spectral windows of length larger than N^{-1} (up to logarithmic corrections). Finally, we give a new formal calculation to characterize the diffusion process undergone by the eigenvalues of a covariance matrix $H := X^*TX$ when the entries of X are undergoing independent Brownian motions and show that the process is only independent of the eigenvectors when $T = c\mathbb{1}$ for some $c \in \mathbb{R}$.

Contents

1. Introduction

A *random matrix* is a random variable that takes values in the space of matrices [4]. The most common example of a random matrix is more concrete than the above definition: construct a random $N \times N$ matrix $H_N = (h_{ij})$ by sampling its entries h_{ij} according to some know probability distributions and enforce a structural constraint (such as symmetry) to ensure its eigenvalues are real. We refer to the sequence of matrices $(H_N)_{N=1}^{\infty}$ as an *ensemble*. Simplifying this construction further, we can require that the entries of the matrix are independent of one another. Remarkably, under few assumptions about the distributions of the entries, when the dimension of random matrix becomes large its *spectral statistics* can be described very precisely. In particular, if the matrix entries are properly normalized, then as the dimension of the matrix tends to infinity its spectrum (basically, a histogram of the eigenvalues) will fill out a compact region of the real line with positive density

The location and density of its eigenvalues can be specified at both a global scale, that is for intervals of order one, and a local scale, that is for intervals of the same order as the inverse of the dimension (up to logarithmic factors). With a *local law*, which describes the density of the eigenvalues of a random matrix on a short scale, it is possible to say how close the spectrum of a random matrix is likely to be to its limiting profile. It is this type of result we present in this paper. There are many more spectral statistics: for example, we can say exact things about the eigenvectors, the spectral spacing, the correlations between eigenvalues, and the distribution of the largest eigenvalue. Interestingly, the asymptotic spectral statistics are often independent of the distribution of the entries and behave in the same way as the Gaussian distribution (which is a canonical point of comparison), a property referred to as *universality*.

Eugene Wigner's seminal work in the 1950s is considered by many to be the beginning of random matrix theory [22]. Wigner introduced random matrices to model the spectra of heavy atoms. He postulated that the gaps between the lines in the spectrum of a heavy atom should resemble the spacings between the eigenvalues of a random matrix; in Section 7 we see very clearly how the spacing between eigenvalues behaves like electrostatic repulsion. The idea of universality can be found in Wigner's original work, in which he postulated that the spacings of the spectrum should not depend on anything but the symmetry class of the underlying evolution. The first ingredient for the modern approach to proving universality for random matrices is a local law, so a natural extension to this research would be to prove universality for this paper's ensemble.

While Wigner's work formulated random matrix theory in more modern terms, it is possible to trace the origins of random matrix theory back to the 1920s. The statistician, John Wishart, worked in multivariate statistics and he introduced random matrices to analyze large samples of data [21]. Suppose we have some centered M dimensional data from which we take N independent, identically distributed samples and form the $M \times N$ *data matrix*

$$X := [\mathbf{x}_1, \ldots, \mathbf{x}_N],$$

where

$$\mathbf{x}_i := (x_{1i}, \ldots x_{Mi})^* \in \mathbb{R}^M.$$

Suppose we wish to understand correlations between the data, so we define the $M \times M$ *population covariance matrix*

$$[\Sigma_{ij}] = \Sigma := \mathbb{E}\mathbf{x}_1\mathbf{x}_1^*,$$

so that

$$\Sigma_{ij} = \mathbb{E}x_{i1}x_{j1},$$

that is the covariance between the ith and jth component of the distribution we are drawing our samples from. How well can we estimate Σ with the *sample covariance matrix*

$$\underline{H} := \frac{1}{N} X X^* ?$$

In the classical case, where M is fixed and N is very large, we can use the law of large numbers to conclude that

$$\underline{H} \to \Sigma,$$

since

$$\underline{H}_{ij} = \frac{1}{N} \sum_{k=1}^{N} x_{ik} x_{jk} \to \mathbb{E} x_{i1} x_{j1} = \Sigma_{ij}.$$

So, we have convergence in each entry. Denote the spectra of the two matrices as follows: let $(\sigma_\alpha)_{\alpha=1}^{M}$ be the ordered spectrum of the population covariance matrix and $(\underline{\lambda}_\alpha)_{\alpha=1}^{M}$ be the ordered spectrum of the sample covariance matrix. Then, a simple corollary is that $\underline{\lambda}_\alpha$ converges to σ_α with at most Gaussian fluctuation on the order $N^{-1/2}$ by the central limit theorem. So, not only can we estimate the population covariance matrix with the sample covariance matrix but we have an idea of how many samples we need for a desired level of accuracy.

Now we move to the high-dimensional case; instead of fixing M we allow it to be a function of N that goes to infinity with N such that

$$\lim_{N \to \infty} \frac{M_N}{N} = \gamma \in (0, \infty).$$

This setting is naturally motivated by high-dimensional data, such as genomic sequences, wireless communication, and finance, where the data are high-dimensional and the sample sizes are limited. Random matrices have many application in these areas; see, for example, [2] for a summary.

Often what we are interested in for applications (such as principal component analysis) are not the entries of the matrix Σ but its eigenvalues or its inverse; or even more specifically, its largest eigenvalues. So, we focus on convergence of the spectrum and not the entries of the matrix. It is thus natural to ask, does the spectrum of the sample covariance matrix approximate the spectrum of the population covariance matrix? The answer is no, not in the high-dimensional case. Even if we deal with the simplest case and assume that each entry x_{ij} has mean zero, variance one, and that they are all independent, which is referred to as the null covariance case, we obtain a negative answer. In this case, it is clear $\Sigma = \mathbb{1}$ but the spectrum of the sample covariance matrix converges to a deterministic distribution, called the Marchenko-Pastur distribution, which is supported on a compact interval containing 1 almost surely. The distribution μ_{mp} is exposed at length in Subsection 2.3, but for now we note that the density of the Marchenko-Pastur distribution has a closed form expression and is supported on the compact interval $[\lambda_-, \lambda_+]$ possibly with a point mass at 0: the density is given by

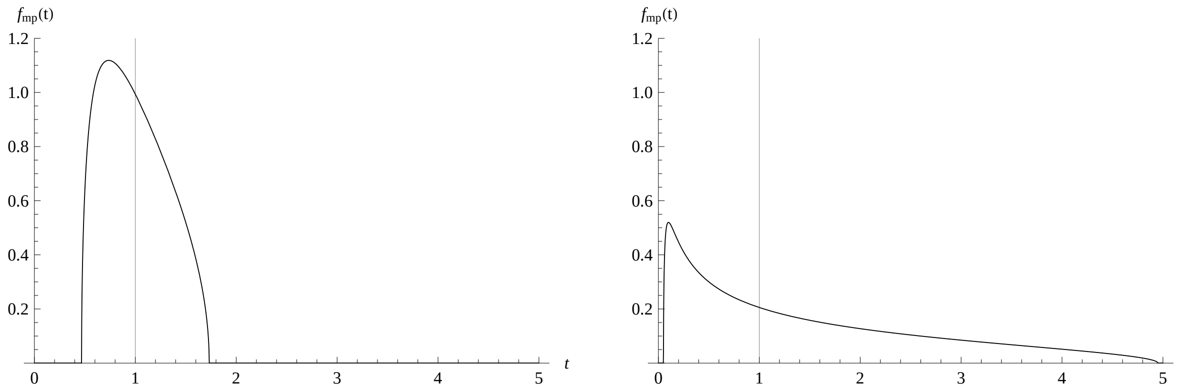
$$f_{\mathrm{mp}}(t) = \frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - t)(t - \lambda_-)}}{\gamma t} \mathbb{I}_{[\lambda_-, \lambda_+]} \mathrm{d}t$$

where

$$\lambda_\pm := (1 \pm \sqrt{\gamma})^2.$$

We have plotted the function in Figure 1 and, as expected, its shape depends on γ.

There has been much work on covariance matrices after Wishart. Marchenko and Pastur proved a global law for generalized covariance matrices in [17], and thus the distribution carries their names. With a quite distinct approach Silverstein and Bai proved a global law for generalized covariance matrices in [19, 20]



(a) The Marchenko-Pastur density for $\gamma = 0.1$.   (b) The Marchenko-Pastur density for $\gamma = 1.5$.

Figure 1: The figure shows the Marchenko-Pastur density for two different values of γ. Notice that as γ becomes larger the length of the density's support increases. The vertical line at 1 is where we would naively expect the eigenvalues of a null sample covariance matrix to concentrate but we can see that the eigenvalues spread out on a compact interval containing 1.

under milder assumption than [17]. Recently, Pillai and Yin have proved universality for null case covariance matrices X^*X in [16], and in doing so they had to prove a local law for the ensemble. However, they assume that the entries of the matrix X have sub-exponential decay. Sub-exponential decay of the matrix entries is a common assumption when proving local laws, but it was weakened in the papers [7, 8], where Erdős, Knowles, Yau, and Yin prove universality (and thus a local law also) for sparse matrices. Sparse matrices, which include as a special case the adjacency matrices of Erdős-Rényi random graphs, have entries whose higher moments decay at a rate proportional to inverse powers of the sparseness parameter q and not $N^{1/2}$. We combine these two research directions in Section 3 of this paper by defining sparse covariance matrices and stating our results for the ensemble.

We can summarize the paper's content briefly as follows: in Section 2 we review some of the basic tools required to prove local laws for random matrices using the resolvent method. In particular, we state and prove the basic properties of the Stieltjes transform and explain why it is particularly useful in random matrix theory. We discuss resolvent formalism by stating the resolvent identities and proving the most commonly used observations about resolvents. Then we discuss the Marchenko-Pastur distribution in more detail and state the properties we need for our main proof. Finally, we summarize the general method for using resolvents.

In Section 3, we state our main result, a weak local law of sparse covariance matrices, which generalizes the weak local law proved in [16] and holds on the optimal scale of $\eta \gtrsim M^{-1}$ uniformly to the edge. We prove this result in Section 6, and the proof follows the strategy outlined in Subsection 2.4. Covariance matrices are unusual, as they have two closely related ensembles: $H := X^*X$ and $\underline{H} := XX^*$, where X is an $M \times N$ sparse matrix with independent entries. As we show in Section 3, the nonzero eigenvalues of H and $\underline{H}$ are identical, despite the difference in their dimension. This implies the following relationship between their Stieltjes transforms, denoted by m_N and $\underline{m}_M$ respectively:

$$m_N(z) = \left(\frac{M}{N} - 1\right)\frac{1}{z} + \frac{M}{N}\underline{m}_M(z), \tag{1.1}$$

which tells us that controlling m_N is equivalent to controlling $\underline{m}_M$. However, in our proof, because of the form of the large deviation estimate, Lemma 5.1, we are forced to control both the entries of $G(z) := (H - z\mathbb{1})^{-1}$ and $\underline{G}(z) := (\underline{H} - z\mathbb{1})^{-1}$ simultaneously. Since no simple formula such as (1.1) is available for the matrix entries of G and $\underline{G}$, controlling both types of entry simultaneously is nontrivial, and to our knowledge this has not been done before.

Our next result is a leading order bound on the moments of the entries of sparse covariance matrices. We calculate this bound in Section 4 using simple combinatorial arguments, and find that the moments of the off-diagonal entries of H decay inversely proportional to q^2 (for $q < M^{1/4}$) and that the moments of the on-diagonal entries of H are order one. The bound on the off-diagonal entries plateaus when $q \geqslant M^{1/4}$ to a decay inversely proportional to $M^{1/2}$, which essentially tells us when $q \geqslant M^{1/4}$ the matrix H is no longer sparse. Our proof in Section 6 relies on a large deviation estimate, which was first proved in [7], so we state and reprove the large deviation estimate in Section 5 along with a simplified proof for one of the large deviation estimates originally given in [9].

In Section 7 we present a model of Dyson's Brownian motion for generalized covariance matrices X^*TX. The section's results are both positive and negative. We present a new formal calculation of the joint distribution of the diffusion process undergone by the eigenvalues of a covariance matrix when the entires of the data matrix X are undergoing standard Brownian motion. Amazingly, the joint distribution of the eigenvalues decouples from their corresponding eigenvectors when $T = c\mathbb{1}$ for some $c \in \mathbb{R}$. This result is originally due to Bru [3]. Our new result in this section is negative: when $T \neq c\mathbb{1}$ for all $c \in \mathbb{R}$, the joint distribution of the eigenvalues does not decouple from their corresponding eigenvectors—the eigenvalues depend on the position of the eigenvectors at time t in a nontrivial way. Finally, in Section 8, we review the derivation of the self-consistent equation for generalized covariance matrices, which was first performed in [19].

2. Preliminaries

The work of this section is to introduce the key tools of the complex-analytic approach to random matrix theory: the Stieltjes transform (in Subsection 2.1), resolvent formalism (in Subsection 2.2), and some basic facts about the global distribution (in our case, the Marchenko-Pastur distribution in 2.3). Throughout the section we explain how these fundamental tools work together and in Subsection 2.4 we outline the general strategy for proving local laws with the complex-analytic approach. This general method is applied to sparse covariance matrices in Section 6, but first a note on notation:

Notation 2.1 (Constants and order notation). Throughout the paper we do not keep track of the values of constants explicitly. We use c and C to denote generic positive constants which may only depend on the constants in assumptions such as (3.1), (3.2), and (3.3); in particular c and C do not depend on the dimension N or the sparseness parameter q defined in Section 3. Typically, we use c to denote a small constant and C to denote a large constant. The fundamental parameter in our model is N (or equivalently M), and the notations $\gg$, $\ll$, $o(\cdot)$, and $\mathcal{O}(\cdot)$ always refer to the limit as $N \to \infty$ (or equivalently $M \to \infty$). We write $f(N) \ll g(N)$ when $f(N) = o(g(N))$ and $f(N) \sim g(N)$ when $C^{-1}f(N) \leqslant g(N) \leqslant Cf(N)$ for all $N \geqslant N_0$ for some N_0.

Notation 2.2 (Matrices and vectors). We use uppercase letters for matrices, such as X and H, and bold lowercase letters for vectors, such as $\mathbf{v}$ and $\mathbf{x}$. In particular $\mathbb{1}$ denotes the identity matrix whose dimension will be clear from the context. Note this is distinct from the indicator function of a set $A \subseteq \mathbb{C}$, which is denoted by $\mathbf{1}_A$ and where $\mathbf{1}_z \equiv \mathbf{1}_{\{z\}}$ for $z \in \mathbb{C}$. We use lowercase greek letters, such as α and β, for indices that enumerate the spectral parameters of a matrix (such as eigenvalues and eigenvectors), whereas indices such as i, j, k, and l enumerate the rows, columns, and entries of a matrix. For a vector $\mathbf{v}$ we denote the kth component by $\mathbf{v}(k)$. So, when we use $\mathbf{x}_j$ to denote the j column of the matrix X, we have the equivalent notations $\mathbf{x}_j(k) \equiv x_{kj}$. Note that for a matrix A we use A^* to denote conjugate transpose. Even when the matrix A has real entries we still use A^* rather than the regular transpose $A^\top$. Also, we use the notation

$$\sum_i^{(\mathbb{T})} := \sum_{\substack{i=1 \\ i\notin\mathbb{T}}}^{N}, \tag{2.1}$$

where $\mathbb{T} \subseteq \{1, \ldots, N\}$. Note that this is distinct from the sum

$$\sum_{i\neq j} := \sum_{1\leqslant i\neq j\leqslant N} \equiv \sum_{i=1}^{N}\sum_{\substack{j=1 \\ i\neq j}}^{N}. \tag{2.2}$$

2.1. The Stieltjes transform. In this subsection we discuss the Stieltjes transform which, is a key tool in random matrix theory and is used extensively throughout this paper. We define the transform, state some of its important properties, and discuss how it is used in random matrix theory. We explain that studying convergence in Stieltjes transforms is equivalent to studying convergence in probability distributions, but due to the nice analytic properties of the Stieltjes transform the former is often preferable.

Definition 2.3 (Stieltjes transform). *Let $\mu \in \mathcal{M}(\mathbb{R})$ be a Borel measure on $\mathbb{R}$ and let $z := E + \mathrm{i}\eta \in \mathbb{C}$. Then we define the Stieltjes transform of μ, denoted by $m_\mu : \mathbb{C}\backslash\operatorname{supp}(\mu) \to \mathbb{C}$, as*

$$m_\mu(z) := \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{t-z}. \tag{2.3}$$

In particular, if $\mu \in \mathcal{P}(\mathbb{R})$ is a probability distribution ($\mu(\mathbb{R}) = 1$) and the random variable X is distributed according to μ, then

$$m_\mu(z) = \mathbb{E}\left(\frac{1}{X - z}\right). \tag{2.4}$$

While $m_\mu(z)$ is defined for all $z \in \mathbb{C}\backslash \operatorname{supp}(\mu)$, if we restrict the function's domain to the upper half-plane $\mathbb{H}$, then we have $m_\mu : \mathbb{H} \to \mathbb{H}$, since

$$\operatorname{Im} m_\mu = \eta \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{(t - E)^2 + \eta^2} \geqslant \frac{1}{\eta} > 0. \tag{2.5}$$

Moreover, m_μ is analytic on $\mathbb{H}$. From now on we only consider the Stieltjes transforms of probability distribution μ and we restrict the transform's domain to $z := E + \mathrm{i}\eta \in \mathbb{H}$, so that $\eta > 0$. There is a very simple bound for the Stieltjes transform, namely

$$|m_\mu(z)| \leqslant \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{|t - z|} = \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{\sqrt{(t - E)^2 + \eta^2}} \leqslant \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{\eta} = \frac{1}{\eta}. \tag{2.6}$$

If the probability distribution μ is compactly supported, then we can swap the limit and integration to find

$$\lim_{z \to \infty} -z m_\mu(z) = \lim_{z \to \infty} \int_{\mathbb{R}} \frac{z\, \mu(\mathrm{d}t)}{z - t} = \int_{\mathbb{R}} \mu(\mathrm{d}t) = 1. \tag{2.7}$$

Even when μ is not compactly supported, we have

$$\lim_{\eta \to \infty} -\mathrm{i}\eta\, m_\mu(\mathrm{i}\eta) = \lim_{\eta \to 0} \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{1 - \frac{t}{\mathrm{i}\eta}} = 1, \tag{2.8}$$

by the dominated convergence theorem. These fundamental properties make the Stieltjes transform a nice, well-behaved, analytic object that is amenable to the tools of complex analysis, which in turn makes it easier for us to study convergence in probability distributions.

We can also relate the Stieltjes transform m_μ to the moments of the measure μ. Suppose additionally that μ is compactly supported, $\operatorname{supp}(\mu) \subseteq [-C, C]$ for some $C \in \mathbb{R}$, and define the kth moment of μ as

$$\mu_k := \int_{\mathbb{R}} t^k\, \mu(\mathrm{d}t) = \mathbb{E}X^k \tag{2.9}$$

for integers $k \geqslant 0$. Then, trivially, $|\mu_k| \leqslant C^k$ and for each k the Stieltjes transform admits the following asymptotic expansion in the neighborhood of infinity given by

$$m_\mu(z) = -\sum_{i=0}^{k} \frac{\mu_i}{z^{i+1}} - o\left(\frac{1}{z^{k+1}}\right). \tag{2.10}$$

Informally, $m_\mu(\infty) = 0$ and we may think of μ_k as Laurent coefficients at ∞, so m_μ is like a moment generating function. This is an important fact for random matrices as the moments of random matrices are related to the combinatorial structure of labelled planar graphs, which enforces some recursive relations on the moments. These recursive relations lead directly to algebraic equations for the Stieltjes transform, which in turn help to estimate the moments. See [1] for more details on this use of the Stieltjes transform.

Under some mild assumption we can recover the probability distribution μ from its Stieltjes transform m_μ. This statement is made precise in the following proposition.

Proposition 2.4 (Inversion of the Stieltjes transform). *Let $\mu \in \mathcal{P}(\mathbb{R})$, then for any $a < b$*

$$\lim_{\eta \to 0} \frac{1}{\pi} \int_a^b \operatorname{Im} m_\mu(z) \mathrm{d}E = \mu(a,b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\}. \tag{2.11}$$

Proof. By definition, we have

$$\frac{1}{\pi} \operatorname{Im} m_\mu(z) = \frac{1}{\pi} \int_{\mathbb{R}} \operatorname{Im} \frac{1}{t-z} \mu(\mathrm{d}t) = \int_{\mathbb{R}} \frac{1}{\pi} \frac{\eta}{(t-E)^2 + \eta^2} \mu(\mathrm{d}t),$$

which is the density of the convolution $\mu * P_\eta$ at E, where

$$P_\eta(\mathrm{d}E) := \frac{1}{\pi} \frac{\eta}{E^2 + \eta^2} \mathrm{d}E$$

is the Cauchy distribution. Then, using Fubini's theorem, we see

$$\begin{aligned}
\lim_{\eta \to 0} \frac{1}{\pi} \int_a^b \operatorname{Im} m_\mu(z) \, \mathrm{d}E &= \frac{1}{\pi} \lim_{\eta \to 0} \int_a^b \int_{\mathbb{R}} \frac{\eta}{(t-E)^2 + \eta^2} \mu(\mathrm{d}t) \mathrm{d}E \\
&= \frac{1}{\pi} \lim_{\eta \to 0} \int_{\mathbb{R}} \eta \int_a^b \frac{\mathrm{d}E}{(t-E)^2 + \eta^2} \mu(\mathrm{d}t) \\
&= \frac{1}{\pi} \lim_{\eta \to 0} \eta \int_{\mathbb{R}} \frac{1}{\eta} \left[-\arctan\left(\frac{t-E}{\eta} \right) \right]_a^b \mu(\mathrm{d}t) \\
&= \frac{1}{\pi} \lim_{\eta \to 0} \int_{\mathbb{R}} \arctan\left(\frac{t-a}{\eta} \right) - \arctan\left(\frac{t-b}{\eta} \right) \mu(\mathrm{d}t).
\end{aligned}$$

Now, we take the limit of the integrand and use the properties of the arctangent to find

$$\lim_{\eta \to 0} \left(\arctan\left(\frac{t-a}{\eta} \right) - \arctan\left(\frac{t-b}{\eta} \right) \right) = \begin{cases} 0 & \text{if } \; t < a \\ \frac{\pi}{2} & \text{if } \; t = a \\ \pi & \text{if } \; a < t < b \\ \frac{\pi}{2} & \text{if } \; t = b \\ 0 & \text{if } \; b < t \end{cases} \tag{2.12}$$

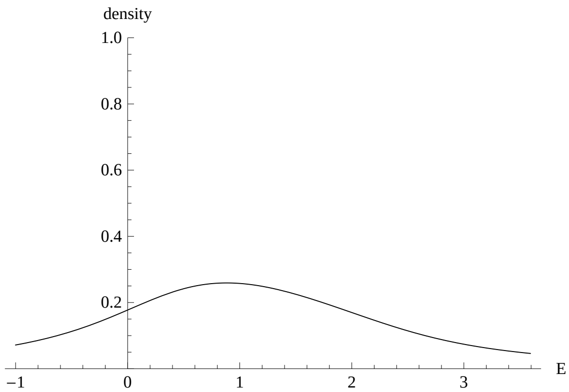
and thus

$$\lim_{\eta \to 0} \frac{1}{\pi} \int_a^b \operatorname{Im} m_\mu(z) \, \mathrm{d}E = \frac{1}{\pi} \int_{\mathbb{R}} \left(\frac{\pi}{2} \mathbf{1}_a(t) + \frac{\pi}{2} \mathbf{1}_b(t) + \pi \mathbf{1}_{(a,b)}(t) \right) \mu(\mathrm{d}t) = \mu(a,b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\},$$

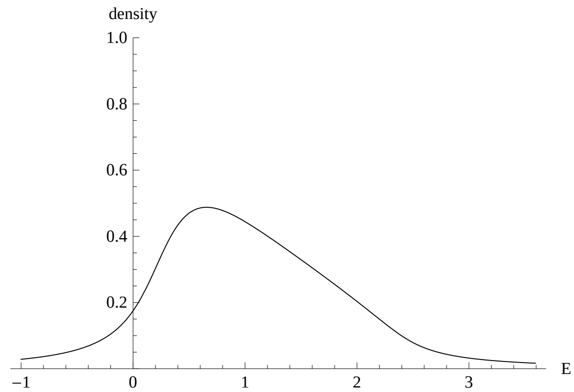
by the dominated convergence theorem. ■

It is easy from here to recover the density of μ, if it admits one. From the interpretation as a convolution in the proof of Proposition 2.4, it is possible to think of η as a resolution which gives a sharper picture of the original probability distribution μ as η gets smaller. Following the convention in quantum mechanics, we use the letter E as the real part of the spectral parameter due to its interpretation as an energy level.

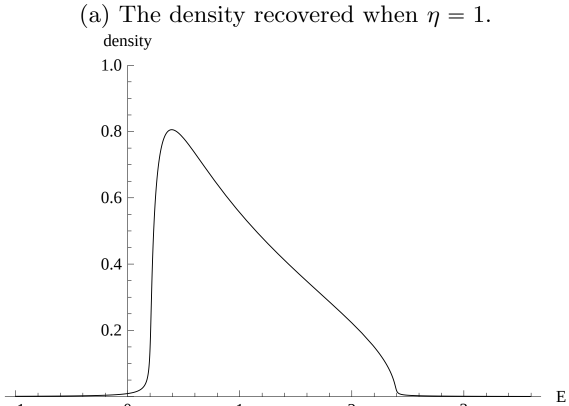
Next, we present two immediate corollaries of Proposition 2.4.



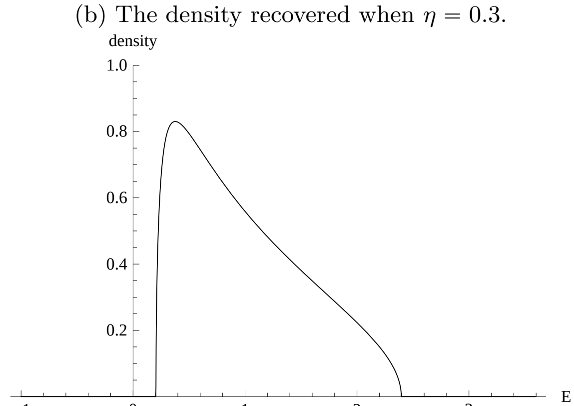
(a) The density recovered when $\eta = 1$.



(b) The density recovered when $\eta = 0.3$.



(c) The density recovered when $\eta = 0.01$.



(d) The density recovered when $\eta = 10^{-10}$.

Figure 2: The figures show numerical examples of the densities recovered using Proposition 2.4 as η approaches zero for the Stieltjes transform of the Marchenko-Pastur distribution when $\gamma = 0.3$. In Figure 2a the density is almost completely smoothed out by the Poisson kernel but in Figure 2d we recover the density very accurately.

Corollary 2.5. *Let $z := E + \mathrm{i}\eta$. If μ admits a density $\frac{\mathrm{d}\mu}{\mathrm{d}E}$, then*

$$\frac{1}{\pi} \lim_{\eta \to 0} \operatorname{Im} m_\mu(z) = \frac{\mathrm{d}\mu(E)}{\mathrm{d}E}. \tag{2.13}$$

So we have point-wise convergence.

Corollary 2.6. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$, then*

$$\mu = \nu \quad \textit{if and only if} \quad m_\mu = m_\nu. \tag{2.14}$$

As we shall see, a sequence Stieltjes transforms can be produced from a random matrices ensemble, and thus the goal of understanding the limiting properties of the ensemble can be achieved through understanding the limiting properties of the sequence of Stieltjes transforms. However, limits of Stieltjes transforms of probability distributions are not always Stieltjes transforms of probability distributions, as the following simple example shows: define

$$m_N(z) = m_{\delta_N}(z) := \int_{\mathbb{R}} \frac{\delta_N(t)}{t-z} = \frac{1}{N-z}.$$

Thus, $\lim_{N\to\infty} m_N = 0$, which clearly cannot be the Stieltjes transform of any probability distribution. We wish to understand the circumstances under which the limit of the Stieltjes transforms of a sequence of probability distribution is always the Stieltjes transform of a probability distribution. The condition is simple and can be understood intuitively as specifying that the sequence of probability distribution should not lose any mass to infinity in the limit. The next proposition formalizes this idea and can be found in [15].

Proposition 2.7. *Let $(\mu_N)_{N=1}^{\infty}$ be a sequence contained in $\mathcal{P}(\mathbb{R})$ and denote their Stieltjes transforms by $(m_N(z))_{N=1}^{\infty}$ respectively. Suppose that pointwise for all $z := E + \mathrm{i}\eta \in \mathbb{H}$,*

$$\lim_{N\to\infty} m_N(z) = m(z) \tag{2.15}$$

for some function $m(z)$. Then there exists $\mu \in \mathcal{P}(\mathbb{R})$, with a Stieltjes transform $m_\mu(z) = m(z)$, if and only if

$$\lim_{\eta\to\infty} -\mathrm{i}\eta\, m(\mathrm{i}\eta) = 1, \tag{2.16}$$

in which case $\mu_N \to \mu$ in distribution.

Proof. The forward direction follows from the calculation in (2.8): if $m(z) = m_\mu(z)$ for some $\mu \in \mathcal{P}(\mathbb{R})$, then (2.7) holds as this is a property of the Stieltjes transform of any probability distribution.

Helly's selection theorem tells us that some subsequence μ_{N_k} converges vaguely to a possibly defective measure μ. Now using the fact that $(z-E)^{-1}$ is continuous and vanishes at infinity, we find

$$m_{N_k}(z) \to m_\mu(z)$$

pointwise for all $z \in \mathbb{H}$.

Thus, $m_\mu = m$, and so the limits of all subsequences of $(\mu_N)_{N=1}^{\infty}$ have the same Stieltjes transforms m. Moreover, by condition (2.16) it follows that μ is not defective and in fact $\mu(\mathbb{R}) = 1$. In more detail,

$$1 = \lim_{\eta\to\infty} -\mathrm{i}\eta\, m(\mathrm{i}\eta) = \lim_{\eta\to\infty} -\mathrm{i}\eta \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{t-\mathrm{i}\eta} = \lim_{\eta\to\infty} \int_{\mathbb{R}} \frac{\mu(\mathrm{d}t)}{1-\frac{t}{\mathrm{i}\eta}} = \int_{\mathbb{R}} \mu(\mathrm{d}t)$$

by the dominated convergence theorem. Then by the uniqueness of Stieltjes transforms implied by Proposition 2.4, all subsequences of $(\mu_N)_{N=1}^{\infty}$ have the same limit, and thus $\mu_N \to \mu$ in distribution. ∎

Before we explain the Stieltjes transform's relationship with matrices we need to make a definition.

Definition 2.8 (Empirical distribution of a matrix). *Let A be an $N\times N$ Hermitian matrix and denote its (real) eigenvalues by*

$$(\lambda_\alpha)_{\alpha=1}^{N}\,. \tag{2.17}$$

Then the empirical distribution of A is the (random) probability distribution on $\mathbb{R}$ defined by

$$\mu_A := \frac{1}{N}\sum_{\alpha=1}^{N} \delta_{\lambda_\alpha}, \tag{2.18}$$

That is a point mass of size $1/N$ at each eigenvalue of A. In particular, the number of eigenvalues in an interval I is given by

$$N \int_I \mu_A(\mathrm{d}t). \tag{2.19}$$

Using Definition 2.8, we find that the Stieltjes transform of the empirical distribution of A is given by

$$m_A(z) := \int_{\mathbb{R}} \frac{\mu_A(\mathrm{d}t)}{t-z} = \frac{1}{N} \sum_{\alpha=1}^{N} \frac{1}{\lambda_\alpha - z}. \tag{2.20}$$

From now on we refer to $m_A(z)$ as the Stieltjes transform of the matrix A and use the slightly abusive notation $m_A(z)$ instead of $m_{\mu_A}(z)$.

2.2. Resolvent formalism. When studying the spectrum of operators on Hilbert spaces (and more general spaces), we can use resolvent formalism to apply concepts from complex analysis. The operators we are interested in are matrices and the resolvent has an intricate relationship with the Stieltjes transform, which we have already demonstrated as a useful tool for studying the convergence of probability distributions on the real line.

Definition 2.9 (Resolvent). *Let A be an $N \times N$ Hermitian matrix. Then denote the resolvent (or Green function) of A by $G : \mathbb{C}\backslash\sigma(A) \to GL_N(\mathbb{C})$, which is defined as*

$$G \equiv G(z) := \frac{1}{A - z\mathbb{1}}. \tag{2.21}$$

Moreover, we denote the (i,j) matrix entry of $G(z)$ by $G_{ij} \equiv G_{ij}(z)$.

Thus, we can also write the Stieltjes transform of A as the trace of the resolvent of A:

$$m_N(z) = \frac{1}{N} \sum_\alpha \frac{1}{\lambda_\alpha - z} = \frac{1}{N} \operatorname{Tr} \frac{1}{A - z\mathbb{1}} = \frac{1}{N} \operatorname{Tr} G(z) = \frac{1}{N} \sum_i G_{ii}(z), \tag{2.22}$$

since $(A - z\mathbb{1})^{-1}$ is a rational function of A and so its eigenvalues are $((\lambda_\alpha - z)^{-1})_{\alpha=1}^N$. This expression is another advantage of the Stieltjes transform for random matrices; it allows us to use the algebraic properties of matrices and all the resolvent identities. To state and prove the resolvent identities we need to define the minor of a matrix.

Definition 2.10 (Minors). *Let $\mathbb{T} \subseteq \{1, \ldots, N\}$ and let $A = (a_{ij})$ be an $N \times N$ matrix. We denote by $A^{(\mathbb{T})}$ the $(N - |\mathbb{T}|) \times (N - |\mathbb{T}|)$ Hermitian matrix obtained by removing all the columns and rows of A indexed by $i \in T$. Note that we keep the original names of the indices for entries of A; for example the $(2,2)$ entry of $A^{(1)}$ is still denoted as a_{22}.*

Given a minor matrix $A^{(\mathbb{T})}$, we denote its eigenvalues and eigenvectors by

$$\left(\lambda_\alpha^{(\mathbb{T})}\right)_{\alpha=1}^{N-|\mathbb{T}|} \quad \text{and} \quad \left(\mathbf{v}_\alpha^{(\mathbb{T})}\right)_{\alpha=1}^{N-|\mathbb{T}|} \tag{2.23}$$

respectively. Similarly, we denote the resolvent and Stieltjes transform of the minor matrix by

$$G^{(\mathbb{T})} \equiv G^{(\mathbb{T})}(z) := \frac{1}{A^{(\mathbb{T})} - z\mathbb{1}} \quad \text{and} \quad m^{(\mathbb{T})} \equiv m^{(\mathbb{T})}(z) := \frac{1}{N} \sum_{\alpha=1}^{N-|\mathbb{T}|} \frac{1}{\lambda_\alpha^{(\mathbb{T})} - z} = \frac{1}{N} \operatorname{Tr} G^{(\mathbb{T})}(z) \tag{2.24}$$

respectively. Note that we use the normalization $1/N$, instead of the more natural $(N - |\mathbb{T}|)^{-1}$ in (2.24); this is a convenience for later applications. Regardless, the cardinality of $\mathbb{T}$ is often less than some fixed constant, and thus $N \sim N - |\mathbb{T}|$. Furthermore, we abbreviate $(\{i\})$ as (i) and $(\{i\} \cup \mathbb{T})$ as $(i\mathbb{T})$ in our notation for minors.

Now we have defined the minor of a matrix we can state the Schur complement formula, which relates the diagonal entries of the inverse of a matrix to the block decomposition of the matrix; it is essential to proving the resolvent identities which follow it.

Lemma 2.11 (Schur complement formula). *Let D be an $N \times N$ Hermitian matrix. Then we can decompose D as*

$$D := \left[\begin{array}{c|c} A & B^* \\ \hline B & C \end{array}\right] \tag{2.25}$$

for A, B, and C an $m \times m$, $(N-m) \times m$, and $(N-m) \times (N-m)$ complex matrices respectively. Suppose that D is invertible, then it follows that

$$\left[\left(D^{(\mathbb{T})}\right)^{-1}\right]_{ij} = \left[\left(\left(A - B^* C^{-1} B\right)^{(\mathbb{T})}\right)^{-1}\right]_{ij} \tag{2.26}$$

for $1 \leqslant i, j \leqslant m$ and $i, j \notin \mathbb{T}$.

Proof. Without loss of generality, we assume $\mathbb{T} = \varnothing$. The proof for nonempty $\mathbb{T}$ is identical. The proof uses a block Gaussian elimination by multiplying the matrix A from the right with the block-lower triangular matrix defined by

$$L := \left[\begin{array}{c|c} \mathbb{1} & 0 \\ \hline -C^{-1}B & \mathbb{1} \end{array}\right],$$

where 0 represents the $m \times (N-m)$ matrix of all zeros and $\mathbb{1}$ is $m \times m$ identity matrix, and C^{-1} exists by assumption. It is easy to see

$$L^{-1} = \left[\begin{array}{c|c} \mathbb{1} & 0 \\ \hline C^{-1}B & \mathbb{1} \end{array}\right].$$

So, performing a block matrix multiplication, we find

$$\begin{aligned} DL &= \left[\begin{array}{c|c} A & B^* \\ \hline B & C \end{array}\right] \left[\begin{array}{c|c} \mathbb{1} & 0 \\ \hline -C^{-1}B & \mathbb{1} \end{array}\right] \\ &= \left[\begin{array}{c|c} A - B^*C^{-1}B & B^* \\ \hline 0 & C \end{array}\right] \\ &= \left[\begin{array}{c|c} \mathbb{1} & B^*C^{-1} \\ \hline 0 & \mathbb{1} \end{array}\right] \left[\begin{array}{c|c} A - B^*C^{-1}B & 0 \\ \hline 0 & C \end{array}\right]. \end{aligned}$$

Thus, we may conclude

$$\left[\begin{array}{c|c} A & B^* \\ \hline B & C \end{array}\right] = \left[\begin{array}{c|c} \mathbb{1} & B^*C^{-1} \\ \hline 0 & \mathbb{1} \end{array}\right] \left[\begin{array}{c|c} A - B^*C^{-1}B & 0 \\ \hline 0 & C \end{array}\right] \left[\begin{array}{c|c} \mathbb{1} & 0 \\ \hline C^{-1}B & \mathbb{1} \end{array}\right],$$

and therefore we have the following expression for the inverse of D

$$D^{-1} = \left[\begin{array}{c|c} \mathbb{1} & 0 \\ \hline -C^{-1}B & \mathbb{1} \end{array}\right] \left[\begin{array}{c|c} \left(A - B^*C^{-1}B\right)^{-1} & 0^* \\ \hline 0 & C^{-1} \end{array}\right] \left[\begin{array}{c|c} \mathbb{1} & -B^*C^{-1} \\ \hline 0 & \mathbb{1} \end{array}\right].$$

Finally, we have

$$\left[D^{-1}\right]_{ij} = \left[\left(A - B^* C^{-1} B\right)^{-1}\right]_{ij}$$

for $a \leqslant i, j \leqslant m$. ■

Remark 2.12. Lemma 2.11 is easily extended to the case where A is not in the top left-hand corner of the matrix D. One defines the resulting decomposition of D in the same way and the proof is identical, but notating the proof is more complicated.

The next lemma states the resolvent identities. The form of the resolvent identities depends on the matrix ensemble in question; for example, the identities are slightly different for Wigner matrices and covariance matrices. While the form of the resolvent identities depends on the particular ensemble, here we state the most general form of the resolvent identities to demonstrate their wide applicability. In Lemma 6.5, we state the resolvent identities for the specific case of covariance matrices, which we need for the proof of the local Marchenko-Pastur law in Section 6.

Lemma 2.13 (General resolvent identities). *For an $N \times N$ Hermitian matrix $A = (a_{ij})$, we let $G_{ij}^{(\mathbb{T})}(z)$ be as defined in Definition 2.10 and suppose $z \in \mathbb{C}\backslash\sigma(A)$. Then for the diagonal elements of $G(z)$, when $i \notin \mathbb{T}$,we have*

$$G_{ii}^{(\mathbb{T})}(z) = \frac{1}{a_{ii} - z - \left(\mathbf{a}_i^{(i\mathbb{T})}\right)^* G^{(i\mathbb{T})}(z)\mathbf{a}_i^{(i\mathbb{T})}}. \tag{2.27}$$

For the off-diagonal elements of $G(z)$, when $i \neq j$ and $i, j \notin \mathbb{T}$, we have

$$G_{ij}^{(\mathbb{T})}(z) = -G_{ii}^{(\mathbb{T})}(z) G_{jj}^{(i\mathbb{T})}(z) \left(a_{ij} - \left(\mathbf{a}_i^{(ij\mathbb{T})}\right)^* G^{(ij\mathbb{T})}(z)\mathbf{a}_j^{(ij\mathbb{T})}\right). \tag{2.28}$$

Finally, when $i \neq k$, $j \neq k$ and $i, j, k \notin \mathbb{T}$,

$$G_{ij}^{(\mathbb{T})}(z) = G_{ij}^{(k\mathbb{T})}(z) + \frac{G_{ik}^{(\mathbb{T})}(z) G_{kj}^{(\mathbb{T})}(z)}{G_{kk}^{(\mathbb{T})}(z)}. \tag{2.29}$$

Proof. We only need elementary linear algebra for this proof. Throughout the prove we assume without loss of generality that $\mathbb{T} = \varnothing$, as the proof is identical for nonempty $\mathbb{T}$. First we prove (2.27). For ease of notation we perform the calculation for $i = 1$ but the following is true for general i. Then

$$A := \left[\begin{array}{c|c} a_{11} & \left(\mathbf{a}_1^{(1)}\right)^* \\ \hline \mathbf{a}_1^{(1)} & A^{(1)} \end{array}\right] \tag{2.30}$$

and so

$$A - z\mathbb{1} := \left[\begin{array}{c|c} a_{11} - z & \left(\mathbf{a}_1^{(1)}\right)^* \\ \hline \mathbf{a}_1^{(1)} & A^{(1)} - z\mathbb{1} \end{array}\right].$$

Therefore, using the Schur complement formula stated in Lemma 2.11, we immediately see

$$G_{11} = \frac{1}{a_{11} - z - \left(\mathbf{a}_1^{(1)}\right)^* (A^{(1)} - z\mathbb{1})^{-1} \mathbf{a}_1^{(1)}} = \frac{1}{a_{11} - z - \left(\mathbf{a}_1^{(1)}\right)^* G^{(1)}(z)\mathbf{a}_1^{(1)}}.$$

Now we show (2.28). For easy of notation let $i = 1$ and $j = 2$. Consider the following block decomposition

$$A - z\mathbb{1} = \left[\begin{array}{c|c|c} a_{11} - z & a_{12} & \left(\mathbf{a}_1^{(12)}\right)^* \\ \hline a_{21} & a_{22} - z & \left(\mathbf{a}_2^{(12)}\right)^* \\ \hline \mathbf{a}_1^{(12)} & \mathbf{a}_2^{(12)} & A^{(12)} - z\mathbb{1} \end{array}\right].$$

Using the Schur complement formula, Lemma 2.11, with $m = 2$, we get

$$G_{ks} = \left[\left[\begin{array}{c|c} a_{11} - z - \left(\mathbf{a}_1^{(12)}\right)^* G^{(12)}\mathbf{a}_1^{(12)} & a_{12} - \left(\mathbf{a}_1^{(12)}\right)^* G^{(12)}\mathbf{a}_2^{(12)} \\ \hline a_{21} - \left(\mathbf{a}_2^{(12)}\right)^* G^{(12)}\mathbf{a}_1^{(12)} & a_{22} - z - \left(\mathbf{a}_2^{(12)}\right)^* G^{(12)}\mathbf{a}_2^{(12)} \end{array}\right]^{-1}\right]_{ks}. \tag{2.31}$$

Define

$$K_{ks} := a_{ks} - \mathbf{1}_{k=s} z - \left(\mathbf{a}_k^{(12)}\right)^* G^{(12)}\mathbf{a}_s^{(12)}$$

and

$$K := K_{11}K_{22} - K_{12}K_{21}.$$

Thus, the expression (2.31), gives us

$$G_{11} = \frac{K_{22}}{K}, \quad G_{22} = \frac{K_{11}}{K}, \quad G_{12} = -\frac{K_{12}}{K}, \quad \text{and} \quad G_{21} = -\frac{K_{21}}{K}.$$

So, we see

$$G_{12} = -\frac{K_{12}}{K} = -\frac{K_{12}}{K_{22}}\frac{K_{22}}{K} = -G_{11}\frac{K_{12}}{K_{22}},$$

and by using the first resolvent identity (2.27) with $i = 2$ on $G^{(1)}$, we find

$$= -G_{11}\frac{K_{12}}{K_{22}} = -G_{11}G_{22}^{(1)}K_{12},$$

that is,

$$G_{12} = -G_{11}G_{22}^{(1)}\left(a_{12} - \left(\mathbf{a}_1^{(12)}\right)^* G^{(12)}\mathbf{a}_2^{(12)}\right).$$

Now, we prove (2.29). Again for easy of notion, we prove it for $i = j = 1$ and $k = 2$. Note

$$\frac{K_{11}K_{22}}{K^2} - \frac{K_{11}}{K}\frac{1}{K_{11}} = \frac{K_{12}K_{21}}{K^2}, \tag{2.32}$$

and thus, by (2.32), we find

$$G_{11}G_{22} - G_{22}\frac{1}{K_{11}} = G_{12}G_{21}.$$

Since, using the first resolvent identity (2.27) with $i = 1$ on $G^{(2)}$, $K_{11}^{-1} = G_{11}^{(2)}$, we have

$$G_{11} - G_{11}^{(2)} = \frac{G_{12}G_{21}}{G_{22}},$$

which completes the proof. ■

Along with the identities in Lemma 2.13, the following is a fundamental property of the resolvent and we make use of it frequently during the proof in Section 6. The Lemma is true for any resolvent, which trivially includes the resolvents of minors.

LEMMA 2.14 (WARD IDENTITY). *Let $z := E + \eta$ and let A be an $N \times N$ Hermitian matrix. Denote the resolvent of A by G. Then*

$$\sum_{j=1}^{N} |G_{ij}(z)|^2 = \frac{1}{\eta} \operatorname{Im} G_{ii}(z) \tag{2.33}$$

and, as a simple corollary,

$$\sum_{i,j=1}^{N} |G_{ij}(z)|^2 = \frac{1}{\eta} \operatorname{Im} \operatorname{Tr} G(z). \tag{2.34}$$

PROOF. Denote the eigenvalues and eigenvectors of A by

$$(\lambda_\alpha)_{i=1}^N \quad \text{and} \quad (\mathbf{v}_\alpha)_{i=1}^N$$

respectively. We use an eigendecompoisition on G: note that

$$\left(\frac{1}{\lambda_\alpha - z}\right)_{\alpha=1}^{N} \tag{2.35}$$

are the eigenvalues of G with corresponding eigenvectors $(\mathbf{v}_\alpha)_{\alpha=1}^N$, thus the entires of G are given by

$$G_{ij} = \sum_\alpha \frac{\overline{\mathbf{v}_\alpha(i)}\mathbf{v}_\alpha(j)}{\lambda_\alpha - z}, \quad \text{and} \quad \overline{G_{ij}} = \sum_\alpha \frac{\mathbf{v}_\alpha(i)\overline{\mathbf{v}_\alpha(j)}}{\lambda_\alpha - \bar{z}}.$$

So, we see

$$\begin{aligned}
\sum_j |G_{ij}|^2 &= \sum_j \overline{G_{ij}} G_{ij} \\
&= \sum_j \sum_\alpha \frac{\overline{\mathbf{v}_\alpha(i)}\mathbf{v}_\alpha(j)}{\lambda_\alpha - z} \sum_\beta \frac{\mathbf{v}_\beta(i)\overline{\mathbf{v}_\beta(j)}}{\lambda_\beta - \bar{z}} \\
&= \sum_{\alpha,\beta} \frac{\overline{\mathbf{v}_\alpha(i)}\mathbf{v}_\beta(i)}{|\lambda_\alpha - z|^2} \sum_j \overline{\mathbf{v}_\alpha(j)}\mathbf{v}_\beta(j) \\
&= \sum_\alpha \frac{|\mathbf{v}_\alpha(i)|^2}{|\lambda_\alpha - z|^2} \\
&= \frac{1}{\eta} \operatorname{Im} G_{ii},
\end{aligned} \tag{2.36}$$

since

$$\operatorname{Im} \frac{1}{\lambda_\alpha - z} = \frac{\eta}{|\lambda_\alpha - z|^2}.$$

Then summing equation (2.36) over i yields

$$\sum_{i,j} |G_{ij}(z)|^2 = \frac{1}{\eta} \operatorname{Im} \sum_i G_{ii} = \frac{1}{\eta} \operatorname{Im} \operatorname{Tr} G,$$

which completes the proof. ■

The next lemma is essential and it provides the impetus to study the minors of matrices: the eigenvalues of a minor matrix $A^{(i)}$ are interlaced with (and thus controlled by) the eigenvalues of the matrix A. With this observation we can recursively relate the Stieltjes transforms, resolvents, eigenvalues, etc. of a matrix with the same of its minors. The resolvent identities allow us to make equations which express the resolvent of a matrix in terms of the resolvents of its minors, and putting this together with interlacing allows us to control the error terms in the equations.

PROPOSITION 2.15 (EIGENVALUE INTERLACING). *Block decompose an $N \times N$ Hermitian random matrix $A = (a_{ij})$ into its jth minor and the remaining elements. Assume the distributions of the entires of A are smooth. Denote the eigenvalues of A and $A^{(j)}$ by*

$$(\lambda_\alpha)_{\alpha=1}^N \quad \text{and} \quad \left(\lambda_\alpha^{(j)}\right)_{\alpha=1}^{N-1}. \tag{2.37}$$

respectively, ordered from smallest to largest. Then the eigenvalues of A and $A^{(j)}$ are strictly interlaced almost surely:

$$\lambda_1 < \lambda_1^{(j)} < \lambda_2 < \lambda_2^{(j)} < \cdots < \lambda_{N-1} < \lambda_{N-1}^{(j)} < \lambda_N. \tag{2.38}$$

PROOF. For easy of notation assume $j = 1$, as the proof is identical for general j. Suppose that λ is one of the eigenvalues of A and let $\mathbf{v} := (v_1, \ldots, v_N)^*$ be the corresponding normalized eigenvector, so that

$$A\mathbf{v} = \lambda\mathbf{v}. \tag{2.39}$$

From the continuity of the distribution it also follows that $v_1 \neq 0$ almost surely. From the decomposition (2.30) and the equation (2.39) we get the pair of equations

$$av_1 + \mathbf{a}^*\mathbf{v}^{(1)} = \lambda v_1 \quad \text{and} \quad \mathbf{a}v_1 + A^{(1)}\mathbf{v}^{(1)} = \lambda\mathbf{v}^{(1)},$$

where $\mathbf{v}^{(1)} := (v_2, \ldots, v_N)^*$. Rearranging the above equation we find

$$\mathbf{a}^*\mathbf{v}^{(1)} = (\lambda - a)\, v_1 \quad \text{and} \quad \mathbf{v}^{(1)} = \left(\lambda\mathbb{1} - A^{(1)}\right)^{-1} \mathbf{a}v_1$$

and then substituting the second equation into the first we see

$$(\lambda - a)\, v_1 = \mathbf{a}^* \left(\lambda\mathbb{1} - A^{(1)}\right)^{-1} \mathbf{a}v_1. \tag{2.40}$$

Using a spectral representation of $A^{(1)}$, where $\mathbf{v}_\alpha^{(1)}$ are the corresponding normalized eigenvectors of the eigenvalues $\lambda_\alpha^{(1)}$, we find

$$(\lambda - a)\, v_1 = \frac{v_1}{N} \sum_{\alpha=1}^{N-1} \frac{\left|\sqrt{N}\mathbf{a} \cdot \mathbf{v}_\alpha^{(1)}\right|^2}{\lambda - \lambda_\alpha^{(1)}}.$$

Since $v_1 \neq 0$ almost surely, we find

$$\lambda - a = \frac{1}{N} \sum_{\alpha=1}^{N-1} \frac{\left|\sqrt{N}\mathbf{a} \cdot \mathbf{v}_\alpha^{(1)}\right|^2}{\lambda - \lambda_\alpha^{(1)}}. \tag{2.41}$$

Moreover, $\mathbf{v}_\alpha^{(1)}$ and $\mathbf{a}$ are independent and $\left|\sqrt{N}\mathbf{a}\cdot\mathbf{v}_\alpha^{(1)}\right|^2$ is strictly positive almost surely. So, from equation (2.41) we may conclude that $\lambda \neq \lambda_\alpha^{(1)}$ for all α almost surely. Also note that the function

$$\Phi(\lambda) := \frac{1}{N}\sum_{\alpha=1}^{N-1}\frac{\left|\sqrt{N}\mathbf{a}\cdot\mathbf{v}_\alpha^{(1)}\right|^2}{\lambda-\lambda_\alpha^{(1)}} \tag{2.42}$$

is strictly decreasing from ∞ to $-\infty$ in the interval $\left(\lambda_{\alpha-1}^{(1)},\lambda_\alpha^{(1)}\right)$. Therefore, there is exactly one solution to the equation

$$\lambda - a = \Phi(\lambda)$$

in the interval $\left(\lambda_{\alpha-1}^{(1)},\lambda_\alpha^{(1)}\right)$, since $\lambda - a$ is increasing in the same interval. By a similar argument we can show there is exactly one solution in each of the intervals $(-\infty,\lambda_1^{(1)})$ and $(\lambda_{N-1}^{(1)},\infty)$. ■

The following is a simple consequence of Proposition 2.15 and integration by parts; it states how interlacing affects the Stieltjes transforms of a matrix and its minors.

Corollary 2.16. *Let $z := E+\mathrm{i}\eta \in \mathbb{H}$. Denote the Stieltjes transform of A and $A^{(j)}$ by $m_N(z)$ and $m_{N-1}^{(j)}(z)$ respectively. Then, because the eigenvalues of A and $A^{(j)}$ are interlaced, we have the following bound on the difference of their Stieltjes transform*

$$\left|m_N(z) - m_{N-1}^{(j)}(z)\right| \leqslant \frac{C}{N\eta}. \tag{2.43}$$

Proof. Let $\mu \equiv \mu_N$ and $\mu^{(j)} \equiv \mu_{N-1}^{(j)}$ denote the empirical distributions of A and $A^{(j)}$ respectively, and define $F(t) := \mu\left((-\infty,t]\right)$ and $F^{(j)}(t) := \mu^{(j)}\left((-\infty,t]\right)$. From Lemma 2.15 we know that the eigenvalues of A and $A^{(j)}$ are interlaced, thus we have the following uniform bound

$$\sup_t \left|NF(t) - NF^{(j)}(t)\right| \leqslant 1. \tag{2.44}$$

Integrating each integral by parts in the second step, we find

$$\begin{aligned}
\left|Nm(z) - Nm^{(j)}(z)\right| &= \left|\int_{\mathbb{R}}\frac{N\mathrm{d}F(t)}{t-z} - \int_{\mathbb{R}}\frac{N\mathrm{d}F^{(j)}(t)}{t-z}\right| \\
&= \left|\left[\frac{NF(t)-NF^{(j)}(t)}{t-z} - \right]_{-\infty}^{\infty} + \int_{\mathbb{R}}\frac{NF(t)-NF^{(j)}(t)}{(t-z)^2}\mathrm{d}t\right| \\
&= \left|\int_{\mathbb{R}}\frac{NF(t)-NF^{(j)}(t)}{(t-z)^2}\mathrm{d}t\right| \\
&\leqslant \left|\int_{\mathbb{R}}\frac{\mathrm{d}t}{(t-z)^2}\right| \\
&\leqslant \int_{\mathbb{R}}\frac{\mathrm{d}t}{|t-z|^2} \\
&\leqslant \frac{C}{\eta},
\end{aligned}$$

where the bound (2.44) is used for the fourth line. ■

2.3. The Marchenko-Pastur distribution. In this section we define the Marchenko-Pastur distribution, which was introduced by Marchenko and Pastur in [17] to describe the global limiting spectrum of generalized covariance matrices, and review some of its important properties. Throughout the entire paper (and this subsection in particular), $\sqrt{\cdot}$ denotes the complex square root with a branch cut on the negative real axis.

DEFINITION 2.17 (MARCHENKO-PASTUR DISTRIBUTION). *The distribution is defined with the density function*

$$\frac{\mathrm{d}\nu_{mp}(t)}{\mathrm{d}t} := \frac{\sqrt{(\lambda_+ - t)(t-\lambda_-)}}{2\pi\gamma t}\mathbb{1}_{[\lambda_-,\lambda_+]} \tag{2.45}$$

where

$$\lambda_\pm := (1\pm\sqrt{\gamma})^2\,, \tag{2.46}$$

so that

$$\mu_{mp}(A) := \begin{cases} \nu_{mp}(A) & \text{if } 0<\gamma\leqslant 1 \\ (1-\gamma)\,\delta_0 + \nu_{mp}(A) & \text{if } \gamma > 1 \end{cases} \tag{2.47}$$

for any measurable set $A\subseteq\mathbb{R}$.

From this definition, we calculate the Stieltjes transform of the Marchenko-Pastur distribution and derive a self-consistent equation for the transform in Lemma 2.18.

LEMMA 2.18. *Define $m_{mp}\equiv m_{mp}(z)$ as the Stieltjes transform of the Marchenko-Pastur distribution μ_{mp}. Then*

$$m_{mp}(z) = \frac{1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2\gamma z}. \tag{2.48}$$

Moreover, $m_{mp}(z)$ satisfies the self-consistent equation

$$m_{mp}(z) = \frac{1}{1-\gamma-z\gamma m_{mp}(z)-z}. \tag{2.49}$$

PROOF. The following proof is based on Lemma 3.11 of [2]. For now assume $\gamma<1$; we will deal with the remaining cases at the end of the proof. By definition, we have

$$m_{\mathrm{mp}}(z) = \int_{\lambda_-}^{\lambda_+}\frac{1}{t-z}\frac{\sqrt{(\lambda_+-t)(t-\lambda_-)}}{2\pi t\gamma}\mathrm{d}t.$$

Using the substitution $t = 1+\gamma+2\sqrt{\gamma}\cos s$ and then the substitution $\zeta=\exp(\mathrm{i}s)$, we see

$$\begin{aligned} m_{\mathrm{mp}}(z) &= \int_0^\pi \frac{2}{\pi}\frac{\sin^2 s}{(1+\gamma+2\sqrt{\gamma}\cos s)(1+\gamma+2\sqrt{\gamma}\cos s - z)}\mathrm{d}s \\ &= \frac{1}{\pi}\int_0^{2\pi}\frac{\left(\frac{e^{\mathrm{i}s}-e^{-\mathrm{i}s}}{2\mathrm{i}}\right)^2}{(1+\gamma+\sqrt{\gamma}(e^{\mathrm{i}s}+e^{-\mathrm{i}s}))(1+\gamma+\sqrt{\gamma}(e^{\mathrm{i}s}+e^{-\mathrm{i}s})-z)}\mathrm{d}s \\ &= \frac{\mathrm{i}}{4\pi}\oint_{|\zeta|=1}\frac{(\zeta-\zeta^{-1})^2}{\zeta(1+\gamma+\sqrt{\gamma}(\zeta+\zeta^{-1}))(1+\gamma+\sqrt{\gamma}(\zeta+\zeta^{-1})-z)}\mathrm{d}\zeta \\ &= \frac{\mathrm{i}}{4\pi}\oint_{|\zeta|=1}\frac{(\zeta^2-1)^2}{\zeta((1+\gamma)\zeta+\sqrt{\gamma}(\zeta^2+1))((1+\gamma)\zeta+\sqrt{\gamma}(\zeta^2+1)-z\zeta)}\mathrm{d}\zeta, \end{aligned}$$

where we used the fact that $\mathrm{d}t = -2\sqrt{\gamma}\sin s\,\mathrm{d}s$ and $\mathrm{d}s = -\frac{\mathrm{i}}{\zeta}\mathrm{d}\zeta$. Notice that the integrand has five simple poles at the points

$$\zeta_0 := 0, \quad \zeta_1 := \frac{-(1+\gamma)+(1-\gamma)}{2\sqrt{\gamma}} = -\sqrt{\gamma}, \quad \zeta_2 := \frac{-(1+\gamma)-(1-\gamma)}{2\sqrt{\gamma}} = -\frac{1}{\sqrt{\gamma}},$$

$$\zeta_3 := \frac{-(1+\gamma)+z+\sqrt{(1+\gamma-z)^2-4\gamma}}{2\sqrt{\gamma}}, \quad \text{and} \quad \zeta_4 := \frac{-(1+\gamma)+z-\sqrt{(1+\gamma-z)^2-4\gamma}}{2\sqrt{\gamma}}.$$

It is easy to calculate the residues at each of the five poles and find them equal to

$$\frac{1}{\gamma}, \quad -\frac{1-\gamma}{\gamma z}, \quad \frac{1-\gamma}{\gamma z}, \quad \frac{\sqrt{(1+\gamma-z)^2-4\gamma}}{\gamma z}, \quad \text{and} \quad -\frac{\sqrt{(1+\gamma-z)^2-4\gamma}}{\gamma z} \tag{2.50}$$

respectively. Now, we have to decide whether the singularities are inside the unit circle and thus which residues to count. Noting that $\zeta_3\zeta_4 = 1$, we see that only one of the residues corresponding to ζ_3 and ζ_4 should be counted. By the definition of the complex square root, we know that the real and imaginary parts of $\sqrt{(1+\gamma-z)^2-4\gamma}$ and $z-(1+\gamma)$, hence

$$|\zeta_3| > 1 \quad \text{and} \quad |\zeta_4| < 1.$$

Moreover,

$$|\zeta_1| = |-\sqrt{\gamma}| < 1 \quad \text{and} \quad |\zeta_2| = |-1/\sqrt{\gamma}| > 1.$$

Thus, by Cauchy's residue theorem, we obtain

$$\begin{aligned} m_{\mathrm{mp}}(z) &= -\frac{1}{2}\left(\frac{1}{\gamma} + \frac{\gamma-1}{\gamma z} - \frac{\sqrt{(1+\gamma-z)^2-4\gamma}}{\gamma z}\right) \\ &= \frac{1-\gamma-z+\sqrt{(z+\gamma-1)^2-4\gamma z}}{2\gamma z} \\ &= \frac{1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2\gamma z}, \end{aligned} \tag{2.51}$$

which completes the proof when $\gamma < 1$.

When $\gamma > 1$ the Marchenko-Pastur distribution also has a point mass of size $1-\gamma$ at zero, so the integral has an extra term we must take into account. Moreover, when $\gamma > 1$, we have $|\zeta_1| = \left|-\sqrt{\gamma}\right| > 1$ and $|\zeta_2| = \left|-1/\sqrt{\gamma}\right| < 1$, and thus the residue at ζ_2 should be taken into consideration. After some manipulation, we find equation (2.51) is still true. The case when $\gamma = 1$ is true because of the equation's continuity in γ.

We can prove equation (2.49) by calculating and using equation (2.51):

$$
\begin{aligned}
-\frac{1}{m_{\mathrm{mp}}} &= -\frac{2\gamma z}{1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}} \\
&= -\frac{2\gamma z\left(1-\gamma-z-\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)}{\left(1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)\left(1-\gamma-z-\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)} \\
&= -\frac{2\gamma z\left(1-\gamma-z-\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)}{(1-\gamma-z)^2+(\lambda_+-z)(z-\lambda_-)} \\
&= -\frac{2\gamma z\left(1-\gamma-z-\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)}{(1-\gamma-z)^2-(1-\gamma-z)^2+4\gamma z} \\
&= \frac{\gamma-1+z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2} \\
&= \frac{2\gamma-2+2z-\gamma+1-z+\mathrm{i}z\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2} \\
&= \gamma-1+z+z\gamma\frac{1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2\gamma z} \\
&= \gamma-1+z+z\gamma m_{\mathrm{mp}},
\end{aligned}
$$

which is equivalent to equation (2.49). ■

As we have just seen $m_{\mathrm{mp}}(z)$ is a solution for equation (2.49); the next lemma states that it is the unique solution which is a function from the upper half-plane to the upper half-plane—in particular, it is the only function that is the Stieltjes transform of some distribution, as Stieltjes transforms are functions from the upper half-plane to the upper half-plane. This is a converse to Lemma 2.18

LEMMA 2.19. *Suppose that $s : \mathbb{H} \to \mathbb{H}$ is a complex function from the upper half-plane to the upper half-plane, which satisfies the equation*

$$s(z) + \frac{1}{z+\gamma-1+z\gamma s(z)} = 0. \tag{2.52}$$

Then,

$$s \equiv m_{mp}. \tag{2.53}$$

PROOF. The proof is very simple, as (2.52) is a quadratic equation in s:

$$z\gamma s^2 + (z+\gamma-1)s + 1 = 0.$$

Now using the quadratic formula, we get

$$s = \frac{1-\gamma-z\pm\sqrt{(z+\gamma-1)^2-4z\gamma}}{2z\gamma} = \frac{1-\gamma-z\pm\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2\gamma z}$$

and since $s : \mathbb{H} \to \mathbb{H}$, we must choose the positive square root to get

$$s = \frac{1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2\gamma z}. \tag{2.54}$$

This completes the proof: $s \equiv m_{\mathrm{mp}}$. ■

REMARK 2.20. There is an alternate proof of Lemma 2.18 which uses the observation made in equation (2.10). By looking at the moment generating function for the Marchenko-Pastur distribution, one can derive equation (2.49) using some simple combinatorics of graphs. From there it is possible to solve equation (2.49) using the quadratic formula and the fact that $m(z)$ is a function from $\mathbb{H}$ to $\mathbb{H}$ to select the correct root, thus avoiding any complex analysis. As a reference, see [1] page 20, for example.

In fact, it is possible to see a priori that the solution (2.54) of equation (2.52) is the Stieltjes transform of the Marchenko-Pastur distribution using the inversion formula in Proposition 2.4. In more detail: note that

$$\begin{aligned}
\operatorname{Im}(1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}) &= -\eta+\operatorname{Im}\left(\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right), \\
\operatorname{Re}(1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}) &= 1-\gamma-E+\operatorname{Re}\left(\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right), \\
\operatorname{Im}\left(\frac{1}{2\gamma z}\right) &= -\frac{\eta}{2\gamma\,|z|^2}, \\
\text{and}\quad \operatorname{Re}\left(\frac{1}{2\gamma z}\right) &= \frac{E}{2\gamma\,|z|^2}.
\end{aligned}$$

Thus

$$\begin{aligned}
\operatorname{Im}\left(\frac{1-\gamma-z+\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}}{2\gamma z}\right) &= \frac{E\operatorname{Re}\left(\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)-\eta(1-\gamma)-\eta\operatorname{Im}\left(\mathrm{i}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)}{2\gamma|z|^2} \\
&= \frac{\operatorname{Re}\left(\bar{z}\sqrt{(\lambda_+-z)(z-\lambda_-)}\right)-\eta(1-\gamma)}{2\gamma|z|^2},
\end{aligned}$$

which when we take the limit $\eta\searrow 0$ and divide by π yields the density

$$\frac{1}{\pi}\frac{\operatorname{Re}\left(E\sqrt{(\lambda_+-E)(E-\lambda_-)}\right)}{2\gamma E^2} = \frac{1}{2\pi}\frac{\sqrt{(\lambda_+-E)(E-\lambda_-)}}{\gamma E}.$$

The Stieltjes transform of the Marchenko-Pastur distribution has some nice analytic properties on a reasonable subset of the upper half-plane. These properties are essential for controlling various quantities during the proof in Section 6. Define the region

$$\mathfrak{R} := \{z := E+\mathrm{i}\eta\in\mathbb{C} : \mathbb{1}_{\gamma>1}(\lambda_-/5)\leqslant E\leqslant 5\lambda_+ \text{ and } M^{-1}\leqslant\eta\leqslant 10(1+\gamma)\}\subseteq\mathbb{H}, \tag{2.55}$$

which we shall motivate further in Section 3, but for now note that the energy parameter E is restricted to being close to the Marchenko-Pastur distribution and η is at most order 1. For $z := E+\mathrm{i}\eta$, we also introduce the edge parameter

$$\kappa\equiv\kappa(z) := \min\{|\lambda_+-E|\,,|E-\lambda_-|\}, \tag{2.56}$$

which records how close the energy of the spectral parameter z is to the edges of the support of the Marchenko-Pastur distribution. Note that when η is small κ has similar behavior to the magnitude of

$$(\lambda_+-z)(z-\lambda_-),$$

which is an important term in the density of the Marchenko-Pastur distribution and its Stieltjes transform. We now collect some of the properties of the Marchenko-Pastur distribution in the region $\mathfrak{R}$ in the following lemma.

Lemma 2.21. *For $z := E + \mathrm{i}\eta \in \mathfrak{R}$, we have the following uniform bounds:*

$$|m_{mp}(z)| \sim 1, \tag{2.57}$$

which tells us $m_{mp}(z)$ is order 1 throughout $\mathfrak{R}$;

$$\left|1 - m_{mp}(z)^2\right| \sim \sqrt{\kappa + \eta}; \tag{2.58}$$

$$|1 - \gamma - z - z\gamma m_{mp}| \geqslant c > 0, \tag{2.59}$$

which tells us that the denominator in (2.49) is not too small throughout $\mathfrak{R}$;

$$\operatorname{Im} m_{mp}(z) \sim \begin{cases} \frac{\eta}{\sqrt{\kappa+\eta}} & \text{if } \kappa \geqslant \eta \text{ and } |E| \notin [\lambda_-, \lambda_+] \\ \sqrt{\kappa+\eta} & \text{if } \kappa \leqslant \eta \text{ or } |E| \in [\lambda_-, \lambda_+], \end{cases} \tag{2.60}$$

which tells us how the density of the Marchenko-Pastur distribution behaves near its edges—in particular, when η is small, we see square root decay as we approach the edges of the support. Furthermore,

$$\frac{\operatorname{Im} m_{mp}(z)}{N\eta} \geqslant \mathcal{O}\left(\frac{1}{N}\right) \tag{2.61}$$

and

$$\frac{\partial}{\partial \eta} \frac{\operatorname{Im} m_{mp}(z)}{\eta} \leqslant 0. \tag{2.62}$$

Proof. The proofs are elementary calculations using the expression (2.48) and the self-consistent equation (2.49). ■

2.4. The general strategy. We can now explain how one puts together the tools in Section 2 to prove a local law for an ensemble of random matrices $H \equiv H_N$ and thus produce an overview of the proof in Section 6. To begin with it is desirable that the ensemble's spectrum is supported on an interval of order 1. This can often be achieved by with the following heuristic: use a moment calculation to compute $\operatorname{Tr} H^2$ and normalize the entries of the matrix to make $\operatorname{Tr} H^2$ order N. This heuristic comes from the fact that $\operatorname{Tr} H^2 = \sum_\alpha \lambda_\alpha^2$, where $(\lambda_\alpha)_{\alpha=1}^N$ are the eigenvalues of H. If the eigenvalues λ_α are order 1, then so are the positive λ_α^2, which implies $\sum_\alpha \lambda_\alpha^2$ is order N. We perform a calculation of this form for sparse covariance matrices in Section 4.

Note that by the remarks in Subsection 2.1, if we want to say that for intervals of order $1/N$ (up to poly-logarithmic factors) the density of the eigenvalues of the random ensemble is close to some deterministic density with probability going to 1 and with error term going to 0 as the dimension N goes to infinity, then we can study the same convergence in Stieltjes transforms. Moreover, we wish to do this uniformly for z in a particular region $\mathfrak{R}$ where E is close to the limiting spectrum of the ensemble and η is at most order 1 and at least $1/N$ up to poly-logarithmic factors. We defined such a region for the ensemble we shall consider in (2.55).

A self-consistent equation is an equation which uniquely defines a function from $\mathbb{H}$ to $\mathbb{H}$, in particular it defines the Stieltjes transform of a probability distribution. The equation usually has nice properties such as stability—that is, it can be perturbed and then the solution of the perturbed self-consistent equation is close to the solution of the original equation. Additionally, the solution to the self-consistent equation normally has a number of nice analytic properties in the region $\mathfrak{R}$ mentioned above, such as the properties outlined in Lemma 2.21 for the Stieltjes transform of the Marchenko-Pastur distribution. We prove the self-consistent equation's stability in Lemma 6.12.

Thus, the general strategy is to find a self-consistent equation for the Stieltjes transformation of your ensemble using the resolvent identities outlined in Lemma 2.13. The ensemble's self-consistent equation should have some perturbation term depending on z that is controllable, in the sense that it can be bounded by a function $f(N)$ that goes to zero as N becomes large. Typically, the control is optimal for $f(N) = 1/N$, so these produce strong local laws; for $f(N) = 1/\sqrt{N}$, for example, we can derive weak local laws. However, apart from this perturbation term, the self-consistent equation should closely resemble the self-consistent equation of the Stieltjes transform of some probability distribution. We refer to this as the global self-consistent equation and the ensemble's self-consistent equation as the approximate self-consistent equation. For the ensemble we shall consider, the approximate self-consistent equation is derived in Lemma 6.7.

Often the perturbation term in the approximate self-consistent equation is in the denominator of a term in the equation, simply due to the form of the resolvent identity. So, because of the stability of the global self-consistent equation, the approximate self-consistent equation has a solution that is provably close to the solution of the global self-consistent equation when the perturbation term is small. Naturally, we are then forced to control the size of the perturbation term.

Often the perturbation involves some quadratic form of the entries of the matrix ensemble with coefficients based on the entries of the resolvent. Thus, we use bounds, which we can prove hold with high probability, on the resolvent (see Lemma 6.10) and large deviation estimates (see Section 5) to control the size of the perturbation term. This produces bounds on the perturbation term that hold with high probability (see Lemma 6.11). Often the bounds are easier to prove for large η, as for large η many of the quantities we deal with are bounded and well behaved. This is a simple consequence of the bound (2.6). Once we have a bound on the perturbation term with high probability, the stability of the self-consistent equation tells us that the approximate self-consistent equation's solution is close to the solution of the global self-consistent equation with high-probability.

To show the bounds on the perturbation term are still valid for small η, we use a continuity (or sometimes called bootstrapping) argument, which we perform in Subsection 6.5. From our initial estimate for large η, which holds with high probability, we define a sequence of complex numbers with imaginary components that decrease from large η down to the smallest values of η ($1/N$ up to logarithmic factors). This sequence must have at most polynomially (in N) many points and the gaps between points must have size at most order N^{-8}, which is possible since the original value of large η is order 1. Then, using an induction argument, we prove that for all points in the sequence the perturbation term is small with high probability. The induction relies on the Lipschitz continuity (with Lipschitz constant greater than $\eta^{-2} \geqslant N^2$ in the region $\mathfrak{R}$) of the resolvent and the fact that the points are very close together. Moreover, since there are at most polynomially many points the perturbation term is small on all of the points concurrently with high probability (see Remark 3.4). A similar argument for a lattice covering the whole region $\mathfrak{R}$ works, which proves the bound uniformly on $\mathfrak{R}$ by the Lipschitz continuity of the resolvent.

3. Definitions and results

We begin this section by defining the class of $N \times N$ random matrices $H \equiv H_N = X_N^* X_N$ we consider. The parameter N should be thought of as large and often we shall not explicitly state the N-dependence of objects. We consider the general class of random matrices with sparse entries characterized by the parameter q, which may be N-dependent. The parameter q expresses the singularity of the distributions of the entries of X.

Definition 3.1 (H). *Fix a (possibly N-dependent) parameter $\xi \equiv \xi_N$, satisfying*

$$1 + a_0 \leqslant \xi \leqslant A_0 \log\log M, \tag{3.1}$$

where $a_0 > 0$ and $A_0 \geqslant 10$. This parameter ξ will be used as an exponent in logarithmic corrections as well as probability estimates. We consider $N \times N$ matrices of the form $H = X^*X$, where $X = (x_{ij})$ is an $M \times N$ matrix whose entries are real and independent. We assume that the entries of X satisfy the moment conditions

$$\mathbb{E}x_{ij} = 0, \quad \mathbb{E}|x_{ij}|^2 = \frac{1}{M}, \quad \textit{and } \mathbb{E}|x_{ij}|^p \leqslant \frac{C^p}{Mq^{p-2}} \tag{3.2}$$

for $1 \leqslant i \leqslant M$, $1 \leqslant j \leqslant M$, and $3 \leqslant p \leqslant (\log M)^{A_0 \log\log M}$, where C is a positive constant. We assume that the ratio of M and N, defined as $\gamma_N := N/M_N$ converges to some positive constant γ at the following rate:

$$|\gamma_N - \gamma| \leqslant \frac{C}{M}. \tag{3.3}$$

Moreover, the (possibly N-dependent) parameter $q \equiv q_N$ satisfies

$$(\log M)^{3\xi} \leqslant q \leqslant CM^{1/2}. \tag{3.4}$$

This differs from the case where the entries are Gaussian primarily because the moments of the entries decay slowly. The variance is of order M^{-1} as usually, but the higher moments decay at a rate proportional to inverse powers of q and not $N^{1/2}$ as usual. Thus, unlike Wishart matrices, the entries of the above sparse matrices satisfying Definition 3.1 do not have a natural scale. For a more precise statement see Remark 2.5 of [7].

Remark 3.2. While the matrix X is not directly related to the the adjacency matrix of the *directed Erdős-Rényi random graph*, since it is not square, this is the motivation for the assumptions we make on the moments of the entires of X. We reproduce the motivation given in [7]: suppose (here only) that X is square with $M = N$, then the assumptions on the matrix X are motivated by the adjacency matrix of the *directed Erdős-Rényi random graph*, which is the random graph where each edge exists independently with probability $0 \leqslant p \leqslant 1$. Thus the adjacency matrix, or the *directed Erdős-Rényi matrix*, has independent entries which are equal to 1 with probability p and 0 with probability $1 - p$. For convenience and a more concise statement of our results we will replace p with the new parameter $q \equiv q_N := (Np)^{1/2}$. Moreover, as usual, we will rescale the entries of our matrices so that the bulk eigenvalues typically lie in an interval of size of order 1.

Thus we give the following definition for the matrix Y. Let $Y = (y_{ij})$ be the $N \times N$ matrix whose entries are independent and identically distributed according to

$$y_{ij} := \frac{\beta}{q} \begin{cases} 1 & \text{with probability } \frac{q^2}{N} \\ 0 & \text{with probability } 1 - \frac{q^2}{N}. \end{cases} \tag{3.5}$$

Here the scaling $\beta := (1 - q^2/N)^{-1/2}$ has been introduced for convenience. Notice that the parameter $q \leqslant N^{1/2}$ expresses the sparseness of the matrix (or more generally the singularity of the distribution of x_{ij} and y_{ij}) and it may depend on N. Typically, Y has Nq^2 non vanishing entries, so we find that if $q \ll N^{1/2}$ then the matrix is called sparse.

Notice that the entries of Y are not mean-zero, so we center them and write

$$y_{ij} = x_{ij} + \frac{\beta q}{N}, \tag{3.6}$$

so that $X = (x_{ij})$ is an $N \times N$ matrix with mean-zero entries. It is easy to calculate the following moment bounds on the entries of X:

$$\mathbb{E}x_{ij}^2 = \frac{1}{N} \quad \text{and} \quad \mathbb{E}|x_{ij}|^p \leqslant \frac{1}{Nq^{p-2}} \tag{3.7}$$

for $p \geqslant 2$. In Definition 3.1 we define a more general class of random matrices which contain the above motivating example.

Many of the events we deal with, such as various inequalities holding true, are of very high probability. In particular, they quickly become increasingly likely as N grows and in the limit hold almost surely. However, for finite N adversarial events can occur rarely, so we cannot rule them out. This forces us to assume such adversarial events do not occur, which is true with very high probability. Many of our results will be stated in terms of very high probability events which can be characterized by two positive parameter, ξ and ν, where ξ is subject to assumption (3.1).

Definition 3.3 (High probability events). *Recall the restrictions on ξ in equation* (3.1). *Let Ω be an M-dependent (or equivalently N-dependent) event. We say Ω holds with (ξ, ν)-high probability if*

$$\mathbb{P}(\Omega^c) \leqslant e^{-\nu(\log M)^\xi} \tag{3.8}$$

for $M \geqslant M_0(\nu, a_0, A_0)$. We can generalize this definition: let Ω_0 be a given event, then we say that Ω holds with (ξ, ν)-high probability on Ω_0 if

$$\mathbb{P}(\Omega_0 \cap \Omega^c) \leqslant e^{-\nu(\log M)^\xi} \tag{3.9}$$

for $M \geqslant M_0(\nu, a_0, A_0)$.

Remark 3.4. Just as we mentioned in Notation 2.1, we will allow ν to decrease from one line to another without noting this or introducing a new notation. Since such changes to ν will occur only finitely many times in the proof, the constant remains positive. Hence our results will hold for all $\nu \leqslant \nu_0$ where ν_0 depends only on the constants in (3.2). An interesting and convenient property of high probability events is that the conjunction of polynomially many high probability events is also a high probability event, possibly with a reduction in the constant ν. Let $\{A_i : i \in \mathcal{I}_M\}$, where $|\mathcal{I}_M| \leqslant CM^k$ for some constant C and $k \in \mathbb{N}$, be a collection of events with (ξ_i, ν_i)-high probability respectively. Define $\xi := \min_i \xi_i$ and $\nu := \min_i \nu_i$, then, since $\nu > 0$ and $\xi > 1$,

$$\begin{aligned}
\mathbb{P}\left(\left[\bigcap_{i\in\mathcal{I}} A_i\right]^c\right) &= \mathbb{P}\left(\bigcup_{i\in\mathcal{I}} A_i^c\right) \\
&\leqslant \sum_{i\in\mathcal{I}} \mathbb{P}\left(A_i^c\right) \\
&\leqslant \sum_{i\in\mathcal{I}} e^{-\nu_i(\log M)^{\xi_i}} \\
&\leqslant CN^k e^{-\nu(\log M)^{\xi}} \\
&= e^{\log C + k\log M} e^{-\nu(\log M)^{\xi}} \\
&\leqslant e^{\nu(\log M)^{\xi}}
\end{aligned}$$

for large enough M and with a possible reduction in ν.

We now state and prove a lemma discussing the equivalence of the ensemble defined in Definition 3.1 and its cousin XX^*.

Lemma 3.5. *Suppose that H satisfies Definition 3.1. Then the nonzero eigenvalues of $H = X^*X$ are identical to the nonzero eigenvalues of $\underline{H} := XX^*$. In terms of their empirical distributions, we have*

$$\mu_H = (1-\gamma_N^{-1})\delta_0 + \gamma_N^{-1}\mu_{\underline{H}}. \tag{3.10}$$

Moreover, this implies that the Stieltjes transforms of X^*X and XX^*, denoted by $m_N \equiv m_N(z)$ and $\underline{m}_M \equiv \underline{m}_M(z)$ respectively, are related by

$$m_N(z) = (\gamma_N^{-1} - 1)\frac{1}{z} + \gamma_N^{-1}\underline{m}_M(z). \tag{3.11}$$

Proof. Denote the eigenvalues of H and $\underline{H}$ in increasing order by

$$(\lambda_\alpha)_{\alpha=1}^N \quad \text{and} \quad (\underline{\lambda}_\alpha)_{\alpha=1}^M$$

respectively, and denote their corresponding eigenvectors by

$$(\mathbf{v}_\alpha)_{\alpha=1}^N \quad \text{and} \quad (\underline{\mathbf{v}}_\alpha)_{\alpha=1}^M$$

respectively. Then by definition

$$H\mathbf{v}_\alpha = \lambda_\alpha \mathbf{v}_\alpha \quad \text{and} \quad \underline{H}\,\underline{\mathbf{v}}_\alpha = \underline{\lambda}_\alpha\,\underline{\mathbf{v}}_\alpha.$$

Suppose that $\lambda_\alpha \neq 0$, then it is an eigenvalue of $\underline{H}$ with normalized eigenvector

$$\frac{X\mathbf{v}_\alpha}{\|X\mathbf{v}_\alpha\|},$$

since

$$XX^*\frac{X\mathbf{v}_\alpha}{\|X\mathbf{v}_\alpha\|} = \frac{1}{\|X\mathbf{v}_\alpha\|}X\left(X^*X\right)\mathbf{v}_\alpha = \lambda_\alpha\frac{X\mathbf{v}_\alpha}{\|X\mathbf{v}_\alpha\|}.$$

We can show the opposite direction starting with $\underline{\lambda}_\alpha$ in exactly the same way. One can also use the Sylvester's determinant theorem to show that the nonzero spectra of H and $\underline{H}$ agree. So, for example and without loss of generality, if we assume $M \leqslant N$, then

$$\sigma(H) := (\lambda_1, \ldots, \lambda_N) = \Bigg(\underbrace{0, \ldots, 0}_{N-M}, \underline{\lambda}_1, \ldots, \underline{\lambda}_M\Bigg).$$

Thus,

$$\mu_H = \frac{1}{N} \sum_{\lambda \in \sigma(H)} \delta_\lambda = \frac{N-M}{N} + \frac{1}{N} \sum_{\lambda \in \sigma(\underline{H})} \delta_\lambda = (1 - \gamma_N^{-1})\delta_0 + \gamma_N^{-1} \mu_{\underline{H}}. \tag{3.12}$$

Moving on to the claim in equation (3.11), when we integrate (3.12), we see

$$m_N = \frac{\gamma_N^{-1} - 1}{z} + \gamma_N^{-1} \underline{m}_M, \tag{3.13}$$

which is what we wanted to show. ■

Remark 3.6. Examining relation (3.13) we can see precisely the difference between the spectra of X^*X and XX^*: the coefficient of $1/z$ results from the different number of zero eigenvalues, as can be seen from the equation for the Stieltjes transformation of a matrix in (2.22).

Notation 3.7 (Fundamental objects of X^*X and XX^* and their minors). Lemma 3.5 essentially tells us that the two ensembles H and $\underline{H}$ are equivalent and we can use the two interchangeably. We make use of this observation several times throughout the paper. To empathize which ensemble we are using at any time, we adopt the following notation hereafter: Let $H := X^*X$ be as defined in Definition 3.1, where $H = (h_{ij})$ and $X = (x_{ij})$. We denote the eigenvalues and eigenvectors of H as

$$(\lambda_\alpha)_{\alpha=1}^N \quad \text{and} \quad (\mathbf{v}_\alpha)_{\alpha=1}^N \tag{3.14}$$

respectively. Furthermore, we denote the resolvent and Stieltjes transform of H by $G \equiv G(z) = (G_{ij}(z))$ and $m_N \equiv m_N(z)$ respectively.

We use the underline to denote the corresponding objects for the ensemble of $M \times M$ matrices $\underline{H} = XX^*$: $\underline{H} = (\underline{h}_{ij})$, $(\underline{\lambda}_\alpha)_{\alpha=1}^M$, $(\underline{\mathbf{v}}_\alpha)_{\alpha=1}^M$, $\underline{G} \equiv \underline{G}(z) = (\underline{G}_{ij}(z))$, and $\underline{m}_M \equiv \underline{m}_M(z)$.

The minors of covariance matrices have a special expression in terms of the entries of X. Let $\mathbf{x}_i$ be the ith column of X and $X^{(\mathbb{T})}$ be the $M \times (N - |\mathbb{T}|)$ matrix formed by removing all the columns of X indexed by $i \in \mathbb{T}$. Then, for $i = 1$, we have

$$X = \left[\begin{array}{c|c} \mathbf{x}_1 & X^{(1)} \end{array}\right] \tag{3.15}$$

and thus

$$X^*X = \left[\begin{array}{c|c} \mathbf{x}_1^*\mathbf{x}_1 & \mathbf{x}_1^* X^{(1)} \\ \hline \left(X^{(1)}\right)^* \mathbf{x}_1 & \left(X^{(1)}\right)^* X^{(1)} \end{array}\right]. \tag{3.16}$$

So, in the notation of Subsection 2.2, we have

$$h_{11} = \mathbf{x}_1^*\mathbf{x}_1, \quad \mathbf{h}_1 = \left(X^{(1)}\right)^* \mathbf{x}_1, \quad \text{and} \quad H^{(1)} = \left(X^{(1)}\right)^* X^{(1)}. \tag{3.17}$$

Note that h_{11} and $H^{(1)}$ are independent. In general, we form the minor $H^{(\mathbb{T})}$ using $X^{(\mathbb{T})}$. Following the notation of Definition 2.10, we denote the eigenvalues, eigenvectors, resolvent, and Stieltjes transform of $H^{(\mathbb{T})}$ as

$$\left(\lambda_\alpha^{(\mathbb{T})}\right)_{\alpha=1}^{N-|\mathbb{T}|}, \quad \left(\mathbf{v}_\alpha^{(\mathbb{T})}\right)_{\alpha=1}^{N-|\mathbb{T}|}, \quad G^{(\mathbb{T})} \equiv G^{(\mathbb{T})}(z) = \left(G_{ij}^{(\mathbb{T})}(z)\right), \quad \text{and} \quad m_{N-|\mathbb{T}|}^{(\mathbb{T})} \equiv m_{N-|\mathbb{T}|}^{(\mathbb{T})}(z) \tag{3.18}$$

respectively. We also define the $M \times M$ matrix

$$\underline{H}^{(\mathbb{T})} := X^{(\mathbb{T})}\left(X^{(\mathbb{T})}\right)^*. \tag{3.19}$$

Note that this is *not* the $(\mathbb{T})$ minor of $\underline{H}$. It is merely a rank $|T|$ perturbation of $\underline{H}$. As above we define the corresponding objects

$$\left(\underline{\lambda}_\alpha^{(\mathbb{T})}\right)_{\alpha=1}^{M}, \quad \left(\underline{\mathbf{v}}_\alpha^{(\mathbb{T})}\right)_{\alpha=1}^{M}, \quad \underline{G}^{(\mathbb{T})} \equiv \underline{G}^{(\mathbb{T})}(z) = \left(\underline{G}_{ij}^{(\mathbb{T})}(z)\right), \quad \text{and} \quad \underline{m}_M^{(\mathbb{T})} \equiv \underline{m}_M^{(\mathbb{T})}(z). \tag{3.20}$$

It is also necessary for us to introduce a second type of minor: Let $\mathbf{r}_i$ be the ith row of X and $X^{[\mathbb{T}]}$ be the $(M-|\mathbb{U}|) \times N$ matrix formed by removing all the row of X indexed by $i \in \mathbb{U}$. Then, for $i = 1$, we have

$$X = \left[\begin{array}{c} \mathbf{r}_1 \\ \hline X^{[1]} \end{array}\right]. \tag{3.21}$$

and thus

$$XX^* = \left[\begin{array}{c|c} \mathbf{r}_1\mathbf{r}_1^* & \mathbf{r}_1\left(X^{[1]}\right)^* \\ \hline X^{[1]}\mathbf{r}_1^* & X^{[1]}\left(X^{[1]}\right)^* \end{array}\right]. \tag{3.22}$$

So, in the notation of Subsection 2.2, we have

$$\underline{h}_{11} = \mathbf{r}_1\mathbf{r}_1^*, \quad \underline{\mathbf{h}}_1 = X^{[1]}\mathbf{r}_1^*, \quad \text{and} \quad \underline{H}^{(1)} = X^{[1]}\left(X^{[1]}\right)^*. \tag{3.23}$$

Note that $\underline{h}_{11}$ and $\underline{H}^{[1]}$ are independent. In general, we form the minor $\underline{H}^{[\mathbb{U}]}$ using $X^{[\mathbb{U}]}$. Following the notation of Definition 2.10, we denote the eigenvalues, eigenvectors, resolvent, and Stieltjes transform of $\underline{H}^{[\mathbb{U}]}$ as

$$\left(\underline{\lambda}_\alpha^{[\mathbb{U}]}\right)_{\alpha=1}^{M-|\mathbb{U}|}, \quad \left(\underline{\mathbf{v}}_\alpha^{[\mathbb{U}]}\right)_{\alpha=1}^{M-|\mathbb{U}|}, \quad \underline{G}^{[\mathbb{U}]} \equiv \underline{G}^{[\mathbb{U}]}(z) = \left(\underline{G}_{ij}^{[\mathbb{U}]}(z)\right), \quad \text{and} \quad \underline{m}_{M-|\mathbb{U}|}^{[\mathbb{U}]} \equiv \underline{m}_{M-|\mathbb{U}|}^{[\mathbb{U}]}(z) \tag{3.24}$$

respectively. We also define the $N \times N$ matrix

$$H^{[\mathbb{U}]} := \left(X^{[\mathbb{U}]}\right)^* X^{[\mathbb{U}]}. \tag{3.25}$$

Note that this is *not* the $[\mathbb{U}]$ minor of H. As above we define the corresponding objects

$$\left(\lambda_\alpha^{[\mathbb{U}]}\right)_{\alpha=1}^{N}, \quad \left(\mathbf{v}_\alpha^{[\mathbb{U}]}\right)_{\alpha=1}^{N}, \quad G^{[\mathbb{U}]} \equiv G^{[\mathbb{U}]}(z) = \left(G_{ij}^{[\mathbb{U}]}(z)\right), \quad \text{and} \quad m_N^{[\mathbb{U}]} \equiv m_N^{[\mathbb{U}]}(z). \tag{3.26}$$

Finally, it is sometimes necessary for us to mix these two types of minor. Let $X^{(\mathbb{U},\mathbb{T})}$ be the $(M-|\mathbb{U}|) \times (N-|\mathbb{T}|)$ matrix formed by removing all the row of X indexed by $i \in \mathbb{U}$ and all the columns of X indexed by $j \in \mathbb{T}$. Then define

$$H^{(\mathbb{U},\mathbb{T})} := \left(X^{(\mathbb{U},\mathbb{T})}\right)^* X^{(\mathbb{U},\mathbb{T})} \quad \text{and} \quad \underline{H}^{(\mathbb{U},\mathbb{T})} := X^{(\mathbb{U},\mathbb{T})}\left(X^{(\mathbb{U},\mathbb{T})}\right)^*, \tag{3.27}$$

along with all their related objects.

So, in summary, the underline or lack thereof indicated the order of the matrices X and X^*. The minors are notated with parentheses $(\cdot)$ if they are formed by removing columns of X, they are notated with square brackets $[\cdot]$ if they are formed by removing rows of X, and they are notated with two argument parentheses $(\cdot\,,\cdot)$ if they are formed by removing both rows (in the first coordinate) and columns (in the second coordinate).

We now list our results. We introduce the spectral parameter

$$z := E + \mathrm{i}\eta, \tag{3.28}$$

where $E \in \mathbb{R}$ and $\eta > 0$. For

$$L \equiv L_N \geqslant 8\xi, \tag{3.29}$$

where ξ satisfies the assumption (3.1), define the region

$$\mathfrak{R}_L := \{z := E + \mathrm{i}\eta \in \mathbb{C} : \mathbf{1}_{\gamma>1}(\lambda_-/5) \leqslant E \leqslant 5\lambda_+ \text{ and } (\log M)^L M^{-1} \leqslant \eta \leqslant 10(1+\gamma)\}, \tag{3.30}$$

where $\lambda_\pm$ are given by the Marchenko-Pastur distribution, see equation (2.46). Extend this definition to the region

$$\mathfrak{R} := \{z := E + \mathrm{i}\eta \in \mathbb{C} : \mathbf{1}_{\gamma>1}(\lambda_-/5) \leqslant E \leqslant 5\lambda_+ \text{ and } M^{-1} \leqslant \eta \leqslant 10(1+\gamma)\}. \tag{3.31}$$

which, for all L, encloses $\mathfrak{R}_L$. Note that for $z \in \mathfrak{R}$, Lemma 2.21 tells us $m_{\mathrm{mp}}(z) \sim 1$. Finally, for $z := E + \mathrm{i}\eta$, we introduce the quantity

$$\kappa \equiv \kappa(z) := \min\{|\lambda_+ - E|\,, |E - \lambda_-|\}, \tag{3.32}$$

which records how close the energy of the spectral parameter z is to the spectral edges of the Marchenko-Pastur distribution.

Theorem 3.8. *Let H be as defined in Definition 3.1. Then there are constants $\nu > 0$ and $C > 0$ such that the following statements hold for any ξ satisfying* (3.1) *and L satisfying* (3.29). *The events*

$$\bigcap_{z\in\mathfrak{R}_L} \left\{ \max_{i\neq j} |G_{ij}(z)| \leqslant \frac{C(\log M)^{\xi}}{q} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/2}} \right\}, \tag{3.33}$$

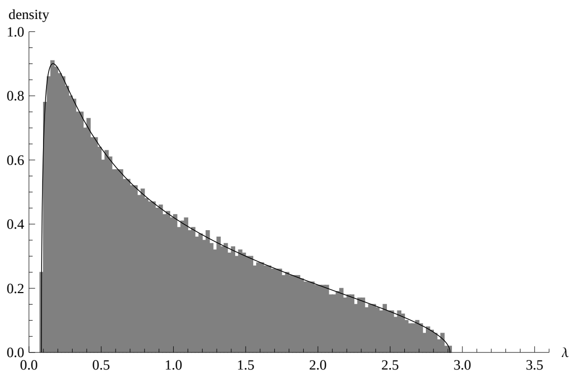
$$\bigcap_{z\in\mathfrak{R}_L} \left\{ \max_{i} |G_{ii}(z) - m_N(z)| \leqslant \frac{C(\log M)^{\xi}}{q} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/2}} \right\}, \tag{3.34}$$

$$\text{and} \quad \bigcap_{z\in\mathfrak{R}_L} \left\{ \max_{i} |G_{ii}(z) - m_{mp}(z)| \leqslant \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}} \right\} \tag{3.35}$$

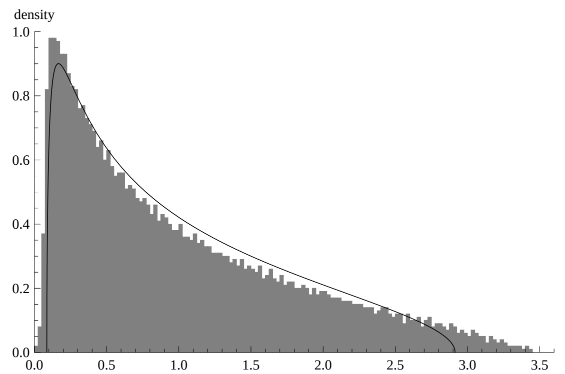
hold with (ξ,ν)-high probability. Furthermore, we have the weak local semicircle law: the event

$$\bigcap_{z\in\mathfrak{R}_L} \left\{ |m_N(z) - m_{mp}(z)| \leqslant \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}} \right\} \tag{3.36}$$

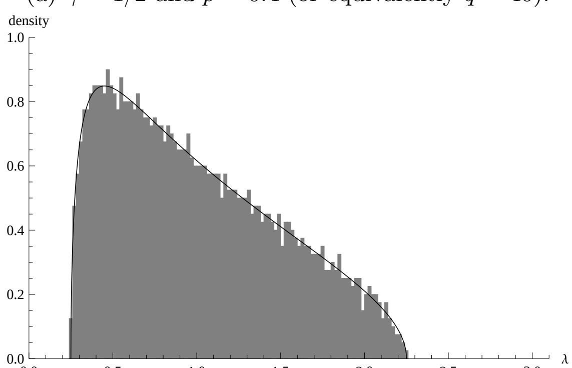
holds with (ξ,ν)-high probability.



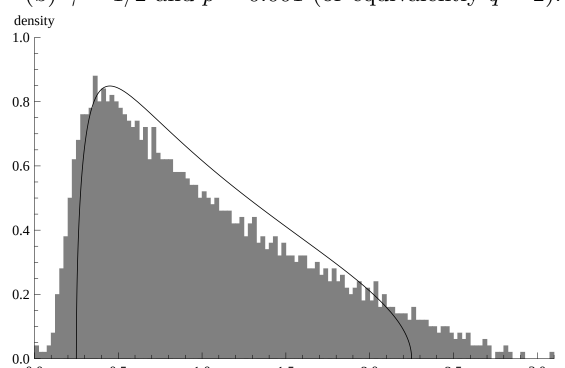
(a) $\gamma = 1/2$ and $p = 0.4$ (or equivalently $q = 40$).



(b) $\gamma = 1/2$ and $p = 0.001$ (or equivalently $q = 2$).



(c) $\gamma = 1/4$ and $p = 0.4$ (or equivalently $q = 20\sqrt{2}$).



(d) $\gamma = 1/4$ and $p = 0.001$ (or equivalently $q = \sqrt{2}$).

Figure 3: This figure shows the spectra of four sparse covariance matrices. The entries were sampled according to a Bernoulli distribution with parameter p and then the matrix was normalized. In each subfigure we plot a histogram of the eigenvalues and the profile predicted by the Marchenko-Pastur distribution, to compare the two. In figure 3a the matrix X has dimension $8,000 \times 4,000$ and q is relatively large, and we see that the Marchenko-Pastur distribution agrees closely with the spectrum. In subfigure 3b the dimensions of X are the same but q is relatively small, and thus the agreement is less accurate. In subfigures 3c and 3d X has dimension $8,000 \times 2,000$, and we observe the same phenomenon for a different value of γ. Note this dependence on q is to be expected given the statement of Theorem 3.8.

4. Moment calculations

In this section we calculate the moments of the entries of the matrix H, as defined in Definition 3.1, up to their leading orders. Before that we prove a (ξ, ν)-high probability bound on the entries of X and calculate $\mathbb{E} \operatorname{Tr} H^2$ as we discussed in Subsection 2.4, which serves as a nice warm up for Lemma 4.3 and the large deviation estimates of Section 5. Through this section we use the following notation for the moments of entries of X,

$$\mu_k := \mathbb{E}|x_{11}|^k. \tag{4.1}$$

for integers $k \geqslant 0$. Note that in Definition 3.1, it is not assumed that the random variables (x_{ij}) are identically distribution, only that they satisfy the common moment bound (3.2). This means that we do not necessarily have $\mu_k = \mathbb{E}\,|x_{ij}|^k$ for general i and j, however since we are interested only in asymptotics in this section, without loss of generality, we ignore this dependence on i and j.

Lemma 4.1. *Choose K large enough so that $K \geqslant C/e$ for C depending on* (3.2)*, then we have with (ξ, ν)-high probability*

$$|x_{ij}| \leqslant \frac{K}{q}. \tag{4.2}$$

Proof. Choose $p := \nu(\log M)^\xi$ and then apply Markov's inequality to show

$$\begin{aligned}
\mathbb{P}\left(|x_{ij}| > \frac{K}{q}\right) &= \mathbb{P}\left(|x_{ij}|^p > \frac{K^p}{q^p}\right) \\
&\leqslant \frac{\mathbb{E}\,|x_{ij}|^p}{K^p/q^p} \\
&\leqslant \frac{C^p}{Mq^{p-2}} \frac{q^p}{K^p} \\
&\leqslant \frac{q^2}{M}\left(\frac{C}{K}\right)^p \\
&\leqslant e^{-\nu(\log M)^\xi},
\end{aligned}$$

where we used the moment bound (3.2) in the third line and the bound (3.4) on q in the last line. ■

Lemma 4.2. *Let $H = (h_{ij})$ be as defined in Definition 3.1. Then*

$$\mathbb{E} \operatorname{Tr} H^2 \sim N. \tag{4.3}$$

Proof. We calculate, using the definition $H = X^*X$, to see

$$h_{ij} = \sum_{k=1}^{M} x_{ki}x_{kj}$$

and thus

$$\left[H^2\right]_{ii} = \sum_{j=1}^{N} h_{ij}h_{ji} = \sum_{j=1}^{N}\left(\sum_{k=1}^{M} x_{ki}x_{kj}\right)\left(\sum_{s=1}^{M} x_{si}x_{sj}\right).$$

Therefore, we have the following expression

$$\mathbb{E}\,\mathrm{Tr}\,H^2 = \sum_{i,j=1}^{N}\sum_{k,s=1}^{M}\mathbb{E}x_{ki}x_{kj}x_{si}x_{sj}.$$

The value of the summand is completely dependent on the assignment of indices i, j, k, and l, since the random variables are centered and independent. So, we find

$$\mathbb{E}x_{ki}x_{kj}x_{si}x_{sj} = \begin{cases} \mu_4 & \text{if } i=j \text{ and } k=s \\ \mu_2^2 & \text{if } i=j \text{ and } k\neq s \\ \mu_2^2 & \text{if } i\neq j \text{ and } k=s \\ 0 & \text{if } i\neq j \text{ and } k\neq s. \end{cases}$$

Hence,

$$\mathbb{E}\,\mathrm{Tr}\,H^2 = \sum_{i=1}^{N}\sum_{k=1}^{M}\mu_4 + \sum_{i=1}^{N}\sum_{k\neq s}\mu_2^2 + \sum_{i\neq j}\sum_{k=1}^{M}\mu_2^2 = MN\mu_4 + MN(N-1)\mu_2^2 + M(M-1)N\mu_2^2,$$

by some simple counting. Therefore, invoking the moment bound (3.2), we find

$$\mathbb{E}\,\mathrm{Tr}\,H^2 \leqslant C^4\frac{N}{q^2} + \gamma_N(M+N-2) \leqslant CN,$$

by the assumption (3.4) on q. Similarly,

$$\mathbb{E}\,\mathrm{Tr}\,H^2 \geqslant MN(M+N-2)N\mu_2^2 \geqslant \gamma_N(M+N-2) \geqslant cN,$$

which completes the proof. ∎

Lemma 4.3 (Moments of the entries of H). *Let $H=(h_{ij})$ be as defined in Definition 3.1. Then we have the following leading order bounds,*

$$\mathbb{E}h_{ij}^p \leqslant \begin{cases} \frac{C^{2p}}{q^{2p}} & \textit{if } i\neq j \textit{ and } q < M^{1/4} \\ \frac{C^{2p}}{M^{p/2}} & \textit{if } i\neq j \textit{ and } q \geqslant M^{1/4} \\ C^{2p} & \textit{if } i=j \end{cases} \tag{4.4}$$

for all $3 \leqslant p \leqslant (\log M)^{A_0 \log\log M}$

Proof. We will now calculate the moments of the entries of $H=(h_{ij})$ to the leading order. There are two cases for the moment calculations: $i=j$, for the diagonal terms, and $i\neq j$, for the off-diagonal terms, which give different asymptotics. This is what we would expect and is true of non-sparse covariance matrices too. We will calculate the first two moments precisely and obtain leading order terms for higher moments. By a simple calculation

$$h_{ij} = \begin{cases} \sum_{k=1}^{M} x_{ki}x_{kj} & \text{if } i\neq j \\ \sum_{k=1}^{M} x_{ki}^2 & \text{if } i=j \end{cases}. \tag{4.5}$$

So for the mean of the entries, we have

$$\mathbb{E}h_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}. \tag{4.6}$$

First, we calculate their variance in the case where $i \neq j$. Since the variables are centered and independent, only the terms where a variable occurs twice contribute to the sum. Thus,

$$\mathbb{E}h_{ij}^2 = \sum_{k_1,k_2=1}^{M} \mathbb{E}x_{k_1 i}x_{k_2 i}x_{k_1 j}x_{k_2 j} = \sum_{k_1,k_2=1}^{M} \mathbb{E}x_{k_1 i}x_{k_2 i} \cdot \mathbb{E}x_{k_1 j}x_{k_2 j} = \sum_{k=1}^{M} \mathbb{E}x_{ki}^2 \cdot \mathbb{E}x_{kj}^2 = M \cdot \mu_2^2 = \frac{1}{M},$$

where we used the fact that $i \neq j$ for the second equality. Second, we calculate their variance in the case where $i = j$. By similar reasoning to above

$$\mathbb{E}h_{ii}^2 = \sum_{k_1,k_2=1}^{M} \mathbb{E}x_{k_1 i}^2 x_{k_2 i}^2 = \sum_{k=1}^{M} \mathbb{E}x_{ki}^4 + \sum_{k_1 \neq k_2} \mathbb{E}x_{k_1 i}^2 \cdot \mathbb{E}x_{k_2 i}^2 = M\mu_4 + M(M-1)\mu_2^2 = M\mu_4 + \frac{M-1}{M}.$$

Summarizing, we get

$$\mathbb{E}h_{ij}^2 = \begin{cases} \frac{1}{M} & \text{if } i \neq j \\ M\mu_4 + \frac{M-1}{M} & \text{if } i = j \end{cases}$$

and using the moment bound from equation (3.2), we see

$$\mathbb{E}|h_{ij}|^2 \leqslant \begin{cases} \frac{1}{M} & \text{if } i \neq j \\ \frac{C^4}{q^2} + 1 & \text{if } i = j \end{cases}. \tag{4.7}$$

More generally, we can look at the pth moments of the entries:

$$\mathbb{E}h_{ij}^p = \begin{cases} \sum_{k_1,\ldots,k_p=1}^{M} \mathbb{E}x_{k_1 i} \cdots x_{k_p i} \cdot \mathbb{E}x_{k_1 j} \cdots x_{k_p j} & \text{if } i \neq j \\ \sum_{k_1,\ldots,k_p=1}^{M} \mathbb{E}x_{k_1 i}^2 \cdots x_{k_p i}^2 & \text{if } i = j \end{cases}. \tag{4.8}$$

First, we deal with the off-diagonal case, $i \neq j$. Since the entries of X are independent, we have

$$\mathbb{E}h_{ij}^p = \sum_{k_1,\ldots,k_p=1}^{M} \left(\mathbb{E}x_{k_1 i} \cdots x_{k_p i}\right)^2. \tag{4.9}$$

It is clear that each assignment of the indices $k_1, \ldots, k_p$ induces a partition (or equivalence relation) Π on the set $\{1, \ldots, p\}$, where two elements, i and j, of the set are in the same block if and only if $k_i = k_j$. Each partition Π is characterized by l blocks with sizes $r_1, \ldots, r_l$ such that

$$r_1 + \cdots + r_l = p.$$

We reorganize the sum in (4.9) so that we are summing over all partitions Π. Note that distinct assignments can yield identical partitions. Moreover, the value of the summand is completely determined by the partition Π. So let n_Π denote the number of assignments to the indices $k_1, \ldots, k_p$ that induce the partition Π. Now, since the entries of X are centered, we can conclude that any summand corresponding to a partition Π with a block of size 1 will be zero. So we may restrict to summing over partitions Π such that $r_k \geqslant 2$ for all $1 \leqslant k \leqslant l$, and thus $l \leqslant p/2$. Then, using the fact that the entries of X are independent, we can rewrite the sum (4.9) as

$$\sum_{\Pi} n_{\Pi} \prod_{k=1}^{l} (\mathbb{E}x_{11}^{r_k})^2 \leqslant \sum_{\Pi} n_{\Pi} \prod_{k=1}^{l} \mu_{r_k}^2 \leqslant \sum_{\Pi} n_{\Pi} \prod_{k=1}^{l} \frac{C^{2r_k}}{M^2 q^{2r_k-4}} = C^{2p} \sum_{\Pi} n_{\Pi} \frac{1}{M^{2l} q^{2p-4l}} \tag{4.10}$$

where we have used the moment bound from equation (3.2). Now we use the trivial bound

$$n_{\Pi} \leqslant M^l. \tag{4.11}$$

Using (4.11) in equation (4.10) we see

$$C^{2p} \sum_{\Pi} n_{\Pi} \frac{1}{M^{2l} q^{2p-4l}} \leqslant C^{2p} \sum_{\Pi} \frac{1}{M^{l} q^{2p-4l}} = C^{2p} \sum_{\Pi} \frac{1}{q^{2p}} \left(\frac{q^4}{M}\right)^l$$

Either $q^4 < M$, in which case the leading order term is when $l = 1$ and we get the bound

$$\frac{C^{2p}}{q^{2p}},$$

since there is one such partition such that $l = 1$. Otherwise $q^4 \geqslant M$, in which case the leading order term is when $l = p/2$ and we get the bound

$$\frac{C^{2p}}{M^{p/2}}$$

since there is one such partition such that $l = p/2$.

Second, we turn to the diagonal case, $i = j$. As before, we will sum over partitions Π, however we must now sum over all partition (not only those with blocks of size at least 2, so it is possible that $p = l$) as each random variable in squared. Thus we may rewrite the second sum of (4.8) as

$$\sum_{\Pi} n_{\Pi} \prod_{k=1}^{l} \mathbb{E}x_{11}^{2r_k} \leqslant \sum_{\Pi} n_{\Pi} \prod_{k=1}^{l} \mu_{2r_k} \leqslant \sum_{\Pi} n_{\Pi} \prod_{k=1}^{l} \frac{C^{2r_k}}{M q^{2r_k-2}} = C^{2p} \sum_{\Pi} n_{\Pi} \frac{1}{M^l q^{2p-2l}} \leqslant C^{2p} \sum_{\Pi} \frac{1}{q^{2p-2l}},$$

where we used the bounds (3.2) and (4.11). Since q grows with M, for M large enough the leading order comes from the term when $l = p$—that is, each block has size 1. There is exactly one such partition. So, the leading order for the diagonal case is

$$C^{2p},$$

as we claimed. ■

Remark 4.4. We can interpret Lemma 4.3 as follows: the dominant term for the diagonal entries comes from the product of the second moments ($l = p$). Whereas, for the off-diagonal entries the dominant term comes from the highest moment of entries of X ($l = 1$) when $q < M^{1/4}$ and the product of the second moments ($l = p/2$) when $q \geqslant M^{1/4}$. The transition point $q = M^{1/4}$ in the moments of the off diagonal entries occurs because the bound

$$\mathbb{E}h_{ij}^p \leqslant \frac{C^{2p}}{q^{2p}}$$

is too strong when $q > M^{1/4}$. That is, we would have the bound

$$\mathbb{E}h_{ij}^p < \frac{C^{2p}}{M^{p/2}},$$

which is a faster decay than is exhibited by choosing X to have Gaussian entries. Thus, we must settle for the weaker bound

$$\mathbb{E}h_{ij}^p \leqslant \frac{C^{2p}}{M^{p/2}}.$$

This plateau in the moment bounds is exactly what we should expect and essentially what we are finding is that covariance matrices are only sparse when $q < M^{1/4}$, rather than $q < M^{1/2}$.

5. Large deviation estimates

In this section we state and prove a large deviation estimate for random variables satisfying the moment conditions outlined in Definition 3.1. The argument's format is typical: in Lemma 5.3 we prove high moment bounds for the sums in question by expanding the sums and using combinatorial arguments to bound them. Then we complete the proof of Lemma 5.1 using Markov's inequality in the same spirt as the proof of Lemma 4.1. This large deviation estimate was first proved in [7] and a simplified proof was presented in [9].

Lemma 5.1. *Let $(a_i)_{i=1}^M$ be centered and independent random variables satisfying*

$$\mathbb{E}|a_i|^p \leqslant \frac{C^p}{Mq^{p-2}} \tag{5.1}$$

for $2 \leqslant p \leqslant (\log M)^{A_0 \log\log M}$. Then there is a $\nu > 0$, depending only on C in (5.1), such that for all ξ satisfying (3.4), and for any $A_i \in \mathbb{C}$ and $B_{ij} \in \mathbb{C}$, we have with (ξ, ν)-high probability

$$\left|\sum_{i=1}^M A_i a_i\right| \leqslant (\log M)^\xi \left[\frac{\max_i |A_i|}{q} + \left(\frac{1}{M}\sum_{i=1}^M |A_i|^2\right)^{1/2}\right], \tag{5.2}$$

$$\left|\sum_{i=1}^M B_{ii}\left(a_i^2 - \mathbb{E}a_i^2\right)\right| \leqslant (\log M)^\xi \, \frac{B_d}{q}, \tag{5.3}$$

$$\text{and} \quad \left|\sum_{1\leqslant i\neq j\leqslant M} a_i B_{ij} a_j\right| \leqslant (\log M)^{2\xi}\left[\frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{1\leqslant i\neq j\leqslant M}|B_{ij}|^2\right)^{1/2}\right], \tag{5.4}$$

where

$$B_d := \max_i |B_{ii}| \quad \text{and} \quad B_o := \max_{i\neq j}|B_{ij}|\,. \tag{5.5}$$

Furthermore, let $(a_i)_{i=1}^M$ and $(b_i)_{i=1}^M$ be independent random variables, each satisfying (5.1). Then there is a constant $\nu > 0$, depending only on C in (5.1), such that for all ξ satisfying (3.1) and $B_{ij} \in \mathbb{C}$ we have with (ξ, ν)-high probability

$$\left|\sum_{i,j=1}^M a_i B_{ij} b_j\right| \leqslant (\log M)^{2\xi}\left[\frac{B_d}{q^2} + \frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{1\leqslant i\neq j\leqslant M}|B_{ij}|^2\right)^{1/2}\right]. \tag{5.6}$$

Remark 5.2. A simple intuition for the bounds in Lemma 5.1 is to think of the two terms on the right-hand side of (5.2) as follows: the first term comes from the special case where the A_i are dominated by a very large maximum term $A := \max_i |A_i|$. In this case, it is easy to see $|Aa_i| \leqslant (\log M)^\xi A q^{-1}$ with (ξ, ν)-high probability. The second term comes from the variance of $\sum_i A_i a_i$. Moreover, the result is optimal, up to factors of $\log M$. However, the powers of q are not optimal, but this has no effect on later applications. The other bounds have similar interpretations.

Note that we assume here that the coefficients A_i and B_{ij} are deterministic. However, the coefficients can be random, so long as they are independent of the random variables $(a_i)_{i=1}^M$ and $(b_i)_{i=1}^M$. In this case we just take a partial expectation, and it is in this sense that we frequently make use of Lemma 5.1.

To prove Lemma 5.1, we need the following lemma, which provides bounds on the higher moments of the sums in question.

LEMMA 5.3. *Let $(a_i)_{i=1}^M$ and $(b_i)_{i=1}^M$ be real, centered, and independent random variables satisfying*

$$\mathbb{E}|a_i|^p \leqslant \frac{C^p}{Mq^{p-2}} \quad \text{and} \quad \mathbb{E}|b_i|^p \leqslant \frac{C^p}{Mq^{p-2}} \tag{5.7}$$

for all $2 \leqslant p \leqslant (\log N)^{A_0 \log\log N}$. Then for all even p satisfying $2 \leqslant p \leqslant (\log M)^{A_0 \log\log M}$ and all $A_i, B_{ij} \in \mathbb{C}$ we have

$$\mathbb{E}\biggl|\sum_{i=1}^M A_i a_i\biggr|^p \leqslant (Cp)^p \left[\frac{\max_i |A_i|}{q} + \left(\frac{1}{M}\sum_{i=1}^M |A_i|^2\right)^{1/2}\right]^p, \tag{5.8}$$

$$\mathbb{E}\biggl|\sum_{i=1}^M B_{ii}\left(a_i^2 - \mathbb{E}a_i^2\right)\biggr|^p \leqslant (Cp)^p \left[\frac{\max_i |B_{ii}|}{q} + \left(\frac{1}{Mq^2}\sum_{i=1}^M |B_{ii}|^2\right)^{1/2}\right]^p, \tag{5.9}$$

$$\mathbb{E}\left|\sum_{i=1}^M a_i A_i b_i\right|^p \leqslant (Cp)^p \left[\frac{\max_i |A_i|}{q^2} + \left(\frac{1}{M^2}\sum_{i=1}^M |A_i|^2\right)^{1/2}\right]^p, \tag{5.10}$$

$$\text{and} \quad \mathbb{E}\left|\sum_{1 \leqslant i \neq j \leqslant M} a_i B_{ij} a_j\right|^p \leqslant (Cp)^{2p} \left[\frac{\max_{i\neq j}|B_{ij}|}{q} + \left(\frac{1}{M^2}\sum_{1\leqslant i\neq j\leqslant M}|B_{ij}|^2\right)^{1/2}\right]^p, \tag{5.11}$$

for some C depending only on the constant in (5.7).

PROOF. It is easier to prove (5.8), (5.9), and (5.10), so we start by showing those. We set $p = 2r$ and compute

$$\mathbb{E}\biggl|\sum_{i=1}^M A_i a_i\biggr|^{2r} = \sum_{i_1,\dots,i_{2r}=1}^M \overline{A}_{i_1} \cdots \overline{A}_{i_r} A_{i_{r+1}} \cdots A_{i_{2r}}\, \mathbb{E}\, a_{i_1} \cdots a_{i_r} a_{i_{r+1}} \cdots a_{i_{2r}}\,. \tag{5.12}$$

It is clear that each assignment of the indices $i_1, \dots, i_{2r}$ induces a partition (or equivalence relation) Π on the set $\{1, \dots, 2r\}$, where two elements, i and j, of the set are in the same block if and only if $i_j = i_k$. Each partition Π is characterized by l blocks with sizes $r_1, \dots, r_l$ such that

$$r_1 + \cdots + r_l = 2r.$$

We reorganize the sum in (5.12) so that we are summing over all partitions Π. Note that distinct assignments can yield identical partitions. Moreover, the value of the summand is completely determined by the partition Π. So let n_Π denote the number of assignments to the indices $i_1, \dots, i_{2r}$ that induce the partition Π. Now, since the random variables a_i are centered, we can conclude that any summand corresponding to a partition Π with a block of size one will be zero. So we may restrict to summing over partitions Π such that $r_i \geqslant 2$ for all $1 \leqslant i \leqslant l$ and $l \leqslant r$. Then, using the fact that the random variables a_i are independent, we can rewrite the sum (5.12) and thus find that the contribution of the partition Π to (5.12) is bounded in absolute value by

$$\sum_{i_1,\dots,i_l} \prod_{s=1}^l |A_{i_s}|^{r_s} \mathbb{E}|a_{i_s}|^{r_s} \leqslant \prod_{s=1}^l \left(\sum_{i=1}^M |A_i|^{r_s} \frac{C^{r_s}}{Mq^{r_s-2}}\right), \tag{5.13}$$

by the bound in equation (5.7). Abbreviating $A := \max_i|A_i|$, we find that (5.13) is bounded by

$$\begin{aligned}
\prod_{s=1}^{l}\biggl((CAq^{-1})^{r_s}A^{-2}M^{-1}q^2\sum_{i=1}^{M}|A_i|^2\biggr) &= (CAq^{-1})^{2r}\biggl(\frac{1}{A^2Mq^{-2}}\sum_{i=1}^{M}|A_i|^2\biggr)^{l} \\
&\leqslant (CAq^{-1})^{2r}\max\biggl\{1,\biggl(\frac{1}{A^2Mq^{-2}}\sum_{i=1}^{M}|A_i|^2\biggr)^{r}\biggr\} \\
&\leqslant C^r\Biggl[\frac{A}{q}+\biggl(\frac{1}{M}\sum_{i=1}^{M}|B_{ii}|^2\biggr)^{1/2}\Biggr]^{2r}.
\end{aligned}$$

Next, it is easy to see that the total number of partitions of $2r$ elements is bounded by $(Cr)^{2r}$, so that we get

$$\mathbb{E}\biggl|\sum_{i=1}^{M}A_ia_i\biggr|^{2r} \leqslant (Cr)^{2r}\Biggl[\frac{A}{q}+\biggl(\frac{1}{M}\sum_{i=1}^{M}|A_i|^2\biggr)^{1/2}\Biggr]^{2r} = (Cp)^p\Biggl[\frac{\max_i|A_i|}{q}+\biggl(\frac{1}{M}\sum_{i=1}^{M}|B_{ii}|^2\biggr)^{1/2}\Biggr]^{p}.$$

This concludes the proof of (5.8).

The proof of (5.9) is very similar but has a slight modification. Let $b_i := a_i^2 - \mathbb{E}a_i^2$, then we set $p = 2r$ and, as before, compute

$$\mathbb{E}\biggl|\sum_{i=1}^{M}B_{ii}b_i\biggr|^{2r} = \sum_{i_1,\dots,i_{2r}=1}^{M}\overline{B}_{i_1i_1}\cdots\overline{B}_{i_ri_r}B_{i_{r+1}i_{r+1}}\cdots B_{i_{2r}i_{2r}}\,\mathbb{E}\,\bar{b}_{i_1}\cdots\bar{b}_{i_r}b_{i_{r+1}}\cdots b_{i_{2r}}\,. \tag{5.14}$$

We reorganize the sum in (5.14) in the same way as before, so that we are summing over all partitions Π. Again, since the random variables b_i are centered, we can conclude that any summand corresponding to a partition Π with a block of size one will be zero. So we may restrict to summing over partitions Π such that $r_i \geqslant 2$ for all $1 \leqslant i \leqslant l$ and $l \leqslant r$. Then, using the fact that the random variables b_i are independent, we can rewrite the sum (5.14) and thus find that the contribution of the partition Π to (5.14) is bounded in absolute value by

$$\sum_{i_1,\dots,i_l}\prod_{s=1}^{l}|B_{i_si_s}|^{r_s}\mathbb{E}|b_{i_s}|^{r_s} \leqslant \prod_{s=1}^{l}\biggl(\sum_{i=1}^{M}|B_{ii}|^{r_s}\frac{C^{r_s}}{Mq^{2r_s-2}}\biggr), \tag{5.15}$$

since

$$\mathbb{E}\,|b_i|^p = \mathbb{E}\left(\left|a_i^2\right|+\left|\mathbb{E}a_i^2\right|\right)^p \leqslant 2^p\mathbb{E}|a_i|^{2p}+2^p\mathbb{E}\,\mathbb{E}|a_i|^{2p} \leqslant \frac{C^p}{Mq^{2p-2}}$$

by the bound in equation (5.7). Abbreviating $B := \max_i|B_{ii}|$, we find that (5.15) is bounded by

$$\begin{aligned}
\prod_{s=1}^{l}\biggl((CBq^{-2})^{r_s}B^{-2}M^{-1}q^2\sum_{i=1}^{M}|B_{ii}|^2\biggr) &= (CBq^{-2})^{2r}\biggl(\frac{1}{B^2Mq^{-2}}\sum_{i=1}^{M}|B_{ii}|^2\biggr)^{l} \\
&\leqslant (CBq^{-2})^{2r}\max\biggl\{1,\biggl(\frac{1}{B^2Mq^{-2}}\sum_{i=1}^{M}|B_{ii}|^2\biggr)^{r}\biggr\} \\
&\leqslant C^r\Biggl[\frac{B}{q^2}+\biggl(\frac{1}{Mq^2}\sum_{i=1}^{M}|B_{ii}|^2\biggr)^{1/2}\Biggr]^{2r}.
\end{aligned}$$

As before, it is easy to see that the total number of partitions of $2r$ elements is bounded by $(Cr)^{2r}$, so that we get

$$\mathbb{E}\left|\sum_{i=1}^{M} B_{ii} b_i\right|^{2r} \leqslant (Cr)^{2r}\left[\frac{B}{q^2}+\left(\frac{1}{Mq^2}\sum_{i=1}^{M}|B_{ii}|^2\right)^{1/2}\right]^{2r} = (Cp)^p\left[\frac{\max_i|B_{ii}|}{q}+\left(\frac{1}{Mq^2}\sum_{i=1}^{M}|B_{ii}|^2\right)^{1/2}\right]^p.$$

This concludes the proof of (5.9).

The proof of (5.10) is very similar to the above, as $(a_i b_i)_{i=1}^M$ are independent centered random variables. However, they satisfy the moment condition

$$\mathbb{E}\,|a_i b_i|^p = \mathbb{E}\,|a_i|^p\,\mathbb{E}\,|b_i|^p \leqslant \frac{C^{2p}}{M^2 q^{2p-4}},$$

by equation (5.7). Thus, the final bound we get is

$$\mathbb{E}\left|\sum_{i=1}^{M} A_i a_i b_i\right|^p \leqslant (Cp)^p\left[\frac{\max_i |A_i|}{q^2}+\left(\frac{1}{M^2}\sum_{i=1}^{M}|A_i|^2\right)^{1/2}\right].$$

The proof of (5.11) is harder. The original argument in [7] uses a combinatorial argument on graphs, but here we present a simpler proof from [9]. If we formulate (5.11) in terms of L^p-norms, we find it is equivalent to

$$\left\|\sum_{i\neq j} a_i B_{ij} a_j\right\|_p^p \leqslant (Cp)^{2p}\left[\frac{\max_{i\neq j}|B_{ij}|}{q}+\left(\frac{1}{M^2}\sum_{1\leqslant i\neq j\leqslant M}|B_{ij}|^2\right)^{1/2}\right]^p.$$

The proof relies on the identity

$$1 = \frac{1}{2^{M-2}}\sum_{I\sqcup J=\mathbb{N}_M}\mathbf{1}_{i\in I}\cdot\mathbf{1}_{j\in J}, \tag{5.16}$$

for any fixed $1\leqslant i\neq j\leqslant N$, where the sum ranges over all partitions of $\mathbb{N}_M := \{1,\ldots,M\}$ into two disjoint sets I and J. Effectively, (5.16) says the number of partitions of $\mathbb{N}_M$ into two disjoint blocks I and J such that $i\in I$ and $j\in J$ is 2^{M-2}. Moreover,

$$\sum_{I\sqcup J=\mathbb{N}_M} 1 = 2^M-2, \tag{5.17}$$

where the sum ranges over nonempty subsets I and J. Effectively, (5.17) says the number of partitions of $\mathbb{N}_M$ into two disjoint blocks I and J is 2^M-2. Note that the ratio between the the second and first number is bounded by 4.

Thus, using the identity (5.16), we see

$$\left\|\sum_{i\neq j} a_i B_{ij} a_j\right\|_p = \left\|\sum_{i\neq j}\sum_{I\sqcup J=\mathbb{N}_M} a_i B_{ij} a_j\frac{\mathbf{1}_{i\in I}\cdot\mathbf{1}_{j\in J}}{2^{M-2}}\right\|_p \leqslant \frac{1}{2^{N-2}}\sum_{I\sqcup J=\mathbb{N}_M}\left\|\sum_{i\in I}\sum_{j\in J} a_i B_{ij} a_j\right\|_p. \tag{5.18}$$

Now, we can use the fact that the families $(a_i)_{i\in I}$ and $(a_j)_{j\in J}$ are independent and the obvious bounds $|I|\leqslant M$ and $|J|\leqslant M$. Define

$$A_i := \sum_{j\in J} B_{ij} a_j$$

and use (5.8) to find

$$\|A_i\|_p \leqslant (Cp)\left[\frac{\max_j|B_{ij}|}{q} + \left(\frac{1}{M}\sum_j^{(i)}|B_{ij}|^2\right)^{1/2}\right]. \tag{5.19}$$

Now, we may rewrite (5.18) as

$$\left\|\sum_{i\neq j} a_i B_{ij} a_j\right\|_p \;\leqslant\; \frac{1}{2^{M-2}} \sum_{I\sqcup J=\mathbb{N}_M} \left\|\sum_{i\in I} a_i A_i\right\|_p .$$

and use (5.8) again to get

$$\frac{1}{2^{M-2}} \sum_{I\sqcup J=\mathbb{N}_M} \left\|\sum_{i\in I} a_i A_i\right\|_p \leqslant \frac{1}{2^{M-2}} \sum_{I\sqcup J=\mathbb{N}_M} (Cp)\left[\frac{\max_i|A_i|}{q} + \left(\frac{1}{M}\sum_{i=1}^{M}|A_i|^2\right)^{1/2}\right]. \tag{5.20}$$

Combining (5.20) with the bound (5.19), we have

$$\left\|\sum_{i\neq j} a_i B_{ij} a_j\right\|_p \;\leqslant\; \frac{1}{2^{M-2}} \sum_{I\sqcup J=\mathbb{N}_M} (Cp)^2\left[\frac{\max_{i\neq j}|B_{ij}|}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right].$$

By equation (5.17), this implies

$$\left\|\sum_{i\neq j} a_i B_{ij} a_j\right\|_p \;\leqslant\; (Cp)^2\left[\frac{\max_{i\neq j}|B_{ij}|}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right],$$

which is equivalent to

$$\mathbb{E}\left|\sum_{i\neq j} a_i B_{ij} a_j\right|^p \;\leqslant\; (Cp)^{2p}\left[\frac{\max_{i\neq j}|B_{ij}|}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right]^p$$

and thus completes the proof. ■

Proof of Lemma 5.1. The proof is a simple application of Lemma 5.3 and Markov's inequality. To begin with, we show (5.2). Let $p := \nu(\log M)^\xi$ and choose ν small enough so that $\nu \leqslant \frac{1}{Ce}$. The probability that

$$\left|\sum_{i=1}^{M} A_i a_i\right| > (\log M)^\xi \left[\frac{\max_i |A_i|}{q} + \left(\frac{1}{M}\sum_{i=1}^{M}|A_i|^2\right)^{1/2}\right]$$

is equivalent to the probability that

$$\left|\sum_{i=1}^{M} A_i a_i\right|^p > (\log M)^{p\xi} \left[\frac{\max_i |A_i|}{q} + \left(\frac{1}{M}\sum_{i=1}^{M}|A_i|^2\right)^{1/2}\right]^p$$

as both sides of the inequality are positive. Moreover, using Markov's inequality, the probability is thus bounded by

$$\frac{\mathbb{E}\left|\sum_{i=1}^{M} A_i a_i\right|^p}{(\log M)^{p\xi}\left[\frac{\max_i|A_i|}{q}+\left(\frac{1}{M}\sum_{i=1}^{M}|A_i|^2\right)^{1/2}\right]^p} \tag{5.21}$$

Now we may use the result (5.8) from Lemma 5.3, to bound (5.21) by

$$\frac{(Cp)^p}{(\log M)^{p\xi}}=\left(\frac{C\nu(\log M)^{\xi}}{(\log M)^{\xi}}\right)^{\nu(\log M)^{\xi}}\leqslant e^{-\nu(\log M)^{\xi}} \tag{5.22}$$

This shows that with (ξ,ν)-high probability

$$\left|\sum_{i=1}^{M} A_i a_i\right|\leqslant(\log M)^{\xi}\left[\frac{\max_i|A_i|}{q}+\left(\frac{1}{M}\sum_{i=1}^{M}|A_i|^2\right)^{1/2}\right]. \tag{5.23}$$

Now, in the same way we show (5.3). Let $p:=\nu(\log M)^{\xi}$ and choose ν small enough so that $\nu\leqslant\frac{1}{2Ce}$. We can bound the first term using Markov's inequality and the bound in (5.9). In more detail, the probability that

$$\left|\sum_{i,j=1}^{M} B_{ij}\left(a_i a_j-\mathbb{E}a_i a_j\right)\right|>(\log M)^{\xi}\frac{B_d}{q}$$

is equivalent to the probability that

$$\left|\sum_{i,j=1}^{M} B_{ij}\left(a_i a_j-\mathbb{E}a_i a_j\right)\right|^p>(\log M)^{\xi p}\frac{B_d^p}{q^p},$$

as both sides of the inequality are positive, and the probability is thus bounded by

$$\frac{\mathbb{E}\left|\sum_{i,j=1}^{M} B_{ij}\left(a_i a_j-\mathbb{E}a_i a_j\right)\right|^p}{(\log M)^{\xi p}\frac{B_d^p}{q^p}}. \tag{5.24}$$

Now we may use the result (5.9) from Lemma 5.3, to bound (5.24) by

$$\begin{aligned}\frac{(Cp)^p}{(\log M)^{\xi p}}\frac{q^p}{B^p}\left[\frac{\max_i|B_{ii}|}{q}+\left(\frac{1}{Mq^2}\sum_{i=1}^{M}|B_{ii}|^2\right)^{1/2}\right]^p &= \frac{(Cp)^p}{(\log M)^{\xi p}}\left[1+\left(\frac{1}{M}\sum_{i=1}^{M}\frac{|B_{ii}|^2}{B_d^2}\right)^{1/2}\right]^p\\ &\leqslant (C\nu)^p\cdot 2^p\\ &\leqslant e^{-\nu(\log M)^{\xi}}.\end{aligned}$$

This shows that with (ξ,ν)-high probability

$$\left|\sum_{i,j=1}^{M} B_{ij}\left(a_i a_j-\mathbb{E}a_i a_j\right)\right|\leqslant(\log M)^{\xi}\frac{B_d}{q}. \tag{5.25}$$

Now we prove (5.4). The argument is similar and again uses Markov's inequality. Let $p := \nu(\log M)^\xi$ and choose ν small enough so that $\nu \leqslant \frac{1}{Ce}$. The probability that

$$\left|\sum_{i\neq j} a_i B_{ij} a_j\right| > (\log M)^{2\xi}\left[\frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right]$$

is equivalent to the probability that

$$\left|\sum_{i\neq j} a_i B_{ij} a_j\right|^p > (\log M)^{2\xi p}\left[\frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right]^p,$$

as both sides of the inequality are positive, and the probability is thus bounded by

$$\frac{\mathbb{E}\left|\sum_{i\neq j} a_i B_{ij} a_j\right|^p}{(\log M)^{2\xi p}\left[\frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right]^p}. \tag{5.26}$$

Now we may use the result (5.11) from Lemma 5.3, to bound (5.26) by

$$\begin{aligned}
\frac{(Cp)^{2p}\left[\frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right]^p}{(\log M)^{2\xi p}\left[\frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right]^p} &= \left(\frac{C^2p^2}{(\log M)^{2\xi}}\right)^p \\
&= \left(\frac{C^2\nu^2(\log M)^{2\xi}}{(\log M)^{2\xi}}\right)^{\nu(\log M)^\xi} \\
&\leqslant (C\nu)^{\nu(\log M)^\xi} \\
&\leqslant e^{-\nu(\log M)^\xi}.
\end{aligned}$$

This shows that with (ξ,ν)-high probability

$$\left|\sum_{i\neq j} a_i B_{ij} a_j\right| \leqslant (\log M)^{2\xi}\left[\frac{B_o}{q} + \left(\frac{1}{M^2}\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right]. \tag{5.27}$$

Finally, we prove (5.6). Using the triangle inequality, we see

$$\left|\sum_{i,j=1}^M a_i B_{ij} b_j\right| \leqslant \left|\sum_{i=1}^M a_i B_{ii} b_i\right| + \left|\sum_{i\neq j}^M a_i B_{ij} b_j\right|.$$

We can now deal with the diagonal and off-diagonal terms separately. Using Markov's inequality and (5.10) from Lemma 5.3 in the same way as above, we see

$$\left|\sum_{i=1}^M a_i B_{ij} b_j\right| \leqslant (\log M)^{\xi}\left[\frac{B_d}{q^2} + \left(\frac{1}{M^2}\sum_{i=1}^M|B_{ii}|^2\right)^{1/2}\right] \leqslant 2(\log M)^\xi\frac{B_d}{q^2} \tag{5.28}$$

with (ξ,ν)-high probability. For the off-diagonal terms, let $A_i := \sum_{j\neq i} B_{ij} b_j$. Then we can use (5.2) from Lemma 5.1, which we have already proved, to see

$$|A_i| \leqslant (\log M)^\xi \left[\frac{B_o}{q} + \left(\frac{1}{M} \sum_{j\neq i} |B_{ij}|^2 \right)^{1/2} \right]$$

with (ξ,ν)-high probability. Since A_i is independent of a_j (it is a linear combination, with deterministic coefficients, of the independent quantities b_j) we may use (5.2) again to see

$$\begin{aligned} \left| \sum_{i\neq j} a_i B_{ij} b_j \right| &= \left| \sum_{i=1}^M A_i a_i \right| \\ &\leqslant (\log M)^\xi \left[\frac{\max_i |A_i|}{q} + \left(\frac{1}{M} \sum_{i=1}^M |A_i|^2 \right)^{1/2} \right] \\ &\leqslant C(\log M)^\xi \left[\frac{B_o}{q} + \left(\frac{1}{M^2} \sum_{i\neq j} |B_{ij}|^2 \right)^{1/2} \right] \end{aligned} \tag{5.29}$$

with (ξ,ν)-high probability, since the conjunction of $\mathcal{O}(M)$ (ξ,ν)-high probability events is also a (ξ,ν)-high probability event, by Remark 3.4. So putting together the bounds (5.28) and (5.29), we get

$$\left| \sum_{i,j=1}^M a_i B_{ij} b_j \right| \leqslant (\log M)^{2\xi} \left[\frac{B_d}{q^2} + \frac{B_o}{q} + \left(\frac{1}{M^2} \sum_{i\neq j} |B_{ij}|^2 \right)^{1/2} \right],$$

where we absorbed the constant C into the smaller constant ν. ■

6. Local Marchenko-Pastur law

The work of this section is to prove a weak local law for the ensemble defined in Definition 3.1. We follow the method outlined in Subsection 2.4. So, our goal is to estimate the following quantities, which we abbreviate for convenience in the following definition.

Definition 6.1. *Let $z := E + \mathrm{i}\eta \in \mathfrak{R}$ by as defined in* (3.31). *Then define*

$$\Lambda_d \equiv \Lambda_d(z) := \max_{1 \leqslant i \leqslant N} |G_{ii}(z) - m_{mp}(z)|, \tag{6.1}$$

$$\Lambda_o \equiv \Lambda_o(z) := \max_{1 \leqslant i \neq j \leqslant N} |G_{ij}(z)|, \tag{6.2}$$

$$\textit{and} \;\; \Lambda \equiv \Lambda(z) := |m_N(z) - m_{mp}(z)|, \tag{6.3}$$

where the subscripts refer to "diagonal" and "off-diagonal" matrix element of the resolvent. Furthermore, we define similar quantities for the minors:

$$\Lambda_d^{(\mathbb{T})} \equiv \Lambda_d^{(\mathbb{T})}(z) := \max_{\substack{1 \leqslant i \leqslant N \\ i \notin \mathbb{T}}} \left| G_{ii}^{(\mathbb{T})}(z) - m_{mp}(z) \right|, \tag{6.4}$$

$$\Lambda_o^{(\mathbb{T})} \equiv \Lambda_o^{(\mathbb{T})}(z) := \max_{\substack{1 \leqslant i \neq j \leqslant N \\ i,j \notin \mathbb{T}}} \left| G_{ij}^{(\mathbb{T})} \right|, \tag{6.5}$$

$$\textit{and} \;\; \Lambda^{(\mathbb{T})} \equiv \Lambda^{(\mathbb{T})}(z) := \left| m_N^{(\mathbb{T})}(z) - m_{mp}(z) \right|. \tag{6.6}$$

The quantities $\underline{\Lambda}_d^{(\mathbb{T})}$, $\underline{\Lambda}_o^{(\mathbb{T})}$, and $\underline{\Lambda}^{(\mathbb{T})}$ are defined in the natural way for the resolvent $\underline{G}^{(\mathbb{T})}$:

$$\underline{\Lambda}_d^{(\mathbb{T})} \equiv \underline{\Lambda}_d^{(\mathbb{T})}(z) := \max_{1 \leqslant i \leqslant M} \left| \underline{G}_{ii}^{(\mathbb{T})}(z) - \left(\gamma m_{mp}(z) + \frac{\gamma - 1}{z} \right) \right|, \tag{6.7}$$

$$\underline{\Lambda}_o^{(\mathbb{T})} \equiv \underline{\Lambda}_o^{(\mathbb{T})}(z) := \max_{1 \leqslant i \neq j \leqslant M} \left| \underline{G}_{ij}^{(\mathbb{T})} \right|, \tag{6.8}$$

$$\textit{and} \;\; \underline{\Lambda}^{(\mathbb{T})} \equiv \underline{\Lambda}^{(\mathbb{T})}(z) := \left| \underline{m}_M^{(\mathbb{T})}(z) - \left(\gamma m_{mp}(z) + \frac{\gamma - 1}{z} \right) \right|, \tag{6.9}$$

where we have adjusted for the point mass at 0. We define similar quantities for the minors $[\mathbb{U}]$ *and* $(\mathbb{U}, \mathbb{T})$.

Recall the definition of $\mathfrak{R}_L$ from (3.30).

Theorem 6.2. *Let H be as defined in Definition 3.1. Then there are constants $\nu > 0$ and $C > 0$ such that the following statements hold for any ξ satisfying* (3.1) *and L satisfying* (3.29). *The events*

$$\bigcap_{z \in \mathfrak{R}_L} \left\{ \Lambda_o(z) \leqslant \frac{C(\log M)^{\xi}}{q} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/2}} \right\}, \tag{6.10}$$

$$\bigcap_{z \in \mathfrak{R}_L} \left\{ \max_i |G_{ii}(z) - m_N(z)| \leqslant \frac{C(\log M)^{\xi}}{q} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/2}} \right\}, \tag{6.11}$$

$$\textit{and} \;\; \bigcap_{z \in \mathfrak{R}_L} \left\{ \Lambda_d(z) \leqslant \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}} \right\} \tag{6.12}$$

hold with (ξ,ν)-high probability. Furthermore, we have the weak local Marchenko-Pastur law: the event

$$\bigcap_{z\in\mathfrak{R}_L}\left\{\Lambda(z)\leqslant\frac{C(\log M)^{\xi}}{\sqrt{q}}+\frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}}\right\} \tag{6.13}$$

holds with (ξ,ν)-high probability.

Roughly, Theorem 6.2 states that

$$|G_{ij}-\delta_{ij}m_N|\lesssim\frac{1}{q}+\frac{1}{\sqrt{N\eta}} \tag{6.14}$$

and

$$|m_N-m_{\mathrm{mp}}|\leqslant\max_i|G_{ii}-m_{\mathrm{mp}}|\lesssim\frac{1}{\sqrt{q}}+\frac{1}{(N\eta)^{1/4}}. \tag{6.15}$$

Note that in (6.14) the diagonal term G_{ii} is compared to the random quantity m_N and not m_{mp}, as it is in (6.15). The bound in (6.15) is not optimal, as we would expect an error term of the form $(N\eta)^{-1}$ rather than $(N\eta)^{-1/4}$; this lack of precision is due to instability near the edge. The bound can be improved if we relax the requirement that our bound be uniform up to the edge. If we only wanted a bound in the bulk, then our method could achieve a stronger bound of the form $(N\eta)^{-1/2}$, at the expense of a term that blows up near the edges. Despite the fact that the bounds are not optimal, and thus the local law is only a weak local law, the bounds hold on the optimal scale of $\eta\gtrsim M^{-1}$, uniformly up to the edge.

6.1. Preliminary lemmas. During the proof of the local Marchenko-Pastur law we will need some basic properties of the Marchenko-Pastur distribution, which are summarized in Subsection 2.3, and some other elementary lemmas, which we state and prove in this section.

LEMMA 6.3. *Let A be an $M\times N$ matrix, such that*

$$(A^*A-z\mathbb{1})\quad\textit{and}\quad(AA^*-z\mathbb{1}) \tag{6.16}$$

are invertible—that is, z is not in the spectra of A^*A or AA^*. Then we have the following identity

$$A\left(A^*A-z\mathbb{1}\right)^{-1}A^*=\mathbb{1}+z\left(AA^*-z\mathbb{1}\right)^{-1}. \tag{6.17}$$

PROOF. The proof is a simple manipulation. We have

$$\begin{aligned}AA^*&=A\left(A^*A-z\mathbb{1}\right)^{-1}\left(A^*A-z\mathbb{1}\right)A^*\\&=A\left(A^*A-z\mathbb{1}\right)^{-1}\left(A^*AA^*-zA^*\right)\\&=A\left(A^*A-z\mathbb{1}\right)^{-1}A^*\left(AA^*-z\mathbb{1}\right),\end{aligned}$$

and multiplying on the right by $\left(AA^*-z\mathbb{1}\right)^{-1}$, we get

$$AA^*\left(AA^*-z\mathbb{1}\right)^{-1}=A\left(A^*A-z\mathbb{1}\right)^{-1}A^*. \tag{6.18}$$

Thus

$$AA^*\left(AA^*-z\mathbb{1}\right)^{-1}=\left(AA^*-z\mathbb{1}+z\mathbb{1}\right)\left(AA^*-z\mathbb{1}\right)^{-1}=\mathbb{1}+z\left(AA^*-z\mathbb{1}\right)^{-1},$$

which with equation (6.18) completes the proof. ∎

Lemma 6.4 (Rank one perturbation formula). *Suppose that A is an invertible $N \times N$ matrix and $\mathbf{q}$ and $\mathbf{r}$ are column vectors of dimension N, such that*

$$1 + \mathbf{r}^* A^{-1} \mathbf{q} \neq 0. \tag{6.19}$$

Then we have the following identity

$$(A + \mathbf{q}\mathbf{r}^*)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{q}\mathbf{r}^* A^{-1}}{1 + \mathbf{r}^* A^{-1}\mathbf{q}}. \tag{6.20}$$

Proof. Again the proof is a simple calculation. Note

$$\begin{aligned} \left(A^{-1} - \frac{A^{-1}\mathbf{q}\mathbf{r}^* A^{-1}}{1 + \mathbf{r}^* A^{-1}\mathbf{q}}\right)(A + \mathbf{q}\mathbf{r}^*) &= \mathbb{1} + A^{-1}\mathbf{q}\mathbf{r}^* - \frac{A^{-1}\mathbf{q}\mathbf{r}^* + A^{-1}\mathbf{q}\left(\mathbf{r}^* A^{-1}\mathbf{q}\right)\mathbf{r}^*}{1 + \mathbf{r}^* A^{-1}\mathbf{q}} \\ &= \mathbb{1} + A^{-1}\mathbf{q}\mathbf{r}^* - \frac{A^{-1}\mathbf{q}\mathbf{r}^* + \left(\mathbf{r}^* A^{-1}\mathbf{q}\right)A^{-1}\mathbf{q}\mathbf{r}^*}{1 + \mathbf{r}^* A^{-1}\mathbf{q}} \\ &= \mathbb{1} + A^{-1}\mathbf{q}\mathbf{r}^* - \frac{1 + \mathbf{r}^* A^{-1}\mathbf{q}}{1 + \mathbf{r}^* A^{-1}\mathbf{q}} A^{-1}\mathbf{q}\mathbf{r}^* \\ &= \mathbb{1}, \end{aligned} \tag{6.21}$$

where we used the fact that $\left(\mathbf{r}^* A^{-1}\mathbf{q}\right)$ is a scalar in the second step. So, multiplying (6.21) on the right by $(A + \mathbf{q}\mathbf{r}^*)^{-1}$ yields (6.20). ∎

We shall also need the following resolvent identities for the matrix elements of the resolvents $G_{ij}^{(\mathbb{T})}$ and $\underline{G}_{ij}^{(\mathbb{T})}$, most of which are a special case of Lemma 2.13.

Lemma 6.5 (Resolvent identities). *Let $G_{ij}^{(\mathbb{T})}$ and $\underline{G}_{ij}^{(\mathbb{T})}$ be as defined in (3.7). Then for the diagonal elements of $G(z)$, we have*

$$G_{ii}(z) = \frac{1}{-z - z\mathbf{x}_i^* \underline{G}^{(i)}(z)\mathbf{x}_i} \quad \text{or equivalently} \quad \mathbf{x}_i^* \underline{G}^{(i)}(z)\mathbf{x}_i = \frac{-1}{zG_{ii}(z)} - 1. \tag{6.22}$$

For the off-diagonal elements of $G(z)$, when $i \neq j$, we have

$$G_{ij}(z) = zG_{ii}(z)G_{jj}^{(i)}(z)\mathbf{x}_i^* \underline{G}^{(ij)}(z)\mathbf{x}_j. \tag{6.23}$$

Finally, when $i \neq k$ and $j \neq k$,

$$G_{ij}(z) = G_{ij}^{(k)}(z) + \frac{G_{ik}(z)G_{kj}(z)}{G_{kk}(z)}. \tag{6.24}$$

For the $\underline{H}$ ensemble, we have the rank one perturbation formulas,

$$\underline{G} = \underline{G}^{(i)} - \frac{\underline{G}^{(i)}\mathbf{x}_i\mathbf{x}_i^*\underline{G}^{(i)}}{1 + \mathbf{x}_i^*\underline{G}^{(i)}\mathbf{x}_i} = \underline{G}^{(i)} + zG_{ii}(z)\underline{G}^{(i)}\mathbf{x}_i\mathbf{x}_i^*\underline{G}^{(i)} \tag{6.25}$$

and

$$\underline{G}^{(i)} = \underline{G} + \frac{\underline{G}\mathbf{x}_i\mathbf{x}_i^*\underline{G}}{1 - \mathbf{x}_i^*\underline{G}\mathbf{x}_i}. \tag{6.26}$$

Moreover, we have the following identity for the trace of the resolvent

$$\operatorname{Tr} G^{(\mathbb{T})}(z) = \frac{M - N + |\mathbb{T}|}{z} + \operatorname{Tr} \underline{G}^{(\mathbb{T})}(z). \tag{6.27}$$

Remark 6.6. Lemma 6.5 remains trivially valid for the minors $H^{(\mathbb{T})}$ and $\underline{H}^{(\mathbb{T})}$ of H and $\underline{H}$, and we frequently use this fact in later proofs. So, stating the results in their most general form: for the diagonal elements of $G^{(\mathbb{U},\mathbb{T})}(z)$, we have

$$G_{ii}^{(\mathbb{U},\mathbb{T})}(z) = \frac{1}{-z - z\left(\mathbf{x}_i^{(\mathbb{U})}\right)^* \underline{G}^{(\mathbb{U},i\mathbb{T})}(z)\mathbf{x}_i^{(\mathbb{U})}} \quad \text{or} \quad \left(\mathbf{x}_i^{(\mathbb{U})}\right)^* \underline{G}^{(\mathbb{U},i\mathbb{T})}(z)\mathbf{x}_i^{(\mathbb{U})} = \frac{-1}{zG_{ii}^{(\mathbb{U},\mathbb{T})}(z)} - 1. \tag{6.28}$$

For the off-diagonal elements of $G^{(\mathbb{U},\mathbb{T})}(z)$, when $i \neq j$, we have

$$G_{ij}^{(\mathbb{U},\mathbb{T})}(z) = zG_{ii}^{(\mathbb{U},\mathbb{T})}(z)G_{jj}^{(\mathbb{U},i\mathbb{T})}(z)\left(\mathbf{x}_i^{(\mathbb{U})}\right)^* \underline{G}^{(\mathbb{U},ij\mathbb{T})}(z)\mathbf{x}_j^{(\mathbb{U})}. \tag{6.29}$$

Finally, when $i \neq k$ and $j \neq k$,

$$G_{ij}^{(\mathbb{U},\mathbb{T})}(z) = G_{ij}^{(\mathbb{U},k\mathbb{T})}(z) + \frac{G_{ik}^{(\mathbb{U},\mathbb{T})}(z)G_{kj}^{(\mathbb{U},\mathbb{T})}(z)}{G_{kk}^{(\mathbb{U},\mathbb{T})}(z)}. \tag{6.30}$$

For the $\underline{H}^{(\mathbb{U},\mathbb{T})}$ ensemble, we have the rank one perturbation formulas,

$$\underline{G}^{(\mathbb{U},\mathbb{T})} = \underline{G}^{(\mathbb{U},i\mathbb{T})} - \frac{\underline{G}^{(\mathbb{U},i\mathbb{T})}\mathbf{x}_i^{(\mathbb{U})}\left(\mathbf{x}_i^{(\mathbb{U})}\right)^*\underline{G}^{(\mathbb{U},i\mathbb{T})}}{1+\left(\mathbf{x}_i^{(\mathbb{U})}\right)^*\underline{G}^{(\mathbb{U},i\mathbb{T})}\mathbf{x}_i^{(\mathbb{U})}} = \underline{G}^{(\mathbb{U},i\mathbb{T})} + zG_{ii}^{(\mathbb{U},\mathbb{T})}(z)\underline{G}^{(\mathbb{U},i\mathbb{T})}\mathbf{x}_i^{(\mathbb{U})}\left(\mathbf{x}_i^{(\mathbb{U})}\right)^*\underline{G}^{(\mathbb{U},i\mathbb{T})} \tag{6.31}$$

and

$$\underline{G}^{(\mathbb{U},i\mathbb{T})} = \underline{G}^{(\mathbb{U},\mathbb{T})} + \frac{\underline{G}^{(\mathbb{U},\mathbb{T})}\mathbf{x}_i^{(\mathbb{U})}\left(\mathbf{x}_i^{(\mathbb{U})}\right)^*\underline{G}^{(\mathbb{U},\mathbb{T})}}{1-\left(\mathbf{x}_i^{(\mathbb{U})}\right)^*\underline{G}^{(\mathbb{U},\mathbb{T})}\mathbf{x}_i^{(\mathbb{U})}}. \tag{6.32}$$

A little thought shows that one can easily prove the corresponding identities in exactly the same way: for the diagonal elements of $\underline{G}^{(\mathbb{U},\mathbb{T})}(z)$, we have

$$\underline{G}_{ii}^{(\mathbb{U},\mathbb{T})}(z) = \frac{1}{-z - z\mathbf{r}_i^{(\mathbb{T})}G^{(i\mathbb{U},\mathbb{T})}(z)\left(\mathbf{r}_i^{(\mathbb{T})}\right)^*} \quad \text{or} \quad \mathbf{r}_i^{(\mathbb{T})}G^{(i\mathbb{U},\mathbb{T})}(z)\left(\mathbf{r}_i^{(\mathbb{T})}\right)^* = \frac{-1}{z\underline{G}_{ii}^{(\mathbb{U},\mathbb{T})}(z)} - 1. \tag{6.33}$$

For the off-diagonal elements of $\underline{G}^{(\mathbb{U},\mathbb{T})}(z)$, when $i \neq j$, we have

$$\underline{G}_{ij}^{(\mathbb{U},\mathbb{T})}(z) = z\underline{G}_{ii}^{(\mathbb{U},\mathbb{T})}(z)\underline{G}_{jj}^{(i\mathbb{U},\mathbb{T})}(z)\mathbf{r}_i^{(\mathbb{T})}G^{(ij\mathbb{U},\mathbb{T})}(z)\left(\mathbf{r}_j^{(\mathbb{T})}\right)^*. \tag{6.34}$$

Finally, when $i \neq k$ and $j \neq k$,

$$\underline{G}_{ij}^{(\mathbb{U},\mathbb{T})}(z) = \underline{G}_{ij}^{(k\mathbb{U},\mathbb{T})}(z) + \frac{\underline{G}_{ik}^{(\mathbb{U},\mathbb{T})}(z)\underline{G}_{kj}^{(\mathbb{U},\mathbb{T})}(z)}{\underline{G}_{kk}^{(\mathbb{U},\mathbb{T})}(z)}. \tag{6.35}$$

For the $H^{(\mathbb{U},\mathbb{T})}$ ensemble, we have the rank one perturbation formulas,

$$G^{(\mathbb{U},\mathbb{T})} = G^{(i\mathbb{U},\mathbb{T})} - \frac{G^{(i\mathbb{U},\mathbb{T})}\left(\mathbf{r}_i^{(\mathbb{T})}\right)^*\mathbf{r}_i^{(\mathbb{T})}G^{(i\mathbb{U},\mathbb{T})}}{1+\mathbf{r}_i^{(\mathbb{T})}G^{(i\mathbb{U},\mathbb{T})}\left(\mathbf{r}_i^{(\mathbb{T})}\right)^*} = G^{(i\mathbb{U},\mathbb{T})} + z\underline{G}_{ii}^{(\mathbb{U},\mathbb{T})}(z)G^{(i\mathbb{U},\mathbb{T})}\left(\mathbf{r}_i^{(\mathbb{T})}\right)^*\mathbf{r}_i^{(\mathbb{T})}G^{(i\mathbb{U},\mathbb{T})} \tag{6.36}$$

and

$$G^{(i\mathbb{U},\mathbb{T})} = G^{(\mathbb{U},\mathbb{T})} + \frac{G^{(\mathbb{U},\mathbb{T})}\left(\mathbf{r}_i^{(\mathbb{T})}\right)^* \mathbf{r}_i^{(\mathbb{T})} G^{(\mathbb{U},\mathbb{T})}}{1 - \mathbf{r}_i^{(\mathbb{T})} G^{(\mathbb{U},\mathbb{T})}\left(\mathbf{r}_i^{(\mathbb{T})}\right)^*}. \tag{6.37}$$

Moreover, we have the following identity for the trace of the resolvent

$$\operatorname{Tr} G^{(\mathbb{U},\mathbb{T})}(z) = \frac{(M - |\mathbb{U}|) - (N - |\mathbb{T}|)}{z} + \operatorname{Tr} \underline{G}^{(\mathbb{U},\mathbb{T})}(z). \tag{6.38}$$

Proof of Lemma 6.5. We use Lemmas 2.13 and 6.3 for this proof. First we prove (6.22). Fix $1 \leqslant j \leqslant N$. Using equation (2.27) from Lemma 2.13, we find

$$G_{ii} = \frac{1}{h_{ii} - z - \left(\mathbf{h}_i^{(i)}\right)^* G^{(i)} \mathbf{h}_i^{(i)}}, \tag{6.39}$$

which we can simplify using the definition of the matrix $H = X^*X$. For ease of notation we will perform the calculation for $j = 1$ but the following is true for general j. Then

$$X = \left[\begin{array}{c|c} \mathbf{x}_1 & X^{(1)} \end{array}\right]$$

and thus

$$X^*X = \left[\begin{array}{c|c} \mathbf{x}_1^*\mathbf{x}_1 & \mathbf{x}_1^* X^{(1)} \\ \hline \left(X^{(1)}\right)^* \mathbf{x}_1 & \left(X^{(1)}\right)^* X^{(1)} \end{array}\right]. \tag{6.40}$$

Substituting this into equation (6.39), we see

$$G_{11} = \frac{1}{\mathbf{x}_1^*\mathbf{x}_1 - z - \mathbf{x}_1^* X^{(1)}\left(\left(X^{(1)}\right)^* X^{(1)} - z\mathbb{1}\right)^{-1}\left(X^{(1)}\right)^* \mathbf{x}_1}.$$

Then we use the identity in Lemma 6.3 for $A = X^{(1)}$,

$$X^{(1)}\left(\left(X^{(1)}\right)^* X^{(1)} - z\mathbb{1}\right)^{-1}\left(X^{(1)}\right)^* = \mathbb{1} + z\left(X^{(1)}\left(X^{(1)}\right)^* - z\mathbb{1}\right)^{-1},$$

to see

$$G_{11} = \frac{1}{-z - z\mathbf{x}_1^* \underline{G}^{(1)} \mathbf{x}_1}. \tag{6.41}$$

Now we prove (6.23). From (2.28), we see

$$G_{ij} = -G_{ii} G_{jj}^{(i)}\left(h_{ij} - \left(\mathbf{h}_i^{(ij)}\right)^* G^{(ij)} \mathbf{h}_j^{(ji)}\right). \tag{6.42}$$

Note that we can produce a similar equation to (6.40) by using the (ij) minor of X. For easy of notation, let $i = 1$ and $j = 2$, then we have

$$X = \left[\begin{array}{c|c|c} \mathbf{x}_1 & \mathbf{x}_2 & X^{(12)} \end{array}\right]$$

and

$$X^*X = \left[\begin{array}{c|c|c} \mathbf{x}_1^*\mathbf{x}_1 & \mathbf{x}_1^*\mathbf{x}_2 & \mathbf{x}_1^* X^{(12)} \\ \hline \mathbf{x}_2^*\mathbf{x}_1 & \mathbf{x}_2^*\mathbf{x}_2 & \mathbf{x}_2^* X^{(12)} \\ \hline \left(X^{(12)}\right)^* \mathbf{x}_1 & \left(X^{(12)}\right)^* \mathbf{x}_2 & \left(X^{(12)}\right)^* X^{(12)} \end{array}\right]. \tag{6.43}$$

When used with (6.43), (6.42) yields

$$G_{12} = -G_{11}G_{22}^{(1)} \left(\mathbf{x}_1^*\mathbf{x}_2 - \mathbf{x}_1^* X^{(12)} \left(\left(X^{(12)}\right)^* X^{(12)} - z\mathbb{1} \right)^{-1} \left(X^{(12)}\right)^* \mathbf{x}_2 \right)$$

So using Lemma 6.3, we get

$$G_{12} = zG_{11}G_{22}^{(1)}\mathbf{x}_1 \left(X^{(12)} \left(X^{(12)}\right)^* - z\mathbb{1} \right)^{-1} \mathbf{x}_2 = zG_{11}G_{22}^{(1)}\mathbf{x}_1 \underline{G}^{(12)}\mathbf{x}_2.$$

The proof is identical for general i and j.

Equation (6.24) is identical to (2.29) from Lemma 2.13. Equation (6.25) follows immediately from Lemma 6.4 with $A = \underline{H}^{(i)} - z\mathbb{1}$ and $\mathbf{q} = \mathbf{r} = \mathbf{x}_i$, since

$$\underline{H} = \underline{H}^{(i)} + \mathbf{x}_i\mathbf{x}_i^*.$$

Using (6.22) completes the proof. Similarly, (6.26) follows immediately from Lemma 6.4 with $A = \underline{H} - z\mathbb{1}$, $\mathbf{q} = \mathbf{x}_i$, and $\mathbf{r} = -\mathbf{x}_i$. Equation (6.27) can be shown in exactly the same way as (3.11) of Lemma 3.5. ■

The following lemma contains the self-consistent resolvent equation, which is essential for our proof.

Lemma 6.7 (Approximate self-consistent equation). *We have the identity*

$$G_{ii}(z) = \frac{1}{1 - \gamma - z\gamma m_N(z) - z - \Delta_i(z)}, \tag{6.44}$$

where

$$\Delta_i \equiv \Delta_i(z) := 1 - \gamma - z\gamma m_N(z) + z\underline{m}_M^{(i)}(z) - \Gamma_i(z) \tag{6.45}$$

and

$$\Gamma_i \equiv \Gamma_i(z) := z\underline{m}_M^{(i)}(z) - z\mathbf{x}_i^* \underline{G}^{(i)}(z)\mathbf{x}_i. \tag{6.46}$$

Proof. Using equation (6.22) of Lemma 6.5, we have

$$G_{ii} = \frac{1}{-z - z\mathbf{x}_i^*\underline{G}^{(i)}\mathbf{x}_i} = \frac{1}{-z - z\underline{m}_M^{(i)} + \Gamma_i}$$

where

$$\Gamma_i := z\underline{m}_M^{(i)} - z\mathbf{x}_i^*\underline{G}^{(i)}\mathbf{x}_i.$$

We can rearrange further, to find

$$G_{ii} = \frac{1}{1 - \gamma - z\gamma m_N - z - \Delta_i},$$

where

$$\Delta_i := 1 - \gamma - z\gamma m_N + z\underline{m}_M^{(i)} - \Gamma_i,$$

which completes the proof. ■

Remark 6.8. During the proof, we show that $\max_i |\Delta_i|$ is bounded. Summing (6.44) over i and normalizing by N^{-1} yields

$$m_N = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1-\gamma - z\gamma m_N - z - \Delta_i}, \tag{6.47}$$

which is like the Marchenko-Pastur equation (2.49) but with a perturbation Δ_i. However, given that this perturbation is small, we can use a Taylor expansion, under certain assumptions, to find that m_N is close to m_{mp} with some small, controllable error term.

6.2. Basic estimates on the event $\boldsymbol{\Omega(z)}$.

Definition 6.9. *For $z := E + \mathrm{i}\eta \in \mathfrak{R}_L$, we introduce the event*

$$\Omega \equiv \Omega(z) := \left\{\Lambda_d(z) + \Lambda_o(z) + \underline{\Lambda}_d(z) + \underline{\Lambda}_o(z) \leqslant C(\log M)^{-\xi}\right\} \tag{6.48}$$

and the control parameter

$$\Psi \equiv \Psi(z) := \sqrt{\frac{\Delta(z) + \operatorname{Im} m_{mp}(z)}{M\eta}}. \tag{6.49}$$

The event Ω is desirable; if it does not hold, then Λ_d, Λ_o, $\underline{\Lambda}_d$, or $\underline{\Lambda}_0$ is too large. Note that the bounds on the event Ω are very weak, however they are often sufficient to prove much stronger bounds. So we use Ω as an assumption set and later show the event holds with (ξ,ν)-high probability, first for η order one and then for smaller η by bootstrapping.

Note that Ψ is a random variable and that on $\mathfrak{R}_L$, we have

$$\eta \geqslant (\log M)^L \frac{1}{M} \geqslant (\log M)^{8\xi}\frac{1}{M}$$

and

$$|m_{\mathrm{mp}}| \sim 1$$

as stated in Lemma 2.21. Thus, on Ω we have

$$\Psi \leqslant C(M\eta)^{-1/2} \leqslant C(\log M)^{-4\xi}, \tag{6.50}$$

since

$$\Delta = |m_N - m_{\mathrm{mp}}| \leqslant \frac{1}{N}\sum_{i=1}^{N} |G_{ii} - m_{\mathrm{mp}}| \leqslant \Lambda_d \leqslant (\log M)^{-\xi}$$

on Ω. Throughout this section we shall make use of the Ward identity

$$\sum_{j=1}^{N} |G_{ij}|^2 = \frac{1}{\eta}\operatorname{Im} G_{ii},$$

which is proved in Lemma 2.14 from Subsection 2.2.

Lemma 6.10 (Rough resolvent bounds on $\Omega(z)$). *Fix $\mathbb{T} \subseteq \{1,\dots,N\}$ such that $|\mathbb{T}| \leqslant 10$ (where 10 can be replaced by any fixed constant). For $z \in \mathfrak{R}$, there is a constant C depending on $|\mathbb{T}|$, such that the following estimates*

$$\max_{i \notin \mathbb{T}} \left| G_{ii}^{(\mathbb{T})}(z) - G_{ii}(z) \right| \leqslant C\Lambda_o(z)^2 \leqslant C(\log M)^{-2\xi}, \tag{6.51}$$

$$c \leqslant \left| G_{ii}^{(\mathbb{T})}(z) \right| \leqslant C, \tag{6.52}$$

$$\text{and} \;\; \Lambda_o^{(\mathbb{T})}(z) \leqslant C\Lambda_o(z) \leqslant C(\log M)^{-\xi} \tag{6.53}$$

hold in $\Omega(z)$. Now fix $\mathbb{T} \subseteq \{1,\dots,N\}$ such that $|\mathbb{T}| \leqslant 10$ (where 10 can be replaced by any fixed constant). Then for $z \in \mathfrak{R}_L$, there is a constant C depending on $|\mathbb{T}|$, such that the following estimates

$$\max_{i,j} \left| \underline{G}_{ij}^{(\mathbb{T})}(z) - \underline{G}_{ij}(z) \right| \leqslant C(\log M)^{-\xi}, \tag{6.54}$$

$$\max_{1 \leqslant i \leqslant M} \left| \underline{G}_{ii}^{(\mathbb{T})}(z) \right| \leqslant C, \tag{6.55}$$

$$\text{and} \;\; \underline{\Lambda}_o^{(\mathbb{T})}(z) \leqslant C(\log M)^{-\xi} \tag{6.56}$$

hold in $\Omega(z)$ with (ξ,ν)-high probability. Similarly, fix $\mathbb{U} \subseteq \{1,\dots,M\}$ such that $|\mathbb{U}| \leqslant 10$ (where 10 can be replaced by any fixed constant). For $z \in \mathfrak{R}$, there is a constant C depending on $|\mathbb{U}|$, such that the following estimates

$$\max_{i \notin \mathbb{U}} \left| \underline{G}_{ii}^{[\mathbb{U}]}(z) - \underline{G}_{ii}(z) \right| \leqslant C\underline{\Lambda}_o(z)^2 \leqslant C(\log M)^{-2\xi}, \tag{6.57}$$

$$c \leqslant \left| \underline{G}_{ii}^{[\mathbb{U}]}(z) \right| \leqslant C, \tag{6.58}$$

$$\text{and} \;\; \underline{\Lambda}_o^{[\mathbb{U}]}(z) \leqslant C\underline{\Lambda}_o(z) \leqslant C(\log M)^{-\xi} \tag{6.59}$$

hold in $\Omega(z)$. Now fix $\mathbb{U} \subseteq \{1,\dots,M\}$ such that $|\mathbb{U}| \leqslant 10$ (where 10 can be replaced by any fixed constant). Then for $z \in \mathfrak{R}_L$, there is a constant C depending on $|\mathbb{U}|$, such that the following estimates

$$\max_{i,j} \left| G_{ij}^{[\mathbb{U}]}(z) - G_{ij}(z) \right| \leqslant C(\log M)^{-\xi}, \tag{6.60}$$

$$\max_{1 \leqslant i \leqslant N} \left| G_{ii}^{[\mathbb{U}]}(z) \right| \leqslant C, \tag{6.61}$$

$$\text{and} \;\; \Lambda_o^{[\mathbb{U}]}(z) \leqslant C(\log M)^{-\xi} \tag{6.62}$$

hold in $\Omega(z)$ with (ξ,ν)-high probability.

Proof. For $\mathbb{T} = \varnothing$, (6.57) and (6.59) are vacuously true. For (6.58), we have $z \in \mathfrak{R}$ and so (2.57) of Lemma 2.21 applies, thus

$$|G_{ii}| \leqslant |G_{ii} - m_{\mathrm{mp}}| + |m_{\mathrm{mp}}| \leqslant \Lambda_d + |m_{\mathrm{mp}}| \leqslant (\log M)^{-\xi} + C \leqslant C,$$

since we are on the event Ω. Similarly, by the reverse triangle inequality,

$$|G_{ii}| \geqslant ||m_{\mathrm{mp}}| - |G_{ii} - m_{\mathrm{mp}}|| \geqslant c - (\log M)^{-\xi} \geqslant c$$

for large enough M.

For $\mathbb{T} \neq \varnothing$, we proceed by inducting for only *finitely many* steps on the cardinality of $\mathbb{T}$. We restrict the induction in this way because then there are only finitely many changes to the constants c and C. Suppose we have the claims (6.57), (6.58), and (6.59) for $\mathbb{T}$ and we wish to show them for $k\mathbb{T}$, where $k \in \{1, \ldots, N\} \backslash \mathbb{T}$ on Ω. From (6.24) of Lemma 6.5, we find

$$\left|G_{ii}^{(k\mathbb{T})} - G_{ii}\right| \leqslant \left|G_{ii}^{(k\mathbb{T})} - G_{ii}^{(\mathbb{T})}\right| + \left|G_{ii}^{(\mathbb{T})} - G_{ii}\right| \leqslant \left|\frac{G_{ik}^{(\mathbb{T})} G_{ki}^{(\mathbb{T})}}{G_{kk}^{(\mathbb{T})}}\right| + C\Lambda_0^2 \leqslant C\Lambda_0^2,$$

where the induction assumption (6.57) in the penultimate step, and the induction assumptions (6.58) and (6.59) in the final step. We can now use the fact that we are on the event Ω and take the maximum over $i \notin \mathbb{T}$ to prove the claim (6.57). Similarly, using (6.24) of Lemma 6.5, we prove

$$\left|G_{ii}^{(k\mathbb{T})}\right| \leqslant \left|G_{ii}^{(\mathbb{T})}\right| + \left|\frac{G_{ki}^{(\mathbb{T})} G_{ik}^{(\mathbb{T})}}{G_{kk}^{(\mathbb{T})}}\right| \leqslant C + C\Lambda_0^2 \leqslant C + C(\log M)^{-2\xi} \leqslant C,$$

since we are on the event Ω. Again using (6.24) of Lemma 6.5, we see with the reverse triangle inequality

$$\left|G_{ii}^{(k\mathbb{T})}\right| \geqslant \left|\left|G_{ii}^{(\mathbb{T})}\right| - \left|\frac{G_{ki}^{(\mathbb{T})} G_{ik}^{(\mathbb{T})}}{G_{kk}^{(\mathbb{T})}}\right|\right| \geqslant c - C(\log M)^{-2\xi} \geqslant c,$$

for large enough M, where we again used the fact we are on the event Ω. Finally, we show (6.59) by noting that

$$\left|G_{ij}^{(k\mathbb{T})}\right| \leqslant \left|G_{ij}^{(\mathbb{T})}\right| + \left|\frac{G_{ik}^{(\mathbb{T})} G_{kj}^{(\mathbb{T})}}{G_{kk}^{(\mathbb{T})}}\right| \leqslant C\Lambda_0 + C\Lambda_0^2 \leqslant C\Lambda_0,$$

since we are on the event Ω, where we used the induction assumptions (6.58) and (6.59) and the fact that Λ_o dominates Λ_o^2 on Ω. Lastly, taking the max over $i \neq j$ and $i, j \notin k\mathbb{T}$ yields (6.59).

The argument for (6.60), (6.61), and (6.62) is similar to above but more complicated. We use induction on the cardinality of $\mathbb{T}$ and point out the base case for (6.60) is trivial and that the base case for (6.62) follows immediately from the assumption that we are on Ω. For the base case of (6.61), we see

$$|\underline{G}_{ii}| \leqslant \left|\underline{G}_{ii} - \left(\gamma m_{\mathrm{mp}} + \frac{\gamma - 1}{z}\right)\right| + \gamma\,|m_{\mathrm{mp}}| + \frac{|\gamma - 1|}{|z|} \leqslant (\log M)^{-\xi} + C \leqslant C,$$

since we are on Ω. Now assume that we have (6.60), (6.61), and (6.62) for $\mathbb{T}$ and we wish to show them for $k\mathbb{T}$, where $k \in \{1, \ldots, M\} \backslash \mathbb{T}$.

Now, we use the rank-one perturbation formula (6.25) of Lemma 6.5 and take the ij entry to see

$$\begin{aligned}
\underline{G}_{ij}^{(\mathbb{T})} - \underline{G}_{ij}^{(k\mathbb{T})} &= zG_{ii}^{(\mathbb{T})} \left[\underline{G}^{(k\mathbb{T})} \mathbf{x}_k \mathbf{x}_k^* \underline{G}^{(k\mathbb{T})}\right]_{ij} \\
&= zG_{ii}^{(\mathbb{T})} \sum_{l,r=1}^{M} x_{lk} \underline{G}_{il}^{(k\mathbb{T})} \underline{G}_{rj}^{(k\mathbb{T})} x_{rk} \\
&= zG_{ii}^{(\mathbb{T})} \left(\sum_{l \neq r}^{M} x_{lk} \underline{G}_{il}^{(k\mathbb{T})} \underline{G}_{rj}^{(k\mathbb{T})} x_{rk} + \sum_{l=1}^{M} \left(x_{lk}^2 - \frac{1}{M}\right) \underline{G}_{il}^{(k\mathbb{T})} \underline{G}_{lj}^{(k\mathbb{T})} + \frac{1}{M} \sum_{l=1}^{M} \underline{G}_{il}^{(k\mathbb{T})} \underline{G}_{lj}^{(k\mathbb{T})}\right).
\end{aligned} \tag{6.63}$$

Note that by (6.58), $G_{ii}^{(\mathbb{T})}$ is order one on Ω, so we can disregard it at the expense of a constant factor, and similarly, z is bounded by a constant in $\mathfrak{R}$. We can control the first term with the large deviation estimate, Lemma 5.1. Thus, with with (ξ, ν)-high probability on Ω

$$\begin{aligned}
\left|\sum_{l\neq r}^{M} x_{lk}\underline{G}_{il}^{(k\mathbb{T})}\underline{G}_{rj}^{(k\mathbb{T})}x_{rk}\right| &\leqslant (\log M)^{2\xi}\left[\frac{\max_{l\neq r}\left|\underline{G}_{il}^{(k\mathbb{T})}\underline{G}_{rj}^{(k\mathbb{T})}\right|}{q} + \left(\frac{1}{M^2}\sum_{l\neq r}\left|\underline{G}_{il}^{(k\mathbb{T})}\underline{G}_{rj}^{(k\mathbb{T})}\right|^2\right)^{1/2}\right] \\
&\leqslant (\log M)^{2\xi}\left[\frac{\max_{l\neq r}\left|\underline{G}_{il}^{(k\mathbb{T})}\underline{G}_{rj}^{(k\mathbb{T})}\right|}{q} + \left(\frac{\operatorname{Im}\underline{G}_{ii}^{(k\mathbb{T})}\operatorname{Im}\underline{G}_{jj}^{(k\mathbb{T})}}{M^2\eta^2}\right)^{1/2}\right] \\
&\leqslant (\log M)^{-\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|^2 + (\log M)^{-6\xi}\left(\left|\underline{G}_{ii}^{(k\mathbb{T})}\right|\left|\underline{G}_{jj}^{(k\mathbb{T})}\right|\right)^{1/2} \\
&\leqslant (\log M)^{-\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|^2 + (\log M)^{-6\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|,
\end{aligned} \tag{6.64}$$

where we used the Ward identity, Lemma 2.14, in the second step twice; and assumption (3.4) on q and the fact that $z \in \mathfrak{R}_L$, so $\eta \geqslant (\log M)^L/M$ in the third step. We also use the large deviation estimate, Lemma 5.1, on the second term of (6.63), to see

$$\begin{aligned}
\left|\sum_{l=1}^{M}\left(x_{lk}^2 - \frac{1}{M}\right)\underline{G}_{il}^{(k\mathbb{T})}\underline{G}_{lj}^{(k\mathbb{T})}\right| &\leqslant (\log M)^{\xi}\frac{\max_l\left|\underline{G}_{il}^{(k\mathbb{T})}\underline{G}_{lj}^{(k\mathbb{T})}\right|}{q} \\
&\leqslant (\log M)^{-2\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|^2
\end{aligned} \tag{6.65}$$

with with (ξ, ν)-high probability, where we used (3.4) again. We see for the third term of (6.63) that

$$\begin{aligned}
\left|\frac{1}{M}\sum_{l=1}^{M}\underline{G}_{il}^{(k\mathbb{T})}\underline{G}_{lj}^{(k\mathbb{T})}\right| &\leqslant \frac{C}{M}\sum_{l=1}^{M}\left(\left|\underline{G}_{il}^{(k\mathbb{T})}\right|^2 + \left|\underline{G}_{lj}^{(k\mathbb{T})}\right|^2\right) \\
&= \frac{C}{M\eta}\left(\operatorname{Im}\underline{G}_{ii}^{(k\mathbb{T})} + \operatorname{Im}\underline{G}_{jj}^{(k\mathbb{T})}\right) \\
&\leqslant C(\log M)^{-8\xi}\max_{l}\left|\underline{G}_{ll}^{(k\mathbb{T})}\right|,
\end{aligned} \tag{6.66}$$

by the definition of $\mathfrak{R}_L$ in (3.30) and the bound on L in (3.29).

From equations (6.63), (6.64), (6.65), and (6.66) we see

$$\begin{aligned}
\left|\underline{G}_{ij}^{(\mathbb{T})} - \underline{G}_{ij}^{(k\mathbb{T})}\right| &\leqslant C(\log M)^{-\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|^2 + C(\log M)^{-6\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right| \\
&\quad + C(\log M)^{-2\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|^2 + C(\log M)^{-8\xi}\max_{l}\left|\underline{G}_{ll}^{(k\mathbb{T})}\right| \\
&\leqslant C(\log M)^{-\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|^2 + C(\log M)^{-6\xi}\max_{l,r}\left|\underline{G}_{lr}^{(k\mathbb{T})}\right|
\end{aligned}$$

with (ξ,ν)-high probability on Ω, where we have isolated the dominant terms. Using the simple inequality $(x+y)^2 \leqslant Cx^2 + Cy^2$, we see,

$$
\begin{aligned}
C(\log M)^{-\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} \right|^2 &\leqslant C(\log M)^{-\xi} \max_{l,r} \left(\left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right| + \left| \underline{G}_{lr}^{(\mathbb{T})} \right| \right)^2 \\
&\leqslant C(\log M)^{-\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right|^2 + C(\log M)^{-\xi} \max_{l,r} \left| \underline{G}_{lr}^{(\mathbb{T})} \right|^2 \\
&\leqslant C(\log M)^{-\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right|^2 + C(\log M)^{-\xi},
\end{aligned}
$$

with (ξ,ν)-high probability on Ω, where we used the induction hypotheses (6.61) and (6.62). Similarly,

$$
\begin{aligned}
C(\log M)^{-6\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} \right| &\leqslant C(\log M)^{-6\xi} \max_{l,r} \left(\left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right| + \left| \underline{G}_{lr}^{(\mathbb{T})} \right| \right) \\
&\leqslant C(\log M)^{-6\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right| + C(\log M)^{-6\xi},
\end{aligned}
$$

by the induction hypotheses (6.61) and (6.62). Putting all this together, we see with (ξ,ν)-high probability on Ω

$$
\left| \underline{G}_{ij}^{(k\mathbb{T})} - \underline{G}_{ij}^{(\mathbb{T})} \right| \leqslant C(\log M)^{-\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right|^2 + C(\log M)^{-6\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right| + C(\log M)^{-\xi},
$$

which immediately implies

$$
\max_{i\neq j} \left| \underline{G}_{ij}^{(k\mathbb{T})} - \underline{G}_{ij}^{(\mathbb{T})} \right| \leqslant C(\log M)^{-\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right|^2 + C(\log M)^{-\xi}
$$

with (ξ,ν)-high probability on Ω, when we take the maximum over all i and j, since $C(\log M)^{-6\xi} = o(1)$. Next we may conclude

$$
\max_{i\neq j} \left| \underline{G}_{ij}^{(k\mathbb{T})} - \underline{G}_{ij}^{(\mathbb{T})} \right| \leqslant C(\log M)^{-\xi} \quad \text{or} \quad \max_{i\neq j} \left| \underline{G}_{ij}^{(k\mathbb{T})} - \underline{G}_{ij}^{(\mathbb{T})} \right| \leqslant C(\log M)^{-\xi} \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right|^2. \tag{6.67}
$$

However, we may rule out the case that

$$
C(\log M)^{\xi} \leqslant \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right| \tag{6.68}
$$

by noting that this is obviously false for η order one by the trivial bound

$$
\max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} - \underline{G}_{lr}^{(\mathbb{T})} \right| \leqslant \max_{l,r} \left| \underline{G}_{lr}^{(k\mathbb{T})} \right| + \max_{l,r} \left| \underline{G}_{lr}^{(\mathbb{T})} \right| \leqslant \frac{2}{\eta}.
$$

Then a simple continuity argument rules out (6.68) for all η. Thus,

$$
\max_{i\neq j} \left| \underline{G}_{ij}^{(k\mathbb{T})} - \underline{G}_{ij}^{(\mathbb{T})} \right| \leqslant C(\log M)^{-\xi}.
$$

So, we see with (ξ,ν)-high probability on Ω

$$
\max_{i,j} \left| \underline{G}_{ij}^{(k\mathbb{T})} - \underline{G}_{ij}^{(\mathbb{T})} \right| \leqslant C(\log M)^{-\xi}. \tag{6.69}
$$

Now we easily find

$$\max_{i,j}\left|\underline{G}_{ij}^{(k\mathbb{T})}-\underline{G}_{ij}\right|\leqslant\max_{i,j}\left|\underline{G}_{ij}^{(k\mathbb{T})}-\underline{G}_{ij}^{(\mathbb{T})}\right|+\max_{i,j}\left|\underline{G}_{ij}^{(\mathbb{T})}-\underline{G}_{ij}\right|\leqslant C(\log M)^{-\xi}$$

with (ξ,ν)-high probability on Ω, by (6.69) and the induction hypothesis (6.60). Similarly, with (ξ,ν)-high probability on Ω

$$\max_{i}\left|\underline{G}_{ii}^{(k\mathbb{T})}\right|\leqslant\max_{i}\left|\underline{G}_{ii}^{(\mathbb{T})}\right|+\max_{i}\left|\underline{G}_{ii}^{(k\mathbb{T})}-\underline{G}_{ii}^{(\mathbb{T})}\right|\leqslant C+C(\log M)^{-\xi}\leqslant C,$$

by (6.69) and the induction hypothesis (6.61), and with (ξ,ν)-high probability on Ω

$$\max_{i\neq j}\left|\underline{G}_{ij}^{(k\mathbb{T})}\right|\leqslant\max_{i\neq j}\left|\underline{G}_{ij}^{(\mathbb{T})}\right|+\max_{i\neq j}\left|\underline{G}_{ij}^{(k\mathbb{T})}-\underline{G}_{ij}^{(\mathbb{T})}\right|\leqslant C(\log M)^{-\xi},$$

by (6.69) and the induction hypothesis (6.62). This completes the whole proof. ■

LEMMA 6.11. *For fixed $z\in\mathfrak{R}_L$, we have with (ξ,ν)-high probability*

$$\Lambda_o(z)\leqslant C(\log M)^{\xi}\frac{1}{q}+C(\log M)^{2\xi}\Psi(z),\tag{6.70}$$

$$\max_i|\Gamma_i(z)|\leqslant C(\log M)^{\xi}\frac{1}{q}+C(\log M)^{2\xi}\Psi(z),\tag{6.71}$$

$$\textit{and}\quad\max_i|G_{ii}(z)-m_N(z)|\leqslant C(\log M)^{\xi}\frac{1}{q}+C(\log M)^{2\xi}\Psi(z)\tag{6.72}$$

on $\Omega(z)$. Similarly, we have with (ξ,ν)-high probability

$$\underline{\Lambda}_o(z)\leqslant C(\log M)^{\xi}\frac{1}{q}+C(\log M)^{2\xi}\Psi(z),\tag{6.73}$$

$$\textit{and}\quad\max_i|\underline{G}_{ii}(z)-\underline{m}_M(z)|\leqslant C(\log M)^{\xi}\frac{1}{q}+C(\log M)^{2\xi}\Psi(z)\tag{6.74}$$

on $\Omega(z)$.

PROOF. First, we prove that (6.70) holds with (ξ,ν)-high probability. Let $i\neq j$, then using the resolvent identity (6.23) from Lemma 6.5, we see

$$\Lambda_o=|z|\max_{i\neq j}\left|G_{ii}G_{jj}^{(i)}\mathbf{x}_i^*\underline{G}^{(ij)}\mathbf{x}_j\right|\leqslant C|z|\max_{i\neq j}\left|\mathbf{x}_i^*\underline{G}^{(ij)}\mathbf{x}_j\right|,$$

where we used (6.58) of Lemma 6.10 to bound the diagonal terms of the resolvent and the fact that we are on the event Ω. It is important to note that $\underline{G}^{(ij)}$ is independent of $\mathbf{x}_i$ and $\mathbf{x}_j$. Thus, using (5.6) from the large deviation estimate, Lemma 5.1, we obtain that there is a constant $\nu>0$ such that for any ξ satisfying assumption (3.1), the inequality

$$\left|\mathbf{x}_i^*\underline{G}^{(ij)}\mathbf{x}_j\right|=\left|\sum_{k,l=1}^{M}x_{ki}\underline{G}_{kl}^{(ij)}x_{lj}\right|\leqslant(\log M)^{2\xi}\left[\frac{\max_k\left|\underline{G}_{kk}^{(ij)}\right|}{q^2}+\frac{\underline{\Lambda}_o^{(ij)}}{q}+\left(\frac{1}{M^2}\sum_{k\neq l}\left|\underline{G}_{kl}^{(ij)}\right|^2\right)^{1/2}\right]\tag{6.75}$$

holds with (ξ,ν)-high probability. Note that on Ω we have $\max_k \left|\underline{G}_{kk}^{(ij)}\right| \leqslant C$ and $\underline{\Lambda}_o^{(ij)} \leqslant C(\log M)^{-\xi}$ by Lemma 6.10. Thus, we can rewrite (6.75) as

$$\left|\mathbf{x}_i^* \underline{G}^{(ij)} \mathbf{x}_j\right| \leqslant C(\log M)^{2\xi} \left[\frac{1}{q^2} + \frac{(\log M)^{-\xi}}{q} + \left(\frac{1}{M^2} \sum_{k \neq l} \left|\underline{G}_{kl}^{(ij)}\right|^2 \right)^{1/2} \right]$$

on Ω with (ξ,ν)-high probability. Using the Ward identity in Lemma 2.14, we see that

$$\frac{1}{M^2} \sum_{k \neq l} \left|\underline{G}_{kl}^{(ij)}\right|^2 \leqslant \frac{1}{M^2 \eta} \operatorname{Im} \operatorname{Tr} \underline{G}^{(ij)}.$$

Now we use (6.27) from Lemma 6.5, to state the bound in terms $G^{(ij)}$:

$$\frac{1}{\eta} \operatorname{Im} \operatorname{Tr} \underline{G}^{(ij)} = \frac{1}{\eta} \operatorname{Im} \operatorname{Tr} G^{(ij)} + \frac{1}{\eta} \operatorname{Im} \frac{N - |\mathbb{T}| - M}{z} = \frac{1}{\eta} \operatorname{Im} \operatorname{Tr} G^{(ij)} + \frac{N - |\mathbb{T}| - M}{|z|^2}. \tag{6.76}$$

Moreover,

$$\frac{1}{M} \operatorname{Tr} G^{(ij)} = C \left(\frac{1}{N} \sum_{k=1}^N G_{kk}^{(ij)} - G_{kk} \right) + C \left(\frac{1}{N} \sum_{k=1}^N G_{kk} - m_{\mathrm{mp}} \right) + C m_{\mathrm{mp}},$$

so on Ω

$$\frac{1}{N} \operatorname{Im} \operatorname{Tr} G^{(ij)} \leqslant C\Lambda_o^2 + C\Lambda + C \operatorname{Im} m_{\mathrm{mp}}, \tag{6.77}$$

by the definition of Λ and the bound (6.57) in Lemma 6.10. Furthermore,

$$\left| \frac{N - |\mathbb{T}| - M}{M^2} \right| \leqslant \frac{C}{M}, \tag{6.78}$$

where C depends on γ. Using the observations in (6.77) and (6.78), we obtain

$$\frac{1}{M^2} \sum_{k \neq l} \left|\underline{G}_{kl}^{(ij)}\right|^2 \leqslant \frac{C \operatorname{Im} m_{\mathrm{mp}} + C\Lambda}{M\eta} + \frac{C\Lambda_o^2}{M\eta} + \frac{C}{|z|^2 M}.$$

from (6.76). However, since $|z| \leqslant C$ on $\mathfrak{R}$, we have

$$\frac{|z|^2}{M^2} \sum_{k \neq l} \left|\underline{G}_{kl}^{(ij)}\right|^2 \leqslant \frac{C \operatorname{Im} m_{\mathrm{mp}} + C\Lambda}{M\eta} + \frac{C\Lambda_o^2}{M\eta} + \frac{C}{M}. \tag{6.79}$$

Now, taking the maximum over $i \neq j$ in equation (6.79), using $|z| \leqslant C$ on $\mathfrak{R}$ again, and absorbing the constants into the front constant, gives us

$$\Lambda_o \leqslant +C(\log M)^{\xi} \frac{1}{q} + C(\log M)^{2\xi} \left[\frac{1}{q^2} + \sqrt{\frac{\operatorname{Im} m_{\mathrm{mp}} + \Lambda}{M\eta} + \frac{\Lambda_o^2}{M\eta} + \frac{1}{M}} \right]$$

on Ω with (ξ,ν)-high probability. However, by the trivial inequality $\sqrt{x+y} \leqslant \sqrt{x} + \sqrt{y}$ for positive x and y (which is easily proved by squaring), we see

$$\Lambda_o \leqslant \frac{C(\log M)^{2\xi}}{\sqrt{M\eta}} \Lambda_o + C(\log M)^{\xi} \frac{1}{q} + C(\log M)^{2\xi} \left[\frac{1}{q^2} + \sqrt{\frac{\operatorname{Im} m_{\mathrm{mp}} + \Lambda}{M\eta} + \frac{1}{M}} \right]$$

on Ω with (ξ,ν)-high probability. Note that

$$\frac{C(\log M)^{2\xi}}{\sqrt{M\eta}} = o(1),$$

in $\mathfrak{R}_L$ by assumption (3.29) on L. Moreover, we can use equation (2.61) to see that the term $\operatorname{Im} m_{\mathrm{mp}}/(M\eta)$ dominates the term $1/M$. Thus, with a possible increase in the constant C, we see

$$\Lambda_o \leqslant o(1)\Lambda_o + C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\left[\frac{1}{q^2} + \sqrt{\frac{\operatorname{Im} m_{\mathrm{mp}} + \Lambda}{M\eta}}\right]$$

on Ω with (ξ,ν)-high probability. Therefore, collecting the Λ_o terms and dividing by their coefficient, we see on Ω with (ξ,ν)-high probability that

$$\Lambda_o \leqslant C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\left[\frac{1}{q^2} + \Psi\right],$$

for large enough M by the definition of Ψ. Finally, we may remove the term $1/q^2$, as it is dominated by the $1/q$ term, to get

$$\Lambda_o \leqslant C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\Psi,$$

This completes the proof of (6.70).

The proof of (6.71) is similar and we omit the level of detail we showed for (6.70). Recall the definition of Γ_i in (6.46). Expanding Γ_i and separating the diagonal and off-diagonal terms, we see

$$\Gamma_i = z\underline{m}_M^{(i)} - z\mathbf{x}_i^*\underline{G}^{(i)}\mathbf{x}_i = \frac{z}{M}\sum_{k=1}^{M}\underline{G}_{kk}^{(i)} - z\sum_{k,l=1}^{M}x_{ki}\underline{G}_{kl}^{(i)}x_{li} = z\sum_{k=1}^{M}\underline{G}_{kk}^{(i)}\left(\frac{1}{M} - x_{ki}^2\right) - z\sum_{k\neq l}x_{ki}\underline{G}_{kl}^{(i)}x_{li}. \quad (6.80)$$

Thus, we can bound (6.80) by using the large deviation estimate, Lemma 5.1, on each of the terms on the right-hand side. In more detail, we use (5.3) on the first term, to see that there is a $\nu > 0$ such that for any ξ satisfying (3.1), such that

$$\left|\sum_{k=1}^{M}\underline{G}_{kk}^{(i)}\left(\frac{1}{M} - x_{ki}^2\right)\right| \leqslant (\log M)^{\xi}\frac{\max_k\left|\underline{G}_{kk}^{(i)}\right|}{q}$$

with (ξ,ν)-high probability. Since $\max_k\left|\underline{G}_{kk}^{(i)}\right| \leqslant C$, by Lemma 6.10, we may write

$$\left|\sum_{k=1}^{M}\underline{G}_{kk}^{(i)}\left(\frac{1}{M} - x_{ki}^2\right)\right| \leqslant C(\log M)^{\xi}\frac{1}{q} \quad (6.81)$$

with (ξ,ν)-high probability on Ω. For the second term we use (5.4) to see,

$$\left|\sum_{k\neq l}x_{ki}\underline{G}_{kl}^{(i)}x_{li}\right| \leqslant C\,(\log M)^{2\xi}\left[\frac{C(\log M)^{-\xi}}{q} + \left(\frac{1}{M^2}\sum_{k\neq l}\left|\underline{G}_{kl}^{(i)}\right|^2\right)^{1/2}\right]$$

with (ξ,ν)-high probability on Ω, since $\underline{\Lambda}_o^{(i)} \leqslant C(\log M)^{-\xi}$ by Lemma 6.10. We deal with the term

$$\frac{1}{M^2}\sum_{k\neq l}\left|\underline{G}_{kl}^{(i)}\right|^2$$

in the same way as before and use the fact that on Ω we have $\Lambda_o \leqslant (\log M)^{-\xi}$, to find

$$\left|\sum_{k\neq l} x_{ki}\underline{G}_{kl}^{(i)}x_{li}\right| \leqslant C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\frac{\Lambda_0}{\sqrt{M\eta}} + C(\log M)^{2\xi}\Psi \tag{6.82}$$

with (ξ,ν)-high probability on Ω. Using the fact $|z|\leqslant C$ on $\mathfrak{R}$, we can conclude from equations (6.81) and (6.82) that

$$|\Gamma_i| \leqslant C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\frac{\Lambda_0}{\sqrt{M\eta}} + C(\log M)^{2\xi}\Psi \tag{6.83}$$

with (ξ,ν)-high probability on Ω. Finally, invoking (6.70) to bound Λ_o and taking the maximum over i, we get

$$\max_i |\Gamma_i| \leqslant C\left[\frac{(\log M)^{\xi}}{q} + (\log M)^{2\xi}\Psi\right].$$

This completes the proof of (6.71).

Finally, we turn to the proof of (6.72). Notice that

$$\max_i |G_{ii}-m_N| = \max_i \left|\frac{1}{N}\sum_{j=1}^N G_{ii}-G_{jj}\right| \leqslant \max_i \frac{1}{N}\sum_{i=1}^N |G_{ii}-G_{jj}| \leqslant \max_{i\neq j}|G_{ii}-G_{jj}|\,.$$

Recall the self-consistent resolvent equation (6.44) and the bound (6.58) of Lemma 6.10, then calculating we see

$$|G_{ii}-G_{jj}| = |G_{ii}|\,|G_{jj}|\,|\Delta_j-\Delta_i| \leqslant C\max_i|\Delta_i|$$

on Ω. It remains to see $\max_i|\Delta_i|$ is small with (ξ,ν)-high probability. Recall the definition of Δ_i in (6.45), then we see

$$z\underline{m}_M^{(i)} = z\gamma_N m_N^{(i)} + \gamma_N - 1 - \frac{|\mathbb{T}|}{M}$$

by equation (6.27). Hence

$$|\Delta_i| \leqslant |\gamma_N-\gamma| + |z|\left|\gamma_N m_N^{(i)} - \gamma m_N\right| + |\Gamma_i|\,.$$

The term $|\gamma_N-\gamma|\leqslant C/M$ by assumption (3.3) on γ_N. We can again use the bound $|z|\leqslant C$ on $\mathfrak{R}$ and then note that

$$\left|\gamma_N m_N^{(i)} - \gamma m_N\right| \leqslant |\gamma_N-\gamma|\left|m_N^{(i)}\right| + \gamma\left|m_N^{(i)}-m_N\right| \leqslant \frac{C}{M} + \frac{\gamma}{N}\sum_k^{(i)}\left|G_{kk}^{(i)}-G_{kk}\right| + \frac{\gamma}{N}|G_{ii}|$$

on Ω, by assumption (3.3) on γ_N and (6.58) from Lemma 6.10. Now, using (6.24) from Lemma 6.5, we see

$$\left|\gamma_N m_N^{(i)} - \gamma m_N\right| \leqslant \frac{C}{M} + C\Lambda_o(z)^2 \tag{6.84}$$

on Ω, by equations (6.57) and (6.58) from Lemma 6.10 again. We can then use (6.70) to bound (6.84). Moreover,

$$|\Gamma_i| \leqslant C\left[\frac{(\log M)^{\xi}}{q} + (\log M)^{2\xi}\Psi\right] \tag{6.85}$$

with (ξ, ν)-high probability on Ω, by (6.71). Thus, by combining the bounds (6.84) and (6.85), we get

$$|\Delta_i| \leqslant C\left[\frac{(\log M)^{\xi}}{q} + (\log M)^{2\xi}\Psi\right]$$

with (ξ, ν)-high probability on Ω.

One can show (6.73) and (6.74) in the same way as above. ■

Note that immediately from (6.72), we see

$$\Lambda_d(z) \leqslant \Lambda(z) + C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\Psi(z) \tag{6.86}$$

on $\Omega(z)$ with (ξ, ν)-high probability. Similarly, from (6.74), we see

$$\underline{\Lambda}_d(z) \leqslant \Lambda(z) + C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\Psi(z) \tag{6.87}$$

on $\Omega(z)$ with (ξ, ν)-high probability, by equation (6.27).

6.3. Stability of the self-consistent equation on $\boldsymbol{\Omega(z)}$. We now expand the self-consistent equation. Fix $z := E + \mathrm{i}\eta \in \mathfrak{R}_L$. It is easy to show that if one has a bound on the perturbation term Δ_i in the self-consistent equation (6.44), that is

$$\max_i |\Delta_i| \leqslant \delta(z) \tag{6.88}$$

for some function δ which is decreasing in η, then with a Taylor expansion we have the following bound

$$\left| m_N - \frac{1}{1-\gamma-z\gamma m_N - z}\right| \leqslant C\delta. \tag{6.89}$$

Equation (6.89) shows that if the perturbation term is small, then m_N approximately satisfies the self-content equation (2.49) with a small error term. The conclusion of Lemma 6.12 is that, in fact, we have the bound

$$\Lambda = |m_N - m_{\mathrm{mp}}| \leqslant C\frac{\delta}{\sqrt{\delta + \eta + \kappa}}, \tag{6.90}$$

which is typically stronger.

Lemma 6.12. *For $z := E + \mathrm{i}\eta \in \mathfrak{R}_L$, define*

$$\delta \equiv \delta(z) := \max_i |\Delta_i(z)| \tag{6.91}$$

and assume $\delta(z)$ is decreasing in η and that on $\Omega(z)$, we have $\delta(z) \leqslant (\log M)^{-2\xi}$ with (ξ, ν)-high probability. Then on $\Omega(z)$ with (ξ, ν)-high probability

$$\Lambda(z) := |m_N(z) - m_{mp}(z)| \leqslant C\frac{\delta(z)}{\sqrt{\delta(z) + \eta + \kappa(z)}}, \tag{6.92}$$

where

$$\kappa \equiv \kappa(z) := \min\{|\lambda_+ - E|, |E - \lambda_-|\}. \tag{6.93}$$

Proof. Recall that for $z \in \mathfrak{R}_L$, we have

$$|1-\gamma-z-z\gamma m_{\mathrm{mp}}| \geqslant c > 0$$

from equation (2.59) of Lemma 2.21. Thus, using the reverse triangle inequality, we get

$$|1-\gamma-z-z\gamma m_N| \geqslant \big||1-\gamma-z-z\gamma m_{\mathrm{mp}}| - C|z|\,|m_{\mathrm{mp}}-m_N|\big| \geqslant c, \tag{6.94}$$

since we are on the event Ω and $|z| \leqslant C$ in $\mathfrak{R}$, so $|m_{\mathrm{mp}}-m_N| = \Lambda \leqslant \Lambda_d \leqslant (\log M)^{-\xi}$. Then expanding

$$m_N = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1-\gamma-z\gamma m_N - z - \Delta_i} = \frac{1}{1-\gamma-z\gamma m_N - z} \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{1}{1-\frac{\Delta_i}{1-\gamma-z\gamma m_N-z}}$$

to the first order, we see

$$\left|m_N - \frac{1}{1-\gamma-z\gamma m_N - z}\right| \leqslant \mathcal{O}\left(\frac{\max_i |\Delta_i|}{|1-\gamma-z\gamma m_N - z|}\right) \leqslant C\delta(z), \tag{6.95}$$

by the bound (6.94).

For some fixed $\Delta \equiv \Delta(z) \in \mathbb{R}$, the equation

$$m(z) - \frac{1}{1-\gamma-z\gamma m(z)-z} = \Delta(z) \tag{6.96}$$

has two distinct solutions which we denote by $m^{\Delta}_{\pm}(z)$. Equation (6.96) can be written equivalently as

$$z\gamma m(z)^2 + (\gamma - 1 + z - z\gamma\Delta)m(z) + 1 + \Delta - \gamma\Delta - z\Delta = 0,$$

so explicit calculation using the quadratic formula shows

$$m^{\Delta}_{\pm}(z) = \frac{1-\gamma-z \pm \mathrm{i}(1+\gamma\Delta)\sqrt{\left(\lambda^{\Delta}_{+}-z\right)\left(z-\lambda^{\Delta}_{-}\right)}}{2\gamma z} + \frac{\Delta}{2} \tag{6.97}$$

where

$$\lambda^{\Delta}_{\pm} := \left(\frac{\sqrt{1+\Delta(\gamma-\gamma^2)} \pm \sqrt{\gamma}}{1+\Delta\gamma}\right)^2. \tag{6.98}$$

This follows by noting

$$(1-\gamma-z+\gamma z\Delta)^2 - 4\gamma z(1+(1-\gamma-z)\Delta) = -(1+\Delta\gamma)^2(\lambda^{\Delta}_{+}-z)(z-\lambda^{\Delta}_{-}).$$

We use the notation $m_{\pm}(z) \equiv m^{0}_{\pm}(z)$ and point out that $m_{+}(z) \equiv m_{\mathrm{mp}}(z)$, as equation (6.97) simplifies to equation (2.48) when $\Delta = 0$. Currently, we have the following conclusion: there exists $\Delta \leqslant C\delta$, such that

$$m_N = m^{\Delta}_{+} \ \text{ or } \ m_N = m^{\Delta}_{-}. \tag{6.99}$$

We take an intermission from the proof of Lemma 6.12 to prove a short lemma about the solutions to equation (6.96), after which we resume the proof. For a fixed energy $\mathbf{1}_{\gamma>1}(\lambda_-/5) \leqslant E \leqslant 5\lambda_+$, we define $\tilde\eta$ as the solution of

$$\delta(E+\mathrm{i}\tilde\eta) = (\log M)^{-\xi}(\kappa+\tilde\eta), \tag{6.100}$$

which is unique as the left-hand side is decreasing in $\tilde{\eta}$ by assumption and the right-hand side is obviously increasing in $\tilde{\eta}$. Note that $\tilde{\eta} \ll 1$ for large enough M. So for any $\eta \geqslant \tilde{\eta}$, we have

$$\delta(z) \leqslant (\log M)^{-\xi}(\kappa+\eta) \leqslant \kappa+\eta, \tag{6.101}$$

and conversely, when $\eta \leqslant \tilde{\eta}$, we have

$$\delta(z) \geqslant (\log M)^{-\xi}(\kappa+\eta). \tag{6.102}$$

Lemma 6.13. *Let $m^{\Delta}_{\pm}(z)$ be as defined in* (6.97) *and let $z := E+\mathrm{i}\eta \in \mathfrak{R}_L$. For sufficiently small Δ, depending on γ, we have*

$$\max_{\pm} \left| m^{\Delta}_{\pm}(z) - m_{\pm}(z) \right| \leqslant C \frac{\Delta}{\sqrt{\eta+\kappa(z)}}. \tag{6.103}$$

Moreover, if $\eta \geqslant \tilde{\eta}$, then

$$\left| m^{\Delta}_{+}(z) - m^{\Delta}_{-}(z) \right| \geqslant c\sqrt{\kappa+\eta} \gg (\log M)^{-\xi} \tag{6.104}$$

and if $\eta \leqslant \tilde{\eta}$, then

$$\left| m^{\Delta}_{+}(z) - m^{\Delta}_{-}(z) \right| \leqslant C(\log M)^{\xi}\sqrt{\delta}. \tag{6.105}$$

Proof of Lemma 6.13. Using the difference of two squares, for Δ small enough, we have

$$\begin{aligned}
\tilde{\Delta} &:= \max_{\pm} \left| \lambda^{\Delta}_{\pm} - \lambda_{\pm} \right| \\
&\leqslant \left| \frac{\sqrt{1+\Delta(\gamma-\gamma^2)} \pm \sqrt{\gamma}}{1+\Delta\gamma} + \frac{1 \pm \sqrt{\gamma} + \Delta(\gamma \pm \gamma\sqrt{\gamma})}{1+\Delta\gamma} \right| \left| \frac{\sqrt{1+\Delta(\gamma-\gamma^2)} \pm \sqrt{\gamma}}{1+\Delta\gamma} - \frac{1 \pm \sqrt{\gamma} + \Delta(\gamma \pm \gamma\sqrt{\gamma})}{1+\Delta\gamma} \right| \\
&\leqslant C \left| \frac{\sqrt{1+\Delta(\gamma-\gamma^2)} - 1 - \Delta(\gamma \pm \gamma\sqrt{\gamma})}{1+\Delta\gamma} \right| \\
&\leqslant C \left| \sqrt{1+\Delta(\gamma-\gamma^2)} - 1 \right| + C\Delta \\
&\leqslant C\Delta,
\end{aligned} \tag{6.106}$$

where we used a Taylor expansion on the term $\sqrt{1+\Delta(\gamma-\gamma^2)}$. Similarly, we find

$$\min_{\pm} \left| \lambda^{\Delta}_{\pm} - \lambda_{\pm} \right| \geqslant c\Delta. \tag{6.107}$$

Therefore, by (6.106), we have

$$\begin{aligned}
\max_{\pm} \left| m^{\Delta}_{\pm} - m_{\pm} \right| &\leqslant \frac{\Delta}{2} + \frac{\left| \sqrt{\left(\lambda^{\Delta}_{+} - z\right)\left(z - \lambda^{\Delta}_{-}\right)} - \sqrt{\left(\lambda_{+} - z\right)\left(z - \lambda_{-}\right)} \right|}{2\gamma|z|} + \frac{\Delta \left| \sqrt{\left(\lambda^{\Delta}_{+} - z\right)\left(z - \lambda^{\Delta}_{-}\right)} \right|}{2|z|} \\
&\leqslant C\Delta + C\Delta\sqrt{\kappa} + C \left| \sqrt{\left(\lambda^{\Delta}_{+} - z\right)\left(z - \lambda^{\Delta}_{-}\right)} - \sqrt{\left(\lambda_{+} - z\right)\left(z - \lambda_{-}\right)} \right|.
\end{aligned}$$

Now, we use the inequality,

$$\left| \sqrt{x+y} - \sqrt{x} \right| \leqslant C \frac{|y|}{\sqrt{|x|+|y|}},$$

which holds for all complex numbers x and y, where we set the quantities $x := (\lambda_+ - z)(z - \lambda_-)$ and $y := (\lambda_+^\Delta - z)(z - \lambda_-^\Delta) - (\lambda_+ - z)(z - \lambda_-)$, to see

$$\begin{aligned}
&\left|\sqrt{\left(\lambda_+^\Delta - z\right)\left(z - \lambda_-^\Delta\right)} - \sqrt{\left(\lambda_+ - z\right)\left(z - \lambda_-\right)}\right| \\
&\qquad \leqslant C \frac{\left|(\lambda_+^\Delta - z)(z - \lambda_-^\Delta) - (\lambda_+ - z)(z - \lambda_-)\right|}{\sqrt{\left|(\lambda_+ - z)(z - \lambda_-)\right| + \left|(\lambda_+^\Delta - z)(z - \lambda_-^\Delta) - (\lambda_+ - z)(z - \lambda_-)\right|}} \\
&\qquad \leqslant C \frac{\Delta}{\sqrt{\kappa + \eta}},
\end{aligned}$$

where the last line follows from the following bounds: $|y| \leqslant C\tilde{\Delta}$ and thus by (6.106), $|y| \leqslant C\Delta$, $|x| \geqslant c\kappa$, and $|y| \geqslant c\eta$.

Now we show (6.104). Note, for $\eta \geqslant \tilde{\eta}$,

$$\left|m_+^\Delta - m_-^\Delta\right| \geqslant c \left|\left(\lambda_+^\Delta - z\right)\left(z - \lambda_-^\Delta\right)\right|^{1/2} \geqslant c\sqrt{\kappa + \eta}.$$

Finally, we show (6.105). Note, for $\eta \leqslant \tilde{\eta}$,

$$\left|m_+^\Delta - m_-^\Delta\right| \leqslant C \left|\left(\lambda_+^\Delta - z\right)\left(z - \lambda_-^\Delta\right)\right|^{1/2} \leqslant C\sqrt{\Delta} \leqslant C(\log M)^\xi \sqrt{\delta(z)},$$

which completes the proof. ∎

Using Lemma 6.13 and equation (6.99), we may conclude that there is some $\Delta \leqslant C\delta$ such that

$$\min\{|m_N - m_{\mathrm{mp}}|, |m_N - m_-|\} \leqslant C \frac{\Delta}{\sqrt{\eta + \kappa}} \leqslant C \frac{\delta}{\sqrt{\eta + \kappa}}. \tag{6.108}$$

Now we must distinguish between the two solutions. If we could see that $m_N = m_+^\Delta$, by arguing that only m_+^Δ (and not m_-^Δ) is a function from $\mathbb{H}$ to $\mathbb{H}$, then we could conclude the proof using Lemma 6.13. However, it is too difficult to see this from the expression (6.97). Despite this difficultly, we can identify m_+^Δ as the correct solution for $\eta \sim 1$, using a different argument. In more detail, let $z_0 := E + \mathrm{i}\eta_0 \in \mathfrak{R}_L$ such that $\eta_0 \geqslant 1$. Using Lemma 6.13, we have the following simple bound

$$\left|m_+^\Delta(z_0) - m_-^\Delta(z_0)\right| \geqslant c\sqrt{\kappa + \eta_0} \geqslant c, \tag{6.109}$$

so for $\eta \sim 1$ we can easily distinguish between the solutions. Then

$$|m_N(z_0) - m_{\mathrm{mp}}(z_0)| = \Lambda(z_0) \leqslant \Lambda_d(z_0) \leqslant (\log M)^{-\xi} \tag{6.110}$$

since we are on the event Ω. Moreover, from Lemma 6.13 and our assumption on δ, we have

$$\left|m_+^\Delta(z_0) - m_+(z_0)\right| \leqslant C \frac{\delta}{\sqrt{\kappa + \eta_0}} \leqslant C(\log M)^{-2\xi}, \tag{6.111}$$

since $\eta_0 \sim 1$. Now note,

$$\left|m_+^\Delta(z_0) - m_N(z_0)\right| \leqslant \left|m_+^\Delta(z_0) - m_+(z_0)\right| + |m_+(z_0) - m_N(z_0)| \leqslant C(\log M)^{-\xi} \tag{6.112}$$

by equations (6.109) and (6.111). So, using the reverse triangle inequality and equations (6.109) and (6.112)

$$\left|m_-^\Delta(z_0) - m_N(z_0)\right| \geqslant \left|\left|m_-^\Delta(z_0) - m_+^\Delta(z_0)\right| - \left|m_+^\Delta(z_0) - m_N(z_0)\right|\right| \geqslant c.$$

Thus, $m_N(z_0) \neq m_-^\Delta(z_0)$ and we must have $m_N(z_0) = m_+^\Delta(z_0)$. Therefore,

$$|m_N(z_0) - m_{\mathrm{mp}}(z_0)| = \left|m_+^\Delta(z_0) - m_+(z_0)\right| \leqslant C\frac{\delta}{\sqrt{\eta_0 + \kappa}} \leqslant C\frac{\delta}{\sqrt{\delta + \eta_0 + \kappa}} \tag{6.113}$$

by Lemma 6.13 and the bound (6.101).

Now, using the continuity of the functions m_N, m_+^Δ, and m_-^Δ in η, we may conclude that $m_N = m_+^\Delta$ for all $\eta \geqslant \tilde{\eta}$ by equation (6.104) of Lemma 6.13. Thus, using (6.103) of Lemma 6.13, we have

$$|m_N(z) - m_{\mathrm{mp}}(z)| = \left|m_+^\Delta(z) - m_+(z)\right| \leqslant C\frac{\delta}{\sqrt{\eta + \kappa}} \leqslant C\frac{\delta}{\sqrt{\delta + \eta + \kappa}}.$$

by equation (6.102) again.

Now, assume $\eta \leqslant \tilde{\eta}$. We show that even if m_N is close to m_-, it is also close to m_+. First assume $m_N = m_+^\Delta$, then, trivially, we see

$$|m_N - m_{\mathrm{mp}}| = \left|m_+^\Delta - m_+\right| \leqslant C\frac{\delta}{\sqrt{\kappa + \eta}}$$

Now assume $m_N = m_-^\Delta$. Thus by the triangle inequality and equation (6.105) of Lemma 6.13, we have

$$\begin{aligned}
|m_N - m_{\mathrm{mp}}| &= \left|m_-^\Delta - m_+\right| \\
&\leqslant \left|m_-^\Delta - m_+^\Delta\right| + \left|m_+^\Delta - m_+\right| \\
&\leqslant C(\log M)^\xi \sqrt{\delta} + C\frac{\delta}{\sqrt{\kappa + \eta}} \\
&\leqslant C\frac{\delta}{\sqrt{\delta + \eta + \kappa}},
\end{aligned}$$

where the last bound is an interpolation of the two pervious terms. ■

6.4. Initial estimates for large η. In order to start the induction in the continuity argument of Subsection 6.5, we need an initial estimate of $\Lambda_d + \Lambda_o$ for large η—that is, we need to prove that $\Omega(E + \mathrm{i}\eta)$ is an event of (ξ, ν)-high probability for $\eta \sim 1$, as from this it follows that the estimates in Lemmas 6.10 and 6.11 all hold with (ξ, ν)-high probability for $\eta \sim 1$.

Lemma 6.14. *Let $\eta := 10(1 + \gamma)$ and specify E so that $z := E + \mathrm{i}\eta \in \mathfrak{R}_L$. Then we have*

$$\Lambda_o(z) + \Lambda_d(z) + \underline{\Lambda}_o(z) + \underline{\Lambda}_d(z) \leqslant C(\log M)^{-\xi} \tag{6.114}$$

with (ξ, ν)-high probability. In other words, $\Omega(z)$ holds with (ξ, ν)-high probability.

Proof. Fix $z := E + \mathrm{i}\eta \in \mathfrak{R}_L$ where $\eta := 10(1 + \gamma)$. We repeatedly make use of the following trivial estimates:

$$\left|G_{ij}^{(\mathbb{T})}\right| \leqslant \frac{1}{\eta},\ \left|\underline{G}_{ij}^{(\mathbb{T})}\right| \leqslant \frac{1}{\eta},\ \left|m_N^{(\mathbb{T})}\right| \leqslant \frac{1}{\eta},\ \left|\underline{m}_M^{(\mathbb{T})}\right| \leqslant \frac{1}{\eta},\ \text{and}\ |m_{\mathrm{mp}}| \leqslant \frac{1}{\eta}, \tag{6.115}$$

where $\mathbb{T} \subseteq \{1, \ldots, N\}$ is arbitrary. When $\eta \sim 1$, (6.115) implies that all these quantities are bounded by constants. These estimates are fairly immediate: the operator norms of G and $\underline{G}$ are bounded by $1/\eta$, since the spectra of H and $\underline{H}$ are real and z is at least η from the real line. Thus,

$$|G_{ij}| = |\mathbf{e}_i^* G \mathbf{e}_j| \leqslant \|\mathbf{e}_i^*\| \|G\| \|\mathbf{e}_j\| \leqslant \frac{1}{\eta}.$$

We can prove exactly the same for $G_{ij}^{(\mathbb{T})}$ and $\underline{G}_{ij}^{(\mathbb{T})}$. The bounds on $m_N^{(\mathbb{T})}$ and $\underline{m}_M^{(\mathbb{T})}$ follow directly from the bounds on the resolvent entires, as they are just averages of the diagonal terms. We demonstrated the bound on m_{mp} in Section 2.

First, we estimate the quantity Λ_o. Let $i \neq j$. Then following the calculation bounding Λ_o in the proof of Lemma 6.11 and using the large deviation estimate, Lemma 5.1, we see that

$$\left|\mathbf{x}_i^* \underline{G}^{(ij)} \mathbf{x}_j\right| = \left|\sum_{k,l=1}^{M} x_{ki} \underline{G}_{kl}^{(ij)} x_{lj}\right| \leqslant (\log M)^{2\xi} \left[\frac{\max_k \left|\underline{G}_{kk}^{(ij)}\right|}{q^2} + \frac{\underline{\Lambda}_o^{(ij)}}{q} + \left(\frac{1}{M^2}\sum_{k\neq l}\left|\underline{G}_{kl}^{(ij)}\right|^2\right)^{1/2}\right] \tag{6.116}$$

with (ξ, ν)-high probability. Note that $\underline{\Lambda}_o^{(ij)} \leqslant C$ and $\max_k |G_{kk}| \leqslant C$ by (6.115). Using (6.115) and the Ward identity from Lemma 2.14, as before, we get

$$\Lambda_o \leqslant C(\log M)^{2\xi}\left(\frac{1}{q^2} + \frac{1}{q} + \sqrt{\frac{\operatorname{Im}\operatorname{Tr}\underline{G}^{(ij)}}{M\eta}}\right)$$

with (ξ, ν)-high probability by taking the max over $i \neq j$, since $z \leqslant C$ on $\mathfrak{R}_L$. Moverover, the term $1/q$ dominates the term $1/q^2$. Finally, with (ξ, ν)-high probability

$$\Lambda_o \leqslant C(\log M)^{2\xi}\left(\frac{1}{q} + \sqrt{\frac{\operatorname{Im}\operatorname{Tr}\underline{G}^{(ij)}}{M\eta}}\right) \leqslant C(\log M)^{-\xi} \tag{6.117}$$

by the bound (3.4) on q.

Now, we need to estimate Λ_d. First, we estimate the error term $|\Delta_i|$ as in the proof of Lemma 6.11, to see

$$|\Delta_i| \leqslant |\gamma_N - \gamma| + |z|\left|\gamma_N m_N^{(i)} - \gamma m_N\right| + |\Gamma_i|$$

and using the large deviation estimate Lemma 5.1, we see recalling (6.80)

$$|\Gamma_i| \leqslant (\log M)^{\xi} \frac{\max_k \left|\underline{G}_{kk}^{(i)}\right|}{q} + C(\log M)^{2\xi}\left[\frac{\Lambda_o^{(i)}}{q} + \left(\frac{1}{M^2}\sum_{k\neq l}\left|\underline{G}_{kl}^{(i)}\right|^2\right)^{1/2}\right] \tag{6.118}$$

with (ξ, ν)-high probability. Simplifying (6.118) in the same way as before, we see

$$|\Gamma_i| \leqslant C(\log M)^{\xi}\frac{1}{q} + C(\log M)^{2\xi}\left[\frac{1}{q} + \frac{1}{\sqrt{M}}\right] \leqslant C(\log M)^{-\xi}$$

with (ξ, ν)-high probability. The term $|\gamma_N - \gamma| \leqslant C/M$ by assumption (3.3) on γ_N. We can again use the bound $|z| \leqslant C$ on $\mathfrak{R}$ and then note that

$$\left|\gamma_N m_N^{(i)} - \gamma m_N\right| \leqslant |\gamma_N - \gamma|\left|m_N^{(j)}\right| + \gamma\left|m_N^{(i)} - m_N\right| \leqslant \frac{C}{M} + \frac{\gamma}{N}\sum_k^{(i)}\left|G_{kk}^{(i)} - G_{kk}\right| + \frac{\gamma}{N}|G_{ii}| \leqslant \frac{C}{M}$$

by assumption (3.3) on γ_N and (6.115). Therefore, with (ξ, ν)-high probability and $\eta \geqslant 10(1+\gamma)$, we see

$$\max_i |\Delta_i| \leqslant C(\log M)^{-\xi}. \tag{6.119}$$

Finally, from the self-consistent resolvent equation (6.44) and the Marchenko-Pastur self-consistent equation (2.49), we find

$$G_{ii} - m_{\mathrm{mp}} = \frac{\Delta_i}{G_{ii}^{-1} m_{\mathrm{mp}}^{-1}}.$$

Now, $m_{\mathrm{mp}} \leqslant 1/(10(1+\gamma))$, so $m_{\mathrm{mp}}^{-1} \geqslant (10(1+\gamma)) \geqslant 10$, and similarly $G_{ii}^{-1} \geqslant 10$

$$\left| \frac{1}{G_{ii}^{-1} m_{\mathrm{mp}}^{-1}} \right| \geqslant 100.$$

Thus, from (6.119) we get

$$|G_{ii} - m_{\mathrm{mp}}| \leqslant C(\log M)^{-\xi},$$

which, when we take the max over i, yields

$$\Lambda_d \leqslant C(\log M)^{-\xi}. \tag{6.120}$$

Now, we can reproduce almost exactly the same argument for $\underline{\Lambda}_o$ and $\underline{\Lambda}_d$. We only outline the proof because the method is almost identical. Basically, one has to remove rows from X rather than columns to produce $X^{[\mathbb{U}]}$. This gives us the minors $\underline{H}^{[\mathbb{U}]}$ of $\underline{H}$ and thus the minors $\underline{G}^{[\mathbb{U}]}$ of $\underline{G}$, and so we can use the resolvent identities from Remark 6.6. Since for η order 1 we only invoke trivial bounds on these minors, this approach actually helps here; we avoid the problem of having to control two types of resolvent simultaneously. Once one has these resolvent identities, it is easy to see how to use the large deviation estimate, Lemma 5.1, and the trivial bounds

$$\left| \underline{G}_{ij}^{[\mathbb{U}]} \right| \leqslant \frac{1}{\eta} \quad \text{and} \quad \left| G_{ij}^{[\mathbb{U}]} \right| \leqslant \frac{1}{\eta}$$

to obtain exactly the same bounds on $\underline{\Lambda}_o$ and $\underline{\Lambda}_d$ as Λ_o and Λ_d.

Finally, summing equations (6.117) and (6.120) and the corresponding bounds on $\underline{\Lambda}_o$ and $\underline{\Lambda}_d$ finishes the proof of (6.114). ■

6.5. Continuity argument: conclusion of the proof of Theorem 6.2. To complete the proof of Theorem 6.2, we use a continuity argument, or what is sometimes called bootstrapping, in η to go from large $\eta = 10(1+\gamma)$ down to the smallest scale of $\eta = M^{-1}(\log M)^L$. To begin with we prove (6.13), so we use Lemma 6.14 for the initial estimate.

Choose any decreasing finite sequence $(\eta_k)_{k=1}^K$ such that $K \leqslant CN^8$, $|\eta_k - \eta_{k+1}| \leqslant N^{-8}$, $\eta_1 := 10(\gamma+1)$, and $\eta_K := M^{-1}(\log M)^L$. For now we fix E such that $\mathbf{1}_{\gamma>1}(\lambda_-/5) \leqslant E \leqslant 5\lambda_+$ and define $z_k := E + \mathrm{i}\eta_k$.

For the initial point z_1, Lemma 6.14 implies that $\Omega(z_1)$ holds with (ξ, ν)-high probability, since the imaginary part of z_1 is order one, $\eta_1 = 10(1+\gamma)$. Thus, using (6.71) from Lemma 6.11, we have with (ξ, ν)-high probability the bound

$$\max_i |\Gamma_i(z_1)| \leqslant C(\log M)^{\xi} \frac{1}{q} + C(\log M)^{2\xi} \Psi(z_1) \leqslant C(\log M)^{\xi} \frac{1}{q} + C(\log M)^{2\xi} \frac{1}{(M\eta)^{1/2}},$$

by (6.50). Since $\max_i |\Gamma_i(z_1)|$ is the dominant term in the perturbation term $\max_i |\Delta_i|$, (which one can see from the calculation at the end of the proof of Lemma 6.11) we can use Lemma 6.12 to find

$$\Lambda(z_1) = |m_N - m_{\mathrm{mp}}| \leqslant \frac{C(\log M)^{\xi} \frac{1}{q} + C(\log M)^{2\xi} \frac{1}{(M\eta)^{1/2}}}{\sqrt{\kappa + \eta + C(\log M)^{\xi} \frac{1}{q} + C(\log M)^{2\xi} \frac{1}{(M\eta)^{1/2}}}} \leqslant \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}}.$$

This estimate proves the claim for the initial point z_1; the next lemma extends the claim to all η_k for $k \leqslant K$.

Lemma 6.15. *Define the events*

$$\Omega_k := \Omega(z_k) \cap \left\{ \Lambda(z_k) \leqslant \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}} \right\}. \tag{6.121}$$

Then

$$\mathbb{P}\left(\Omega_k^c\right) \leqslant 2k e^{-\nu(\log M)^{\xi}}. \tag{6.122}$$

Proof. We use induction on k. We just demonstrated the base case $k = 1$. So, we take as our induction hypothesis that Ω_k holds with (ξ, ν)-high probability. Now we prove the same for $k + 1$. Note that

$$\Omega_{k+1}^c = \left(\Omega_{k+1}^c \cap \Omega_k\right) \cup \left(\Omega_{k+1}^c \cap \Omega_k^c\right)$$

by de Morgan's law. Similarly,

$$\Omega_{k+1}^c = \left(\Omega_{k+1}^c \cap \Omega_k \cap \Omega(z_{k+1})\right) \cup \left(\Omega_{k+1}^c \cap \Omega_k \cap \Omega(z_{k+1})^c\right) \cup \left(\Omega_{k+1}^c \cap \Omega_k^c\right),$$

so that

$$\mathbb{P}\left(\Omega_{k+1}^c\right) \leqslant \mathbb{P}\left(\Omega_{k+1}^c \cap \Omega_k \cap \Omega(z_{k+1})\right) + \mathbb{P}\left(\Omega_k \cap \Omega(z_{k+1})^c\right) + \mathbb{P}\left(\Omega_k^c\right).$$

By the induction hypothesis, we have

$$\mathbb{P}\left(\Omega_k^c\right) \leqslant 2k e^{-\nu(\log M)^{\xi}}.$$

For convenience, define

$$A := \mathbb{P}\left(\Omega_k \cap \Omega(z_{k+1})^c\right) \quad \text{and} \quad B := \mathbb{P}\left(\Omega_k \cap \Omega(z_{k+1}) \cap \Omega_{k+1}^c\right). \tag{6.123}$$

First, we show that A is small. For any i and j using the Lipschitz continuity of the resolvent, we have

$$|G_{ij}(z_k) - G_{ij}(z_{k+1})| \leqslant |z_{k+1} - z_k| \sup_{z \in \mathfrak{R}_L} \left| \frac{\partial G_{ij}(z)}{\partial z} \right| \leqslant N^{-8} \sup_{z \in \mathfrak{R}_L} \frac{1}{(\operatorname{Im} z)^2} \leqslant N^{-6}, \tag{6.124}$$

and similarly,

$$|\underline{G}_{ij}(z_k) - \underline{G}_{ij}(z_{k+1})| \leqslant N^{-6}. \tag{6.125}$$

Therefore, using (6.124), we see

$$\Lambda_d(z_{k+1}) + \Lambda_o(z_{k+1}) \leqslant \Lambda_d(z_k) + \Lambda_o(z_k) + 2N^{-6}.$$

Now we may use the much stronger bounds, (6.70) and (6.86), as we are on the event Ω_k and hence $\Omega(z_k)$. So, with (ξ, ν)-high probability

$$\Lambda_d(z_k) + \Lambda_o(z_k) + 2N^{-6} \leqslant \Lambda(z_k) + C\left[\frac{(\log M)^{\xi}}{q} + (\log M)^{2\xi}\Psi(z_k)\right].$$

We can simplify this bound further, to see on Ω_k with (ξ,ν)-high probability

$$\Lambda_d(z_{k+1}) + \Lambda_o(z_{k+1}) \leqslant \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}} \ll C(\log M)^{-\xi},$$

by the induction hypothesis, (3.4), and (6.50). One can show in almost exactly the same way that

$$\underline{\Lambda}_d(z_{k+1}) + \underline{\Lambda}_o(z_{k+1}) \ll C(\log M)^{-\xi}$$

with (ξ,ν)-high probability. This proves that

$$A \leqslant e^{-\nu(\log M)^{\xi}}.$$

We can bound B similarly using Lemma 6.12, to find for all k

$$\mathbb{P}\left(\Omega_{k+1}^c\right) \leqslant 2e^{-\nu(\log M)^{\xi}} + \mathbb{P}(\Omega_k^c) \leqslant 2(k+1)e^{-\nu(\log M)^{\xi}} \tag{6.126}$$

from which the claim follows. ■

To complete the proof of Theorem 6.2, we use a lattice argument which strengthens the result of Lemma 6.15 to a statement uniform in $z \in \mathfrak{R}_L$. Again, the result relies on the Lipschitz continuity of the map $z \mapsto G_{ij}(z)$, with a Lipschitz constant bounded by $\eta^{-2} \leqslant N^2$.

Corollary 6.16. *There is a constant C such that*

$$\mathbb{P}\left(\bigcup_{z\in\mathfrak{R}_L} \Omega(z)^c\right) + \mathbb{P}\left(\bigcup_{z\in\mathfrak{R}_L} \left\{\Lambda(z) > \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}}\right\}\right) \leqslant e^{-\nu(\log M)^{\xi}}. \tag{6.127}$$

Proof. Let $\mathfrak{L} \subseteq \mathfrak{R}_L$ be a lattice such that $|\mathfrak{L}| \leqslant CN^6$ and for any $z \in \mathfrak{R}_L$ there is a $\tilde{z} \in \mathfrak{L}$ such that $|z - \tilde{z}| \leqslant N^{-3}$. Using the Lipschitz continuity of G, we see

$$|G_{ij}(z) - G_{ij}(\tilde{z})| \leqslant |z - \tilde{z}| \sup_{z\in\mathfrak{R}_L} \left|\frac{\partial G_{ij}(z)}{\partial z}\right| \leqslant N^{-3} \sup_{z\in\mathfrak{R}_L} \frac{1}{(\operatorname{Im} z)^2} \leqslant N^{-1}. \tag{6.128}$$

The same bound is true of $|m(z) - m(\tilde{z})|$. So, using Lemma 6.15, we get

$$\mathbb{P}\left(\bigcap_{z\in\mathfrak{L}} \left\{\Lambda(z) \leqslant \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}}\right\}\right) \geqslant 1 - e^{-\nu(\log M)^{\xi}},$$

for some C large enough and then from equation (6.128), we see

$$\mathbb{P}\left(\bigcup_{z\in\mathfrak{R}_L} \left\{\Lambda(z) > \frac{C(\log M)^{\xi}}{\sqrt{q}} + \frac{C(\log M)^{2\xi}}{(M\eta)^{1/4}}\right\}\right) \leqslant e^{-\nu(\log M)^{\xi}}. \tag{6.129}$$

The other term can be dealt with in the same way. ■

This proves (6.13). To demonstrate (6.10), one can use (6.13), (6.70), and Corollary 6.16. Similarly, one uses (6.13), (6.86), and Corollary 6.16 to show (6.11), and (6.13), (6.72), and Corollary 6.16 to show (6.12).

7. A Brownian motion model for covariance matrices

In this section we present a formal calculation to characterize the diffusion process undergone by the eigenvalues of a covariance matrix $H := X^*TX$ when the entries of X are undergoing independent Brownian motions and T is a nonnegative deterministic $M \times M$ diagonal matrix. The derivation is an exercise in Itō calculus. We fix N throughout this section and do not notate the t-dependence of all quantities during the proof.

The original result of this form is due to Dyson and is commonly called Dyson's Brownian motion—that is, the diffusion process induced on the eigenvalues, denoted by $(\lambda_\alpha(t))_{\alpha=1}^N$ of a symmetric (or Hermitian) $N \times N$ matrix when its entries are independent real (or complex) Brownian motions. The concept was introduced in 1962 by Freeman Dyson in his seminal paper [5]. Dyson's interest in the model was mostly physical and he thought of this as a new type of Coulomb gas (now referred to as log gases): that is, "N point charges executing Brownian motions under the influence of their mutual electrostatic repulsions." The original result of Dyson's paper, presented in modern terms, is that the joint distribution of the eigenvalues of a symmetric (or Hermitian) matrix when its entries are independent real (or complex) Brownian motions is a unique solution to a system of coupled stochastic differential equations,

$$\mathrm{d}\lambda_\alpha(t) = \sqrt{\frac{2}{\beta N}}\mathrm{d}b_\alpha(t) + \frac{1}{N}\left(\sum_\beta^{(\alpha)} \frac{1}{\lambda_\alpha(t) - \lambda_\beta(t)}\right)\mathrm{d}t, \tag{7.1}$$

with initial conditions to specify the initial positions of the eigenvalues, where $(b_\alpha)_{\alpha=1}^N$ are independent standard Brownian motions and $\beta = 1$ (or $\beta = 2$) in the real (or complex) case.. What is significant about the system of coupled stochastic differential equations (7.1) is that the joint distribution of the eigenvalues does not depend on their corresponding eigenvectors—the eigenvectors and eigenvalues decouple.

Here we investigate the corresponding phenomenon for covariance matrices. The results in this section are both negative and positive. The positive result, which was originally proved in [3], is a formal calculation to show that when, for some positive constant c, $T = c\mathbb{1}$ and thus $H = cX^*X$ the processes of the eigenvectors and eigenvalues decouple. In particular, we can interpret Theorem 7.2 as characterizing the processes of the eigenvalues as N independent squared Bessel processes which interact through the last drift term in equation (7.3). Thus, the processes stay positive and they repel each other. The repulsion term is more complicated and interesting than the corresponding term for Dyson's Brownian motion: the pairwise repulsion is not only inversely proportional to the difference of the eigenvalues, but also proportional to the sum of the eigenvalues. For example, when two eigenvalues are close to zero the repulsion can be weak even when the eigenvalues are close to each other. All of these phenomena can be seen easily in the simulations provided by Figure 4.

The negative result states that when T is not a multiple of $\mathbb{1}$, the processes of the eigenvectors and eigenvalues do not decouple. In more detail, the stochastic part of the diffusion process for the eigenvectors depends nontrivially on the components of the eigenvectors at time t. So during the proof, we perform the calculation for general T as defined above but it is only be successful in the case $T = c\mathbb{1}$.

Definition 7.1. *We consider $N \times N$ matrices of the form $H(t) = X(t)^*TX(t)$, where $X = (x_{ij}(t))$ is an $M \times N$ matrix whose entries are real and independent standard Brownian motions and $T := \operatorname{diag}(t_1, \dots t_M)$ whose entries are deterministic nonnegative real numbers. Furthermore, we denote the eigenvalues of H and their corresponding normalized eigenvectors as*

$$(\lambda_\alpha)_{\alpha=1}^N \quad \textit{and} \quad (\mathbf{v}_\alpha)_{\alpha=1}^N \tag{7.2}$$

respectively.

Theorem 7.2. *Let H satisfy Definition 7.1. If $T = c\mathbb{1}$, for some positive constant c, then at each time t, $H(t)$ has N distinct eigenvalues almost surely and the N-dimensional process $(\lambda_1, \ldots, \lambda_N)$ satisfies the coupled system of stochastic differential equations*

$$\mathrm{d}\lambda_\alpha = 2\sqrt{c}\sqrt{\lambda_\alpha(t)}\,\mathrm{d}b_\alpha(t) + cM\mathrm{d}t + \sum_\beta^{(\alpha)} \frac{\lambda_\alpha(t) + \lambda_\beta(t)}{\lambda_\alpha(t) - \lambda_\beta(t)}\,\mathrm{d}t \tag{7.3}$$

for $1 \leqslant \alpha \leqslant N$, where $b_1, \ldots, b_N$ are independent Brownian motions. Note that the process does not depend on the eigenvectors of $H(t)$. Conversely, if T is not a multiple of $\mathbb{1}$, then the N-dimensional process $(\lambda_1, \ldots, \lambda_N)$ does not decouple from the eigenvectors of $H(t)$.

Proof sketch. By definition, we have

$$H\mathbf{v}_\alpha = \lambda_\alpha \mathbf{v}_\alpha \tag{7.4}$$

and

$$\mathbf{v}_\beta^* \mathbf{v}_\alpha = \delta_{\alpha\beta}. \tag{7.5}$$

We want to find out how the eigenvalues evolve when the entries of X are undergoing Brownian motion. We use the notation

$$\dot{f} = \frac{\partial f}{\partial x_{ij}}$$

rather than the usual time derivative, so differentiating equation (7.4) with respect to x_{ij} we get:

$$\dot{\lambda}_\alpha \mathbf{v}_\alpha + \lambda_\alpha \dot{\mathbf{v}}_\alpha = \dot{H}\mathbf{v}_\alpha + H\dot{\mathbf{v}}_\alpha. \tag{7.6}$$

Then differentiating equation (7.5) we get:

$$\dot{\mathbf{v}}_\alpha^* \mathbf{v}_\beta + \mathbf{v}_\alpha^* \dot{\mathbf{v}}_\beta = 0 \text{ and } \dot{\mathbf{v}}_\alpha^* \mathbf{v}_\alpha = 0. \tag{7.7}$$

Multiplying equation (7.6) on the left-hand side by $\mathbf{v}_\alpha^*$, yields

$$\dot{\lambda}_\alpha \mathbf{v}_\alpha^* \mathbf{v}_\alpha + \lambda_\alpha \mathbf{v}_\alpha^* \dot{\mathbf{v}}_\alpha = \mathbf{v}_\alpha^* \dot{H} \mathbf{v}_\alpha + \mathbf{v}_\alpha^* H \dot{\mathbf{v}}_\alpha$$

or equivalently,

$$\dot{\lambda}_\alpha = \mathbf{v}_\alpha^* \dot{H} \mathbf{v}_\alpha, \tag{7.8}$$

since $\lambda_\alpha \mathbf{v}_\alpha^* \dot{\mathbf{v}}_\alpha = \mathbf{v}_\alpha^* H \dot{\mathbf{v}}_\alpha$. Similarly, multiplying equation (7.6) on the left-hand side by $\mathbf{v}_\beta^*$, yields

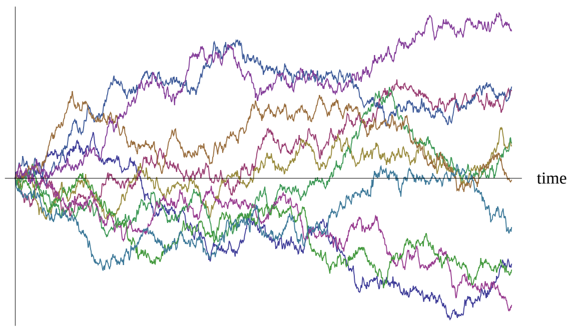
$$\dot{\lambda}_\alpha \mathbf{v}_\beta^* \mathbf{v}_\alpha + \lambda_\alpha \mathbf{v}_\beta^* \dot{\mathbf{v}}_\alpha = \mathbf{v}_\beta^* \dot{H} \mathbf{v}_\alpha + \mathbf{v}_\beta^* H \dot{\mathbf{v}}_\alpha,$$

which implies

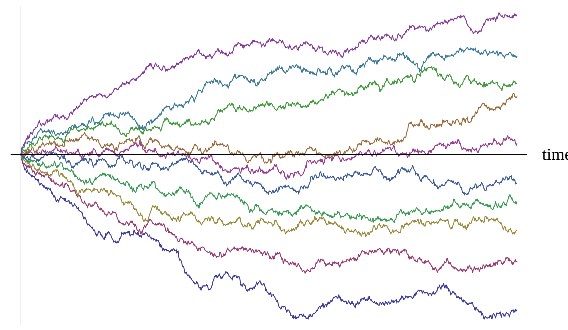
$$\lambda_\alpha \mathbf{v}_\beta^* \dot{\mathbf{v}}_\alpha = \mathbf{v}_\beta^* \dot{H} \mathbf{v}_\alpha + \lambda_\beta \mathbf{v}_\beta^* \dot{\mathbf{v}}_\alpha,$$

or equivalently

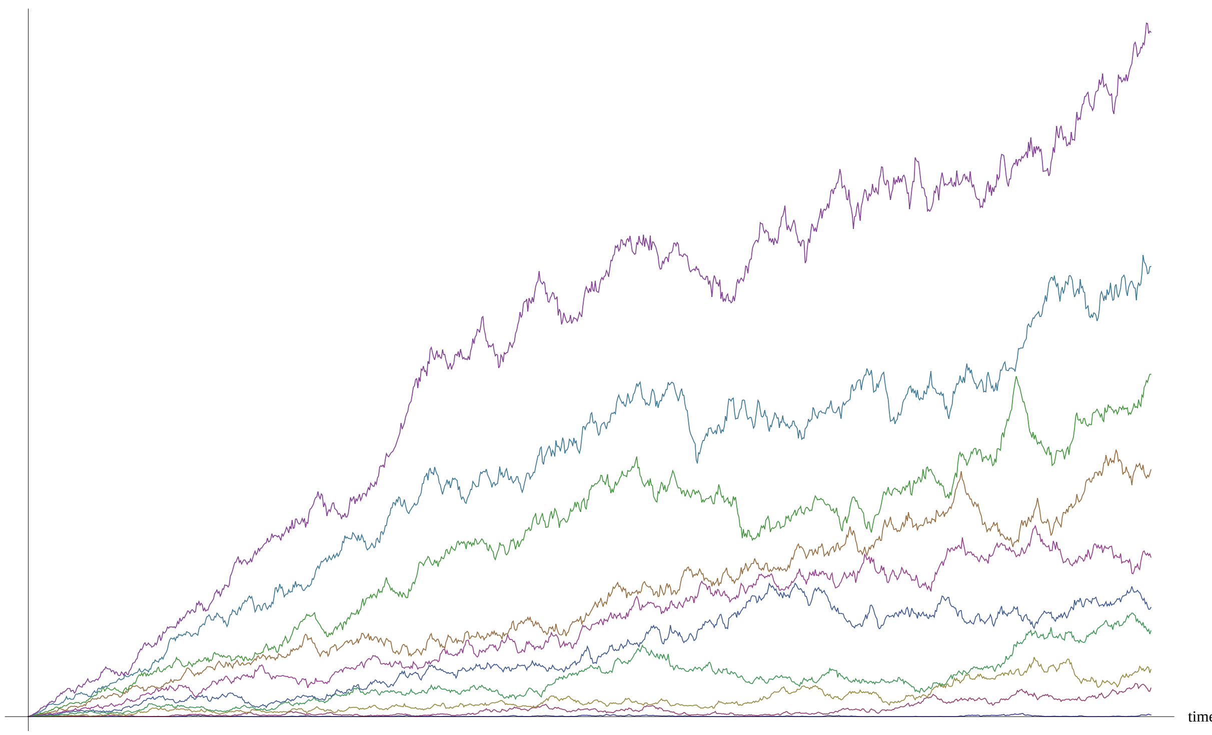
$$\mathbf{v}_\beta^* \dot{\mathbf{v}}_\alpha = \frac{\mathbf{v}_\beta^* \dot{H} \mathbf{v}_\alpha}{\lambda_\alpha - \lambda_\beta}, \tag{7.9}$$



(a) 10 independent Brownian motions started at 0.



(b) Dyson's Brownian motion for a 10×10 matrix.



(c) A simulation of the eigenvalues of H for $N = 10$, where the entries of X are standard Brownian motions started at the origin.

Figure 4: In all simulations $T = \mathbb{1}$. In Figure 4a the processes intersect frequently and do not interact, whereas in Figure 4b the processes do not intersect due to their mutual repulsion. As in Figure 4b, the processes in Figure 4c repel each other but the repulsion is stronger for the larger eigenvalues.

where the second step is by equation (7.5) and $\mathbf{v}_\beta^* H = \lambda_\beta \mathbf{v}_\beta^*$, and the last step is a simple rearrangement as

$\lambda_\beta \neq \lambda_\alpha$ almost surely. Next we can use an eigendecomposition of $\dot{\mathbf{v}}_\alpha$ and equation (7.9) to see

$$\dot{\mathbf{v}}_\alpha = \sum_\beta (\mathbf{v}_\beta \cdot \dot{\mathbf{v}}_\alpha)\mathbf{v}_\beta = \sum_\beta^{(\alpha)} (\mathbf{v}_\beta \cdot \dot{\mathbf{v}}_\alpha)\mathbf{v}_\beta = \sum_\beta^{(\alpha)} \frac{\mathbf{v}_\beta^* \dot{H} \mathbf{v}_\alpha}{\lambda_\alpha - \lambda_\beta}\mathbf{v}_\beta, \tag{7.10}$$

since the α term in the sum, $\mathbf{v}_\alpha \cdot \dot{\mathbf{v}}_\alpha = \dot{\mathbf{v}}_\alpha^* \mathbf{v}_\alpha$ is zero by equation (7.7).

Now we calculate $\dot{H}$ explicitly by looking at its (k,l)th entry:

$$\begin{aligned}
\left[\dot{H}\right]_{kl} &= \left[\frac{\partial X^*}{\partial x_{ij}} T X\right]_{kl} + \left[X^* T \frac{\partial X}{\partial x_{ij}}\right]_{kl} \\
&= \left[\sum_{r=1}^{M} \left(\frac{\partial X^*}{\partial x_{ij}}\right)_{kr} t_r x_{rl}\right]_{kl} + \left[\sum_{r=1}^{M} x_{rk} t_r \left(\frac{\partial X}{\partial x_{ij}}\right)_{rl}\right]_{kl} \\
&= \left[\sum_{r=1}^{M} (\delta_{ir}\delta_{jk})\, t_r x_{rl}\right]_{kl} + \left[\sum_{r=1}^{M} x_{rk} t_r \,(\delta_{ir}\delta_{jl})\right]_{kl} \\
&= t_i \left[x_{il}\delta_{jk} + x_{ik}\delta_{jl}\right]_{kl}.
\end{aligned} \tag{7.11}$$

Thus, equations (7.8) and (7.11) imply

$$\begin{aligned}
\dot{\lambda}_\alpha &= t_i \sum_{r=1}^{N} \sum_{k=1}^{N} \mathbf{v}_\alpha(r) \left[x_{ir}\delta_{jk} + x_{ik}\delta_{jr}\right]_{kr} \mathbf{v}_\alpha(k) \\
&= t_i \sum_{r=1}^{N} \sum_{k=1}^{N} \mathbf{v}_\alpha(r) x_{ir} \delta_{jk} \mathbf{v}_\alpha(k) + \sum_{r=1}^{N} \sum_{k=1}^{N} \mathbf{v}_\alpha(r) x_{ik} \delta_{jr} \mathbf{v}_\alpha(k) \\
&= t_i \sum_{r=1}^{N} \mathbf{v}_\alpha(r) x_{ir} \mathbf{v}_\alpha(j) + \sum_{k=1}^{N} \mathbf{v}_\alpha(j) x_{ik} \mathbf{v}_\alpha(k) \\
&= 2t_i \sum_{r=1}^{N} x_{ir} \mathbf{v}_\alpha(r) \mathbf{v}_\alpha(j).
\end{aligned} \tag{7.12}$$

Similarly, equation (7.10) and (7.11) imply

$$\dot{\mathbf{v}}_\alpha = t_i \sum_\beta^{(\alpha)} \frac{\sum_{r=1}^N x_{ir}[\mathbf{v}_\alpha(r)\mathbf{v}_\beta(j) + \mathbf{v}_\alpha(j)\mathbf{v}_\beta(r)]}{\lambda_\alpha - \lambda_\beta}\mathbf{v}_\beta$$

and thus

$$\dot{\mathbf{v}}_\alpha(k) = t_i \sum_\beta^{(\alpha)} \frac{\sum_{r=1}^N x_{ir}[\mathbf{v}_\alpha(r)\mathbf{v}_\beta(j) + \mathbf{v}_\alpha(j)\mathbf{v}_\beta(r)]}{\lambda_\alpha - \lambda_\beta}\mathbf{v}_\beta(k). \tag{7.13}$$

Now, we want to get the second order partial derivatives of the eigenvalues, but we will not need all of them, only those where we take the derivative with respect to the same entry twice. That is, differentiating equation (7.12) with respect to x_{ij} we get

$$\begin{aligned}
\ddot{\lambda}_\alpha &:= \frac{\partial^2 \lambda_\alpha}{\partial x_{ij}^2} \\
&= 2t_i \sum_{r=1}^{N} \frac{\partial x_{ir}}{\partial x_{ij}} \mathbf{v}_\alpha(r)\mathbf{v}_\alpha(j) + x_{ir}\frac{\partial \mathbf{v}_\alpha(r)}{\partial x_{ij}}\mathbf{v}_\alpha(j) + x_{ir}\mathbf{v}_\alpha(r)\frac{\partial \mathbf{v}_\alpha(j)}{\partial x_{ij}} \\
&= 2t_i \mathbf{v}_\alpha(j)^2 + 2t_i \sum_{r=1}^{N} x_{ir}\dot{\mathbf{v}}_\alpha(r)\mathbf{v}_\alpha(j) + x_{ir}\mathbf{v}_\alpha(r)\dot{\mathbf{v}}_\alpha(j).
\end{aligned} \tag{7.14}$$

We can substitute equation (7.13) into equation (7.14) to see

$$\begin{aligned}
\ddot{\lambda}_\alpha = 2t_i \mathbf{v}_\alpha(j)^2 + 2t_i \sum_{\beta}^{(\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} \sum_{r,s=1}^{N} x_{ir}x_{is}\big[&\mathbf{v}_\alpha(j)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)\mathbf{v}_\beta(r) \\
&+ \mathbf{v}_\alpha(j)^2\mathbf{v}_\beta(r)\mathbf{v}_\beta(s) + \mathbf{v}_\alpha(r)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)^2 + \mathbf{v}_\alpha(j)\mathbf{v}_\alpha(r)\mathbf{v}_\beta(j)\mathbf{v}_\beta(s)\big].
\end{aligned} \tag{7.15}$$

The eigenvalues are functions of the entires of X, so we will now take the Itô derivative of $\lambda_\alpha(x_{11}, \ldots, x_{MN})$ using Itô's lemma. Thus,

$$\begin{aligned}
\mathrm{d}\lambda_\alpha &= \sum_{i=1}^{M}\sum_{j=1}^{N} \frac{\partial \lambda_\alpha}{\partial x_{ij}}\mathrm{d}x_{ij} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{N} \frac{\partial^2 \lambda_\alpha}{\partial x_{ij}\partial x_{kl}}(\mathrm{d}x_{ij})(\mathrm{d}x_{kl}) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{N} \frac{\partial \lambda_\alpha}{\partial x_{ij}}\mathrm{d}x_{ij} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N} \frac{\partial^2 \lambda_\alpha}{\partial x_{ij}^2}\mathrm{d}t.
\end{aligned} \tag{7.16}$$

We can substitute equations (7.12) and (7.15) into equation (7.16) to see

$$\begin{aligned}
\mathrm{d}\lambda_\alpha = 2\sum_{i=1}^{M}\sum_{j,r=1}^{N} t_i x_{ir}\mathbf{v}_\alpha(r)\mathbf{v}_\alpha(j)\mathrm{d}x_{ij} &+ \sum_{i=1}^{M}\sum_{j=1}^{N} t_i \mathbf{v}_\alpha(j)^2 \mathrm{d}t \\
&+ \sum_{\beta}^{(\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} \sum_{i=1}^{M}\sum_{j,r,s=1}^{N} t_i x_{ir}x_{is}\big[\mathbf{v}_\alpha(j)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)\mathbf{v}_\beta(r) \\
&\quad + \mathbf{v}_\alpha(j)^2\mathbf{v}_\beta(r)\mathbf{v}_\beta(s) + \mathbf{v}_\alpha(r)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)^2 + \mathbf{v}_\alpha(j)\mathbf{v}_\alpha(r)\mathbf{v}_\beta(j)\mathbf{v}_\beta(s)\big]\mathrm{d}t.
\end{aligned} \tag{7.17}$$

Let's take this one term at a time. First, note that

$$\sum_{r=1}^{N} \sqrt{t_i}\, x_{ir}\mathbf{v}_\alpha(r) = (T^{1/2}X\mathbf{v}_\alpha)(i)$$

where $(T^{1/2}X\mathbf{v}_\alpha)(i)$ is the ith entry of the vector $T^{1/2}X\mathbf{v}_\alpha$. Suppose $\lambda_\alpha \neq 0$ (which is true almost surely), and define

$$\underline{\mathbf{v}}_\alpha := \frac{T^{1/2}X\mathbf{v}_\alpha}{\|T^{1/2}X\mathbf{v}_\alpha\|},$$

where

$$\|T^{1/2}X\mathbf{v}_\alpha\|^2 = \Big\langle T^{1/2}X\mathbf{v}_\alpha\,,\, T^{1/2}X\mathbf{v}_\alpha \Big\rangle = \langle \mathbf{v}_\alpha\,,\, X^*TX\mathbf{v}_\alpha\rangle = \lambda_\alpha,$$

and thus

$$\sum_{r=1}^{N} \sqrt{t_i} x_{ir} \mathbf{v}_\alpha(r) = (T^{1/2} X \mathbf{v}_\alpha)(i) = \sqrt{\lambda_\alpha} \underline{\mathbf{v}}_\alpha(i). \tag{7.18}$$

Note that $\underline{\mathbf{v}}_\alpha$ is a normalized eigenvector of $T^{1/2}XX^*T^{1/2}$ with eigenvalue λ_α, since

$$T^{1/2}XX^*T^{1/2}\underline{\mathbf{v}}_\alpha = \frac{1}{\|T^{1/2}X\mathbf{v}_\alpha\|} T^{1/2}X\left(X^*TX\right)\mathbf{v}_\alpha = \lambda_\alpha \frac{T^{1/2}X\mathbf{v}_\alpha}{\|X^*TX\|} = \lambda_\alpha \underline{\mathbf{v}}_\alpha.$$

Using the above observation in (7.18) we can see that

$$2\sum_{i=1}^{M}\sum_{j,r=1}^{N} t_i x_{ir}\mathbf{v}_\alpha(r)\mathbf{v}_\alpha(j)\mathrm{d}x_{ij} = 2\sqrt{\lambda_\alpha}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)\mathrm{d}x_{ij}.$$

Define $b_\alpha := \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)x_{ij} = \sqrt{\lambda_\alpha}$, so if we can show that b_α is a standard Brownian motion we know then that the stochastic part of the eigenvalue's process is a squared Bessel process, since $b_\alpha^2 = \lambda_\alpha$. Define $\mathbb{E}_t\left(A\right) := \mathbb{E}\left(A \,|\, \mathcal{F}(t)\right)$, where $\mathcal{F}(t)$ is the σ-algebra of events that can be defined in terms of the process up to time t. Thus, we must show $\mathbb{E}_t\left(\mathrm{d}b_\alpha\right) = 0$ and $\mathbb{E}_t\left(\mathrm{d}b_\alpha \mathrm{d}b_\beta\right) = c\delta_{\alpha\beta}$. For any unit vector $\mathbf{v}$, $\mathbf{v}\cdot\mathrm{d}\mathbf{v} = \mathrm{d}\mathbf{v}\cdot\mathbf{v} = 0$, so we find

$$\begin{aligned}
\mathrm{d}b_\alpha &= \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{t_i}\mathrm{d}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)x_{ij} + \sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathrm{d}\mathbf{v}_\alpha(j)x_{ij} + \sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)\,\mathrm{d}x_{ij} \\
&= \sqrt{\lambda_\alpha}\mathrm{d}\underline{\mathbf{v}}_\alpha\cdot\underline{\mathbf{v}}_\alpha + \sqrt{\lambda_\alpha}\underline{\mathbf{v}}_\alpha\cdot\mathrm{d}\underline{\mathbf{v}}_\alpha + \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)\,\mathrm{d}x_{ij} \\
&= \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)\,\mathrm{d}x_{ij}.
\end{aligned} \tag{7.19}$$

So when we take the expectation and use the facts that $\mathbb{E}_t\left(\mathrm{d}x_{ij}\right) = 0$ and that $\underline{\mathbf{v}}_\alpha$ and $\mathbf{v}_\alpha$ are measurable in $\mathcal{F}(t)$, we see

$$\begin{aligned}
\mathbb{E}_t\left(\mathrm{d}b_\alpha\right) &= \mathbb{E}_t\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)\,\mathrm{d}x_{ij} \\
&= \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{t_i}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)\mathbb{E}_t\left(\mathrm{d}x_{ij}\right) \\
&= 0.
\end{aligned} \tag{7.20}$$

Next, we get

$$\begin{aligned}
\mathbb{E}_t\left(\mathrm{d}b_\alpha\mathrm{d}b_\beta\right) &= \mathbb{E}_t\sum_{i,k=1}^{M}\sum_{j,l=1}^{N}\sqrt{t_i t_k}\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\alpha(j)\underline{\mathbf{v}}_\beta(k)\mathbf{v}_\beta(l)(\mathrm{d}x_{ij})(\mathrm{d}x_{kl}) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{N} t_i\underline{\mathbf{v}}_\alpha(i)\underline{\mathbf{v}}_\beta(i)\mathbf{v}_\alpha(j)\mathbf{v}_\beta(j)\,\mathrm{d}t \\
&= \delta_{\alpha\beta}\sum_{i=1}^{M} t_i\underline{\mathbf{v}}_\alpha(i)^2\,\mathrm{d}t.
\end{aligned} \tag{7.21}$$

Note that the quantity $\sum_{i=1}^M t_i \underline{\mathbf{v}}_\alpha(i)^2$ is only deterministic and independent of the components of the eigenvectors if $t_1 = \ldots = t_M = c$. If T does not satisfy this condition, then the stochastic part of the process $(\lambda_1, \ldots, \lambda_N)$ depends in a nontrivial way on the eigenvectors. This completes the proof of the negative claim in Theorem 7.2. However, when $T = c\mathbb{1}$ we get

$$\sum_{i=1}^{M} t_i \underline{\mathbf{v}}_\alpha(i)^2 = c\delta_{\alpha\beta}\, \mathrm{d}t. \tag{7.22}$$

So, in this special case,

$$2\sum_{i=1}^{M}\sum_{j,r=1}^{N} x_{ir}\mathbf{v}_\alpha(r)\mathbf{v}_\alpha(j)\mathrm{d}x_{ij} = 2\sqrt{c}\sqrt{\lambda_\alpha}\, \mathrm{d}b_\alpha. \tag{7.23}$$

We continue to notate our calculations for general T, to demonstrate how the calculations of other terms work out, but from now on we remain in the special case $T = c\mathbb{1}$. For the second term in (7.17), we see

$$\sum_{i=1}^{M}\sum_{j=1}^{N} t_i \mathbf{v}_\alpha(j)^2 \mathrm{d}t = \mathrm{Tr}\,(T)\,\mathrm{d}t = cM\mathrm{d}t. \tag{7.24}$$

For the third term in (7.17), we shall see

$$\begin{aligned}\sum_{i=1}^{M}\sum_{j,r,s=1}^{N} t_i x_{ir}x_{is}\big[&\mathbf{v}_\alpha(j)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)\mathbf{v}_\beta(r) + \mathbf{v}_\alpha(j)^2\mathbf{v}_\beta(r)\mathbf{v}_\beta(s)\\ &+ \mathbf{v}_\alpha(r)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)^2 + \mathbf{v}_\alpha(j)\mathbf{v}_\alpha(r)\mathbf{v}_\beta(j)\mathbf{v}_\beta(s)\big]\mathrm{d}t\\ &= \lambda_\alpha + \lambda_\beta.\end{aligned} \tag{7.25}$$

Take the first summand in equation (7.25) and use equation (7.18) twice, to see

$$\begin{aligned}\sum_{i=1}^{M}\sum_{j,r,s=1}^{N} \sqrt{t_i}x_{ir}\sqrt{t_i}x_{is}\mathbf{v}_\alpha(j)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)\mathbf{v}_\beta(r) &= \sqrt{\lambda_\alpha}\sum_{i=1}^{M}\sum_{j,r=1}^{N}\sqrt{t_i}x_{ir}\mathbf{v}_\alpha(j)\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\beta(j)\mathbf{v}_\beta(r)\\ &= \sqrt{\lambda_\alpha\lambda_\beta}\sum_{i=1}^{M}\sum_{j=1}^{N}\mathbf{v}_\alpha(j)\underline{\mathbf{v}}_\alpha(i)\mathbf{v}_\beta(j)\underline{\mathbf{v}}_\beta(i)\\ &= \lambda_\alpha\delta_{\alpha\beta}.\end{aligned}$$

Now, take the second of the terms in equation (7.25) and perform a similar calculation, to find

$$\begin{aligned}\sum_{i=1}^{M}\sum_{j,r,s=1}^{N} \sqrt{t_i}x_{ir}\sqrt{t_i}x_{is}\mathbf{v}_\alpha(j)^2\mathbf{v}_\beta(r)\mathbf{v}_\beta(s) &= \sqrt{\lambda_\beta}\sum_{i=1}^{M}\sum_{j,r=1}^{N}\sqrt{t_i}x_{ir}\mathbf{v}_\alpha(j)^2\mathbf{v}_\beta(r)\underline{\mathbf{v}}_\beta(i)\\ &= \lambda_\beta\sum_{i=1}^{M}\sum_{j=1}^{N}\mathbf{v}_\alpha(j)^2\underline{\mathbf{v}}_\beta(i)^2\\ &= \lambda_\beta.\end{aligned}$$

The third term is similar to the second and the fourth term is similar to the first. Therefore,

$$\begin{aligned}\sum_{\beta}^{(\alpha)} \frac{1}{\lambda_\alpha - \lambda_\beta} \sum_{i,j,r,s=1}^{n} x_{ir}x_{is}\big[&\mathbf{v}_\alpha(j)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)\mathbf{v}_\beta(r) \\ &+ \mathbf{v}_\alpha(j)^2\mathbf{v}_\beta(r)\mathbf{v}_\beta(s) + \mathbf{v}_\alpha(r)\mathbf{v}_\alpha(s)\mathbf{v}_\beta(j)^2 + \mathbf{v}_\alpha(j)\mathbf{v}_\alpha(r)\mathbf{v}_\beta(j)\mathbf{v}_\beta(s)\big]\mathrm{d}t \\ &= \sum_{\beta}^{(\alpha)} \frac{\lambda_\alpha + \lambda_\beta}{\lambda_\alpha - \lambda_\beta}\mathrm{d}t.\end{aligned} \tag{7.26}$$

So, putting together equations (7.23), (7.24), and (7.26) we find

$$\mathrm{d}\lambda_\alpha = 2\sqrt{c}\sqrt{\lambda_\alpha}\,\mathrm{d}b_\alpha + cM\mathrm{d}t + \sum_{\beta}^{(\alpha)} \frac{\lambda_\alpha + \lambda_\beta}{\lambda_\alpha - \lambda_\beta}\mathrm{d}t. \tag{7.27}$$

Note that this is a formal calculation, not a rigorous proof. We have ignored the issues associated with intersecting eigenvalues, which can be handled but we have omitted doing this here. ■

8. Derivation of the Marchenko-Pastur equation

In this section we consider matrices of the form $H := T^{1/2}X^*XT^{1/2}$, where $X := (x_{ij})$ is an $M \times N$ real matrix whose entries x_{ij} are independent and identically distributed with $\mathbb{E}x_{11} = 0$ and $\mathbb{E}x_{11}^2 = 1/M$; and T is a deterministic positive-definite $N \times N$ matrix, so that $T := U^*\mathrm{diag}(t_1, \ldots, t_N)U$ with U an $N \times N$ unitary matrix, where the $t_i \geqslant 0$, and the empirical distribution of T converges almost surely in distribution to a probability density function μ_T which is compactly supported on $[0, h]$ for some $h \in \mathbb{R}$ as $N \to \infty$. Furthermore, as in Definition 3.1, we assume $M = M_N$ and that $N/M = \gamma_N \to \gamma \in (0, \infty)$ as $N \to \infty$. As before, all these quantities are N (or equivalently M) dependent but we repress this dependence in our notation. We derive the self-consistent equation for the ensemble H by reworking the proof originally given in [19]; the same derivation is also reproduced in [6].

Definition 8.1. *We define the generalized Marchenko-Pastur equation as*

$$m(z) = \int_{\mathbb{R}} \frac{\mu_T(\mathrm{d}t)}{t(1 - \gamma - \gamma z m(z)) - z}. \tag{8.1}$$

The above equation has a unique solution, denoted by $m_{fc}(z)$, when we stipulate that $m_{fc} : \mathbb{H} \to \mathbb{H}$.

Note that, unlike the case $T = \mathbb{1}$, $m_{\mathrm{fc}}(z)$ does not in general have an explicit formula. Also note that when $T = \mathbb{1}$, we recover the regular Marchenko-Pastur equation (2.49), as $\mu_T = \delta_1$. The uniqueness of the equation's solution is proved in Section 5 of [19].

As was explained for the simpler ensemble in Remark 3.6 and Notation 3.7, there are many simple relationships between the ensembles

$$H := T^{1/2}X^*XT^{1/2} \quad \text{and} \quad \underline{H} := XTX^*, \tag{8.2}$$

so in some sense they are equivalent. The proofs are similar to before and we will keep the same broad idea with our notation. We denote the Stieltjes transform of the matrix H as

$$m_N(z) := \frac{1}{N} \operatorname{Tr}(H - z\mathbb{1})^{-1} = \frac{1}{N} \operatorname{Tr} G(z). \tag{8.3}$$

Similarly, we denote the Stieltjes transform of the matrix $\underline{H}$ as

$$\underline{m}_M(z) := \frac{1}{M} \operatorname{Tr}(\underline{H} - z\mathbb{1})^{-1} = \frac{1}{M} \operatorname{Tr} \underline{G}(z). \tag{8.4}$$

These two transforms are related by

$$m_N(z) = (\gamma_N^{-1} - 1)\frac{1}{z} + \gamma_N^{-1}\underline{m}_M(z) \tag{8.5}$$

or equivalently

$$\underline{m}_M(z) = (\gamma_N - 1)\frac{1}{z} + \gamma_N m_N(z), \tag{8.6}$$

since the nonzero eigenvalues of H and $\underline{H}$ agree, as in Lemma 3.5. Now we derive the self-consistent equation for H. Denote the ith row of X by $\mathbf{x}_i$. Note that these rows are independent of one another. Define the quantities

$$\mathbf{r}_i := \mathbf{x}_i T^{1/2} \quad \text{and} \quad H^{(i)} := \sum_j^{(i)} \mathbf{r}_j^* \mathbf{r}_j = H - \mathbf{r}_i^* \mathbf{r}_i, \tag{8.7}$$

since

$$H = T^{1/2}X^*XT^{1/2} = T^{1/2}\left(\sum_{j=1}^{M}\mathbf{x}_j^*\mathbf{x}_j\right)T^{1/2} = \sum_{j=1}^{M}\mathbf{r}_j^*\mathbf{r}_j.$$

Notice that the $\mathbf{r}_i$ are independent of one another (since the $\mathbf{x}_i$ are independent) and that $\mathbf{r}_i$ is independent of $H^{(i)}$ (which is not a minor but behaves very similarly to a minor). We are forced into this approach, rather than the usual analysis of minors, because the non-diagonal nature of T stops the columns of $T^{1/2}X^*XT^{1/2}$ being independent. Note that when T is diagonal we can simply look at columns $XT^{1/2}$, in which case the analysis is similar to usual, but with these definition we can utilize the independence of X even when T is non-diagonal.

During the derivation we will utilize the following simple lemma to express the resolvent of H in terms of a sum of the resolvents of $H^{(i)}$. This result is analogous to the first resolvent identity in Lemma 6.5.

Lemma 8.2. *Let $\mathbf{q} \in \mathbb{C}^N$ be a row and A be an $N \times N$ matrix such that A and $A + \mathbf{q}^*\mathbf{q}$ are invertible, then we have the following equality*

$$\mathbf{q}(A + \mathbf{q}^*\mathbf{q})^{-1} = \frac{1}{1 + \mathbf{q}A^{-1}\mathbf{q}^*}\mathbf{q}A^{-1}. \tag{8.8}$$

Proof. Note

$$(1 + \mathbf{q}A^{-1}\mathbf{q}^*)\mathbf{q} = \mathbf{q} + \mathbf{q}A^{-1}\mathbf{q}^*\mathbf{q} = \mathbf{q}A^{-1}(A + \mathbf{q}^*\mathbf{q}),$$

then divide by the scalar $1+\mathbf{q}A^{-1}\mathbf{q}^*$ and on the right by $(A+\mathbf{q}^*\mathbf{q})^{-1}$, both of which are valid by assumption. ∎

Now using the definition of H, we see

$$(H - z\mathbb{1}) + z\mathbb{1} = H = \sum_{j=1}^{M}\mathbf{r}_j^*\mathbf{r}_j$$

and multiplying by $(H - z\mathbb{1})^{-1}$ on the right, we obtain

$$\mathbb{1} + zG = \mathbb{1} + z(H - z\mathbb{1})^{-1} = \sum_{j=1}^{M}\mathbf{r}_j^*\mathbf{r}_j(H - z\mathbb{1})^{-1} = \sum_{j=1}^{M}\mathbf{r}_j^*\mathbf{r}_jG. \tag{8.9}$$

Denote

$$G^{(i)} \equiv G^{(i)}(z) := (H^{(i)} - z\mathbb{1})^{-1}.$$

Now we use Lemma 8.2 with $\mathbf{q} = \mathbf{r}_i$ and $A = H^{(i)} - z\mathbb{1}$ on each term of the sum above, to see

$$\mathbf{r}_i(H - z\mathbb{1})^{-1} = \mathbf{r}_i(H^{(i)} - z\mathbb{1} + \mathbf{r}_i^*\mathbf{r}_i^*)^{-1} = \frac{1}{1 + \mathbf{r}_i(H^{(i)} - z\mathbb{1})^{-1}\mathbf{r}_i^*}\mathbf{r}_i(H^{(i)} - z\mathbb{1})^{-1},$$

which we can multiply by $\mathbf{r}^*$ on the left and sum over i, to find

$$\mathbb{1} + zG = \sum_{i=1}^{M}\mathbf{r}_i^*\mathbf{r}_i(H - z\mathbb{1})^{-1} = \sum_{i=1}^{M}\frac{1}{1 + \mathbf{r}_iG^{(i)}\mathbf{r}_i^*}\mathbf{r}_i^*\mathbf{r}_iG^{(i)}, \tag{8.10}$$

where we used (8.9). Now, taking the trace of equation (8.10) (using the trace's linearity) and dividing by M, gives by definition

$$
\begin{aligned}
\gamma_N + \gamma_N z m_N &= \frac{1}{M}\operatorname{Tr}(\mathbb{1}) + z\frac{N}{M}\frac{1}{N}\operatorname{Tr} G \\
&= \frac{1}{M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_i G^{(i)}\mathbf{r}_i^*}\operatorname{Tr}\left[\mathbf{r}_i^*\mathbf{r}_i G^{(i)}\right] \\
&= \frac{1}{M}\sum_{i=1}^{M}\frac{\mathbf{r}_i G^{(i)}\mathbf{r}_i^*}{1+\mathbf{r}_i G^{(i)}\mathbf{r}_i^*} \\
&= 1 - \frac{1}{M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_i G^{(i)}\mathbf{r}_i^*},
\end{aligned}
\tag{8.11}
$$

where the third equality follows by the cyclic property of the trace and the last equality follows from a simple partial fraction. Subtracting 1 from both sides of equation (8.11) and dividing by z yields

$$
\underline{m}_M = (\gamma_N - 1)\frac{1}{z} + \gamma_N m_N = -\frac{1}{z}\frac{1}{M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}G^{(i)}\mathbf{r}_i^*}
\tag{8.12}
$$

by equation (8.6). So, multiplying by T on the right gives us

$$
-z\underline{m}_M T = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}G^{(i)}\mathbf{r}_i^*}T
\tag{8.13}
$$

Next, we prove the following simple resolvent equation.

Lemma 8.3. *Fix $z \in \mathbb{C}$. For any $N \times N$ matrices A and B, such that $(A - z\mathbb{1})$ and $(B - z\mathbb{1})$ are invertible, we have*

$$
(A - z\mathbb{1})^{-1} - (B - z\mathbb{1})^{-1} = -(A - z\mathbb{1})^{-1}(A - B)(B - z\mathbb{1})^{-1}.
\tag{8.14}
$$

Proof. The proof is a simple manipulation. Note

$$
(A - z\mathbb{1})^{-1}(B - z\mathbb{1} + A - B) = (A - z\mathbb{1})^{-1}(A - z\mathbb{1}) = \mathbb{1}
$$

and then multiply on the right by $(B - z\mathbb{1})^{-1}$ to get

$$
(A - z\mathbb{1})^{-1} + (A - z\mathbb{1})^{-1}(A - B)(B - z\mathbb{1})^{-1} = (B - z\mathbb{1})^{-1}
$$

and complete the proof. ∎

Next we use the identity (8.14) and the definition of H to see

$$\begin{aligned}
(-z\underline{m}_M T - z\mathbb{1})^{-1} - G &= -(-z\underline{m}_M T - z\mathbb{1})^{-1}\left((-z\underline{m}_M T) - \sum_{i=1}^{N}(\mathbf{r}_i^*\mathbf{r}_i)\right)G \\
&= -(-z\underline{m}_M T - z\mathbb{1})^{-1}\left[(-z\underline{m}_M T)\,G - \left(\sum_{i=1}^{M}\mathbf{r}_i^*\mathbf{r}_i\right)G\right] \\
&= -(-z\underline{m}_M T - z\mathbb{1})^{-1}\left[\frac{1}{M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}G^{(i)}\mathbf{r}_i^*}TG - \left(\sum_{i=1}^{M}\mathbf{r}_i^*\mathbf{r}_i G\right)\right] \\
&= -(-z\underline{m}_M T - z\mathbb{1})^{-1}\frac{1}{M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_iG^{(i)}\mathbf{r}_i^*}TG \\
&\quad + (-z\underline{m}_M T - z\mathbb{1})^{-1}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_iG^{(i)}\mathbf{r}_i^*}\mathbf{r}_i^*\mathbf{r}_iG^{(i)} \\
&= \frac{1}{z}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_iG^{(i)}\mathbf{r}_i^*}\left[(-\underline{m}_M T - \mathbb{1})^{-1}\mathbf{r}_i^*\mathbf{r}_iG^{(i)} - \frac{1}{M}(-\underline{m}_M T - \mathbb{1})^{-1}TG\right],
\end{aligned}$$

where the second equality is a simple manipulation; the third equality follows from equation (8.13); the fourth follows by equations (8.9) and (8.10); and the last equality follows from factoring the above. Next we take the trace of the above and divide by N, to get

$$\begin{aligned}
\frac{1}{N}\operatorname{Tr}(-z\underline{m}_M T - z\mathbb{1})^{-1} - m_N &= \frac{1}{zN}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_iG^{(i)}\mathbf{r}_i^*} \\
&\quad\times\left(\operatorname{Tr}\left[(-\underline{m}_M T - \mathbb{1})^{-1}\mathbf{r}_i^*\mathbf{r}_iG^{(i)}\right] - \frac{1}{M}\operatorname{Tr}\left[(-\underline{m}_M T - \mathbb{1})^{-1}TG\right]\right) \\
&= \frac{1}{zN}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_iG^{(i)}\mathbf{r}_i^*} \\
&\quad\times\left[\mathbf{r}_iG^{(i)}(-\underline{m}_M T - \mathbb{1})^{-1}\mathbf{r}_i^* - \frac{1}{M}\operatorname{Tr}\left[G(-\underline{m}_M T - \mathbb{1})^{-1}T\right]\right] \\
&= \frac{1}{z\gamma M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_iG^{(i)}\mathbf{r}_i^*}d_i(z),
\end{aligned} \tag{8.15}$$

where

$$d_i(z) := \mathbf{r}_iG^{(i)}(z)(-\underline{m}_M(z)T - \mathbb{1})^{-1}\mathbf{r}_i^* - \frac{1}{M}\operatorname{Tr}\left[G(z)(-\underline{m}_M(z)T - \mathbb{1})^{-1}T\right]. \tag{8.16}$$

Let's examine the first term on the left-hand side of equation (8.15). We see

$$\frac{1}{N}\operatorname{Tr}(-z\underline{m}_M T - z\mathbb{1})^{-1} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{-z - t_i z\underline{m}_M}$$

as $-\underline{m}_M(z)T - \mathbb{1}$ is simply a rational transformation of the spectrum of T. Substituting for $\underline{m}_M(z)$ using equation (8.6), we see from equation (8.15)

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{-z-t_i\left(\gamma_N-1+\gamma_N m_N z\right)}-m_N=\frac{1}{z\gamma M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_i G^{(i)}\mathbf{r}_i^*}d_i$$

and then rearranging yields

$$m_N(z)-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{t_i\left(1-\gamma_N-\gamma_N z m_N(z)\right)-z}=-\frac{1}{z\gamma M}\sum_{i=1}^{M}\frac{1}{1+\mathbf{r}_i G^{(i)}(z)\mathbf{r}_i^*}d_i(z), \tag{8.17}$$

where the right-hand side is an error term which can be controlled. Equation (8.17) is an approximate version of equation (8.1).

9. Acknowledgements

The author would like to thank Professor Horng-Tzer Yau for his guidance and supervision since last summer, and acknowledge the constant patience and help afford by the postdoctoral follows, Alex Bloemendal and Kevin Schnelli, throughout the project. The author also thanks Antti Knowles for his useful comments and for pointing out the simplified proof of (5.11) in the large deviation estimate. This work was kindly supported by a grant from the Harvard College Research Program during the summer of 2012.

References

[1] Anderson, G., Guionnet, A., Zeitouni, O.: *An Introduction to Random Matrices.* Studies in advanced mathematics, **118**, Cambridge University Press, 2009.

[2] Bai, Z.D., Silverstein, J.W.: Spectral Analysis of Large Dimensional Random Matrices. Springer series in statistics, Springer, 2010.

[3] Bru, M.: Diffusions of perturbed principal component analysis. *J. Multivariate Anal.*, **29** 127-136 (1989).

[4] Diaconis, P.: What is... a random matrix? *Notices of the AMS*, **52**:11 1348-1349 (2005).

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

[6] Erdős, L., Farrell, B.: Local eigenvalue density for general MANOVA matrices. arXiv: 1207.0031v1.

[7] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős-Rényi graphs I: local semicircle law. To appear in *Ann. Prob.* Preprint arXiv:1103.1919

[8] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős-Rényi graphs II: eigenvalue spacing and the extreme eigenvalues. To appear in *Comm. Math. Phys.* Preprint arXiv:0911.3687.

[9] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Delocalization and diffusion profile for random band matrices. Preprint arXiv:1205.5669.

[10] Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: The local semicircle law for a general class of random matrices. Preprint arXiv:1212.0164.

[11] Erdős, L., Schlein, B., Yau, H.-T., Yin, J.: The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, **48** 1-46 (2012).

[12] Erdős, L., Yau, H.-T., Yin, J.: Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields* **154** 1-2, 341-407, (2012).

[13] Erdős, P. and Rényi, A.: On random graphs. I. *Publicationes Mathematicae*, **6**, 290–297 (1959).

[14] Erdős, P.; Rényi, A.: The evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **5** 17–61 (1960).

[15] Geronimo, J.; Hill, T.: Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *J. Approx. Theory*, **121** 54-60 (2003)

[16] Pillai, N., Yin, J.: Universality of covariance matrices. arXiv: 1110.2501v5.

[17] Marchenko, V., Pastur, A.: Distribution of eigenvalues for some sets of random matrices. *USSR-Sb.*, **1** 457-483 (1967).

[18] Mehta, M.L.: *Random Matrices.* Academic Press, New York, 1991.

[19] Silverstein, J.W., Bai, Z.D.: On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, **55** (2):175-192, (1995).

[20] Silverstein, J.W.: Strong convergence of the empirical distribution of eigenvalues or large dimensional random matrices. *J. Multivariate Anal.*, **55** (2):331-339, (1995).

[21] Wishart, J.: Generalized product moment distribution in samples. *Biometrika*, **20A** (12): 3252. (1928).

[22] Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.*, **62**, 548–564, (1955).