SPECTRAL PROPERTIES OF RANDOM MATRICES

Anastasis Kafetzopoulos

Doctor of Philosophy

School of Mathematics

Queen Mary University of London

April 2023

Statement of originality

I confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that Queen Mary, University of London has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Anastasis Kafetzopoulos

Date: April 2023

Details of publications:

1. Local Marchenko-Pastur law. [Journal of Mathematical Physics]
2. Universality of correlated covariance matrices. [in preparation]

Abstract

In the first part of this thesis, we give the theoretical foundations of random matrix theory through the definitions of a random matrix, a random probability measure and the corresponding empirical spectral distribution we will be working with. The main technical tool of the first paper is also defined rigorously and analyzed deeply, which is the Stieltjes transform method.

We then use this tool to prove optimal convergence of the empirical spectral distribution of random sample covariance matrices to the deterministic Marchenko-Pastur distribution. We also give new results about the rigidity of the eigenvalues of this random sample covariance matrix as well as about the rate of their convergence.

In the second part of this thesis, we define another important and more general technical tool which works additionally well with non-Hermitian random matrices and that is the Dyson equation method which was used in the second paper. Just like the Stieltjes transform method, it is also defined rigorously and analyzed deeply.

We then prove new local laws about a random matrix model that interpolates between the Marchenko-Pastur distribution, the elliptical law and the circular law. Through our work these local laws can now be considered universal, which means that they are independent of the initial distribution of the random matrix entries.

We finally give an overview of our new results and provide new directions of study.

Acknowledgements

Contents

Bibliography

Chapter 1

Introduction

1.1   Random matrix theory

Random matrix theory is an active research area of modern mathematics combining Mathematical and Theoretical Physics, Mathematical Analysis and Probability, and with numerous applications, for example in Theoretical Physics, Number Theory, Combinatorics and further in Statistics, Financial Mathematics, Biology and Engineering. [4], (ch.3: applications of random matrices).

The main goal of Random Matrix Theory is to provide an understanding of the properties of random matrices, that is matrices that have as elements random variables, real, complex or quaternion. One important direction of investigation is the study of the eigenvalue distribution of such matrices when the matrix dimension is large. Many quantities in physics and other applied areas are modeled by the eigenvalues of such matrices.

The origins of random matrices can be traced back to the works of Wishart in 1928 and James in 1960 in the discipline of Statistics, where random matrices were used to derive distributions for numerous statistics of normal multivariate random variables. A lot of development in this mathematical field was due to the work of Wigner in Nuclear Physics around 1950. Wigner was the first to suggest that the fluctuations of nuclei

resonances can be described in terms of the properties of eigenvalues of very large symmetric matrices with random entries.

An observable in quantum mechanics can be characterized by a self-adjoint linear operator in a Hilbert space, its Hamiltonian, which we may think informally of as a matrix of infinitely many dimensions. This suggests that random matrix theory should be dealing with general properties of the underlying generic Hamiltonians, most importantly such global features as the Hermiticity, the time-inversion invariance as well as other symmetries Hamiltonians may obey from general principles. Wigner hoped that the output of the model, that is the distribution of the eigenvalues for large-dimensional Hamiltonians will be universal and independent of the details of the entries. It would be common to the majority of systems sharing the corresponding symmetries. [46] (section 1.1: random matrices in nuclear physics).

Along those lines Wigner succeeded in calculating nontrivial spectral characteristics of random real symmetric matrices with independent, identically distributed entries, the mean density of the eigenvalues and demonstrated that in the limit of large matrix size it is given quite generally by the semi-circle law.

1.2 The Wigner semi-circle law

Consider a random symmetric $N \times N$ matrix X with independent and identically distributed entries (x_{ij}) for $i \leq j \in \{1, ..., N\}$ and then consider the distribution of its random eigenvalues.

It turns out that as the dimension of the matrix increases the random distribution of the eigenvalues becomes deterministic provided we impose these two conditions on the matrix entries so that they don't get too large:

$$\mathbb{E}[x_{ij}] = 0, \quad \mathrm{Var}[x_{ij}] = 1.$$

Since the dimension is also getting large we need to normalize the matrix so that the

eigenvalue distribution is confined to an interval independent of N. For this reason, we consider the eigenvalues of the random matrix $W := \frac{1}{\sqrt{N}} X$ with

$$\mathbb{E}[w_{ij}] = 0, \quad \text{Var}[w_{ij}] = \frac{1}{N}.$$

It turns out that the relative frequency histogram ρ_N of the positions of the N scaled random eigenvalues for large N, is close to the following deterministic law f:

$$\rho_N \approx f,$$

where f is a probability density supported in $[-2, 2]$, which is a scaled semi-circle centered at $(0, 0)$:

$$f(x) = \frac{1}{2\pi} \sqrt{4 - x^2}.$$
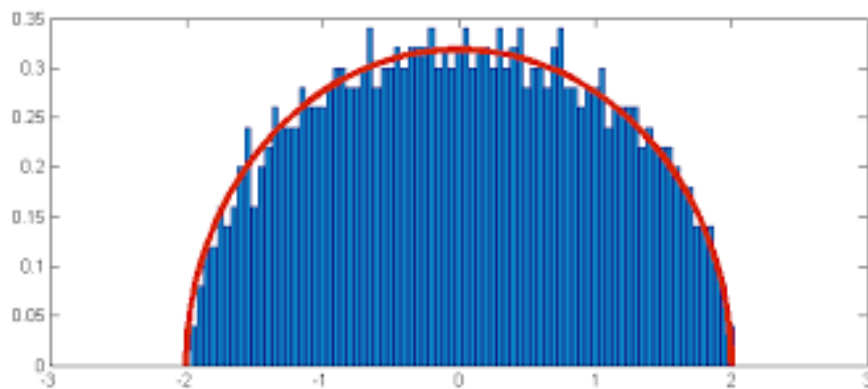
Figure 1.1: Semi-circle law from [27].

This was proved back in 1958 by Wigner, [64]. The approximation details will be clarified in the following chapters.

1.3 The Marchenko-Pastur law

Another famous example of a deterministic distribution comes from the consideration of a large random sample covariance matrix.

For this, we consider firstly a random $N \times M$ matrix X with i.i.d. centered entries with unit variance and then multiply it with its transpose. That means that we are considering the random covariance matrix XX^T and get to study its random eigenvalues, after we normalize it as $W := \frac{1}{N}XX^T$.

For this model, it turns out that as the dimensions N and M increase, the relative frequency histogram μ_N of the positions of the N random eigenvalues of the Hermitian and positive semi-definite square matrix $\frac{1}{N}XX^T$ approaches the following law g:

$$\mu_N \approx g,$$

where g is a probability density supported in $[\lambda_-, \lambda_+]$, which is called the Marchenko-Pastur distribution. The constants λ_-, λ_+ are defines as follows. First, we take the limit:

$$d := \lim_{N,M\to\infty} \frac{M}{N} \in (0,1],$$

and the we define the endpoints:

$$\lambda_\pm := (1 \pm d)^2.$$

The distribution g is then given by the following density,

$$g(x) = \frac{1}{2\pi\lambda} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{x}, \quad x \in [\lambda_-, \lambda_+].$$

This was proved back in 1973 by Marchenko and Pastur, [44]. Again, the approximation details will be clarified in the next chapters.
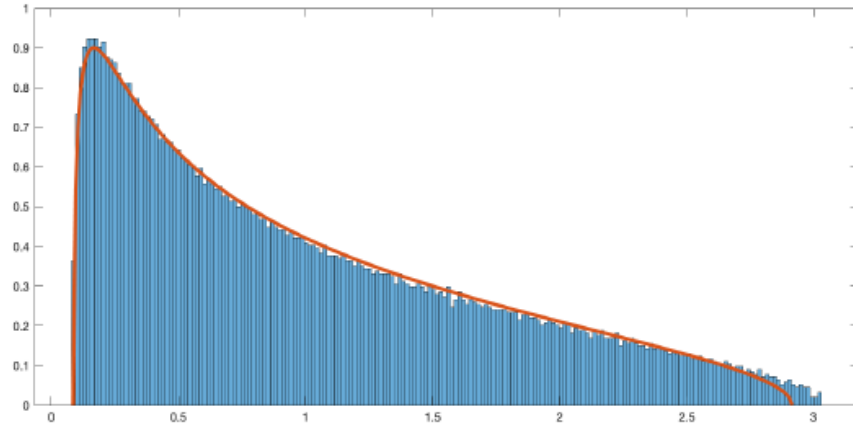
Figure 1.2: Marchenko-Pastur law from [13].

1.4 The circular law

We now turn our attention to non-Hermitian random matrix models which generally have complex eigenvalues.

One famous such example is the case of a random matrix X that has independent identically distributed complex entries, without any required symmetry in them.

This means that we are considering an $N \times N$ random matrix X with i.i.d. complex entries (x_{ij}) such that:

$$\mathbb{E}[\text{Re}(x_{ij})] = 0, \quad \text{Var}[\text{Re}(x_{ij})] = 1/2,$$

$$\mathbb{E}[\text{Im}(x_{ij})] = 0, \quad \text{Var}[\text{Im}(x_{ij})] = 1/2.$$

If we consider now the 2-dimensional relative frequency histogram c_N of the positions of the N random eigenvalues of the scaled matrix $W := \frac{1}{\sqrt{N}} X$ with expectation zero and complex variance $\frac{1}{N}$, we get that:

$$c_N \approx h,$$

where h is the complex uniform distribution on the unit disk, i.e.:

$$h(z) = 1_{|z| \leq 1}.$$
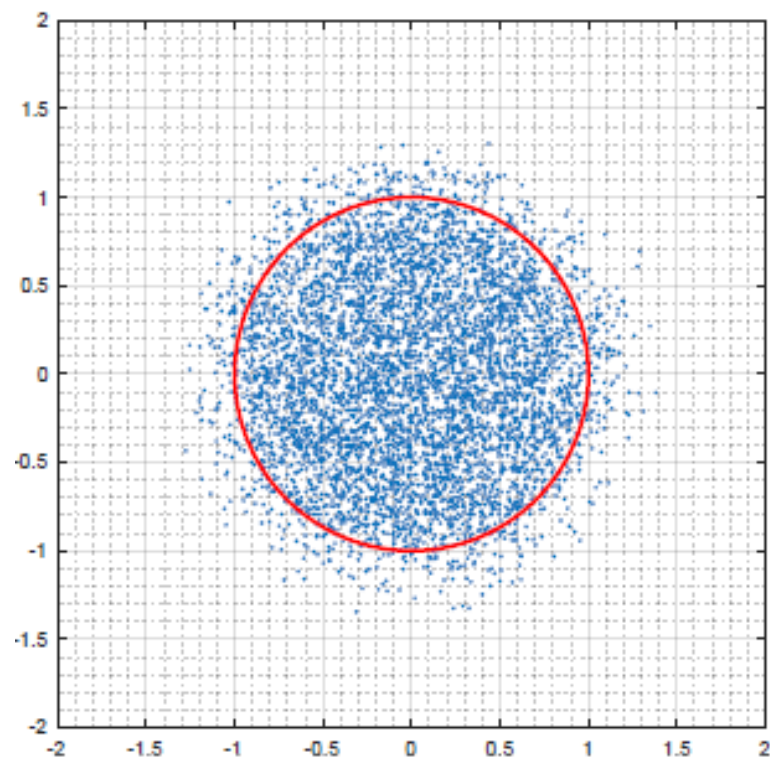
Figure 1.3: Circular law from [12].

This is known as the circular law. For random matrices with a Gaussian distribution for the entries, the circular law was established in 1965 by Ginibre, see [31]. Around 1980, Girko introduced an approach which allowed to establish the circular law for more general distributions of the matrix entries, see [32].

1.5 The elliptic law

The circular law result was extended in 1988 by Sommers, Crisanti, Sompolinsky and Stein to an elliptical law for ensembles of matrices with the properties described next, (see [59]).

Let (ξ_1, ξ_2) be a random vector in $\mathbb{C}^2$ where both ξ_1 and ξ_2 have mean zero and unit complex variance. We say that a random matrix $X = (x_{ij})_{i,j=1}^N$ belongs to the complex elliptic ensemble if the following conditions hold:

- (Independence). $\{x_{ii} : i \geq 1\} \cup \{(x_{ij}, x_{ji}) : 1 \leq i < j\}$ is a collection of independent random elements.

- (Common distribution). Each pair (x_{ij}, x_{ji}), $1 \leq i < j$ is an i.i.d. copy of (ξ_1, ξ_2).

- (Flexibility of the main diagonal). The diagonal elements $\{x_{ii} : i \geq 1\}$ are i.i.d. with mean zero and finite variance.

- (Correlations). We have that $\mathbb{E}[\xi_1 \xi_2] = \rho$, where $\rho \in (-1, 1)$ is a universal correlation constant for the random matrix X.

If d_N is the 2-dimensional relative frequency histogram of the positions of the random eigenvalues of the scaled version of this random matrix model $W := \frac{1}{\sqrt{N}} X$, then we have that:

$$d_N \approx k,$$

where k is the complex uniform distribution on the ellipse centered at $(0, 0)$ and with its edges at $(1 - \rho, 0)$ and $(1 + \rho, 0)$:

$$k(z) := \frac{1}{\pi(1 - \rho^2)} \mathbf{1}_{\left\{ z \in \mathbb{C} \; \middle| \; \frac{(\mathrm{Re}z)^2}{(1+\rho)^2} + \frac{(\mathrm{Im}z)^2}{(1-\rho)^2} \leq 1 \right\}}.$$
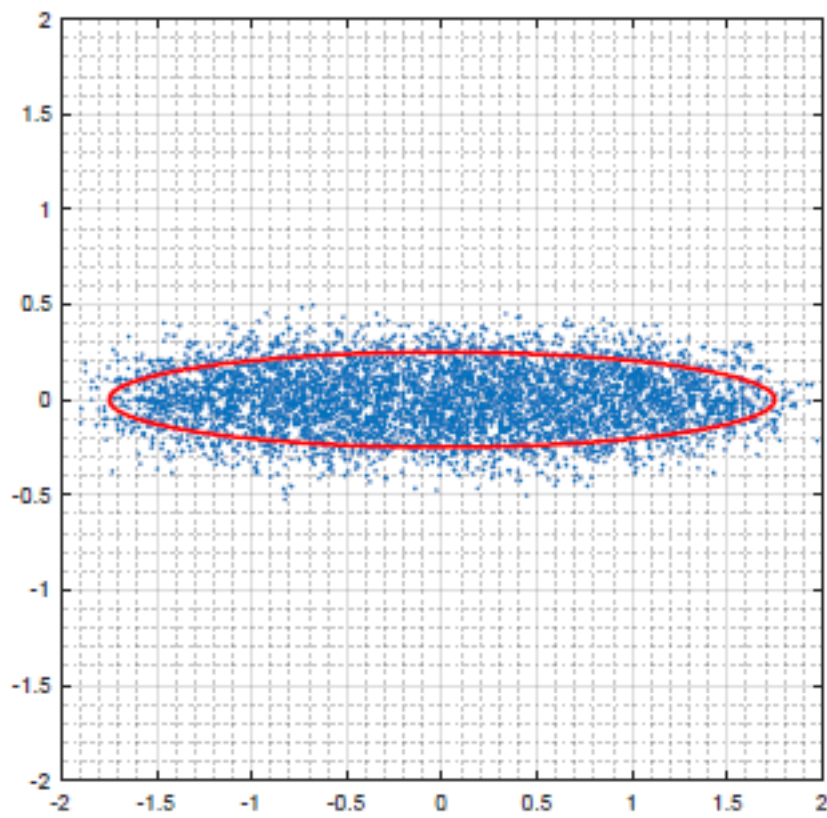
Figure 1.4: Elliptic law from [12].

1.6   Local laws

Since the random matrix W we are usually dealing with is normalized, we have that its spectrum is of constant order, or in other words the operator norm $\|W\|$ is of order 1. Since W has N eigenvalues, the typical separation of the eigenvalues is of order $\frac{1}{N}$. Individual eigenvalues are expected to fluctuate from their mean locations but as it turns out for eigenvalues of random matrices these fluctuation are really small, eigenvalues are rigid.

This allows us to talk about convergence of the eigenvalue distribution in short scales on an interval where there is just a constant $\mathcal{O}(1)$ amount of them in contrast to a large $\mathcal{O}(N)$ amount of them. In general, if the fluctuations of a random process are not small this can turn out to be impossible, as it is for example in the Poisson point process.
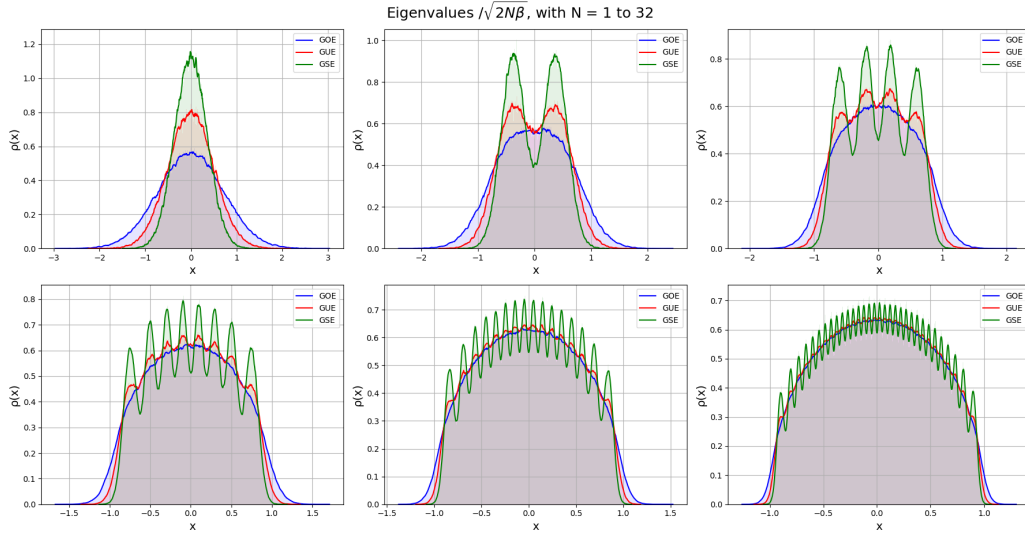
Figure 1.5: Local laws.

This convergence on short scales for the eigenvalue distribution has numerous applications and is technically useful. We give a description of it for our previous models:

<u>Wigner matrix.</u> We will define a certain "zooming mechanism" regarding the "scale" of convergence of the random distribution of the eigenvalues, that is a mechanism for creating peaks at each interval in the real line with a constant amount of eigenvalues and a smoothing-out until the next one. Here, the typical separation of the eigenvalues is of order $\mathcal{O}(\frac{1}{N})$ everywhere in the support except from the edges and this mechanism can create peaks in each such small interval of that order where there is a constant amount of eigenvalues. Near the edges we have a typical separation of order $\mathcal{O}(\frac{1}{N^{2/3}})$. We still get convergence to the semi-circle law with such a "bandwidth" aligned to the random eigenvalue distribution.

<u>Covariance matrix.</u> We will define the same mechanism for the eigenvalue distribution of a random covariance matrix. There is a strong accumulation of eigenvalues near zero when $d = 1$, which creates a singularity in the limiting Marchenko-Pastur distribution which approximates the random eigenvalue distribution. Due to this, the typical separation of eigenvalues there is of smaller order $\mathcal{O}(\frac{1}{N^2})$. Away from this point, the

separation is of typical order $\mathcal{O}(\frac{1}{N})$ but again, near the right edge there must be a smaller amount of eigenvalues and so the separation is of slightly larger order $\mathcal{O}(\frac{1}{N^{2/3}})$. We still get convergence to the Marchenko-Pastur distribution when we "zoom in" to these intervals but now greater care is needed for our mechanism process, especially near the singularity.

Circular law. Here the zooming process is different since the eigenvalues are complex. We use a mechanism for creating peaks arising from small circles around each individual eigenvalue. Here the typical separation of the eigenvalues is of order $\mathcal{O}(\frac{1}{\sqrt{N}})$ everywhere inside the support of the unit disk. This means that we can create a random distribution which gives more mass to circles of diameter $\mathcal{O}(\frac{1}{\sqrt{N}})$ inside the unit disk while also following the random eigenvalue distribution. This would still converge to the the circular law for large N.

Elliptic law. We use the same zooming process as we do with the local circular law to create small circles inside the domain of the ellipse of diameter $\mathcal{O}(\frac{1}{\sqrt{N}})$, which is the typical separation of the random eigenvalues inside the ellipse. The eigenvalue distribution, aligned with this zooming mechanism still converges to the elliptic law.

1.7 Discussion about the new results

This section will be technical and will provide a description about the past results and the new original contributions that have been made during my work as a PhD student at Queen Mary university.

1. Local Marchenko-Pastur law.

Let $X_N = X/\sqrt{N}$, where X is a $N \times N$ matrix with entries $x_{ij} = \mathrm{Re}(x_{ij}) + i\mathrm{Im}(x_{ij})$, where $\mathrm{Re}(x_{ij}), \mathrm{Im}(x_{ij})$ are i.i.d. real random variables with mean 0 and variance $\frac{1}{2}$, so that $\mathbb{E}(x_{ij}) = 0$ and $\mathbb{E}|x_{ij}|^2 = 1$ for $i, j = 1, ..., N$. We assume that the random variables have finite fourth moments.

Let λ_i, $i = 1, ..., N$ denote the eigenvalues of the positive semi-definite matrix $X_N^* X_N$

with $0 \leq \lambda_1 \leq ... \leq \lambda_N$. If we denote by ρ_N the empirical spectral density of the eigenvalues:

$$\rho_N(E) = \frac{1}{N} \# \left\{ i \leq N | s_i \leq E \right\},$$

and by $\rho(E)$ the density of the Marchenko-Pastur distribution:

$$\rho(E) = \frac{1}{2\pi} \sqrt{\frac{4}{E} - 1},$$

we have that $\rho_N \to P$ weakly in probability, where P is the cumulative density of ρ. We show that this convergence holds locally at an optimal scale. Let's comment here about the previous results so far.

In [22], Erdös-Schlein-Yau-Yin established the convergence of the empirical spectral density for general covariance matrices to the Marchenko-Pastur law in the bulk on small intervals for X_N being a $M \times N$ matrix with $M < N$. They used a decomposition by minors for the diagonal elements of the resolvent to establish a self-consistent equation for the Stieltjes transform S_N of ρ_N. Large deviation estimates and a continuity argument were then used to show the convergence of the spectral measure on small intervals (involving polynomial corrections) in the bulk distribution. These methods have been extended to the "hard" edge for $N \times N$ matrices and logarithmic rather than polynomial corrections by Cacciapuoti-Maltsev-Schlein in [14]. More precisely, the authors showed that the fluctuation of the Stieltjes transform $\sqrt{E} S_N$ away from $\sqrt{E} S_\rho$ is on the order of $\sqrt{\frac{\sqrt{E}}{N\eta}}$ and they obtained convergence of the counting function of the eigenvalues everywhere including close to the "hard" edge. Eigenvalue rigidity with polynomial corrections for the "bulk" and "soft" edges for entries with sub-exponential decay can be found in Pillai-Yin [54].

In this work, we obtain optimal bounds on the expectations of high moments of the fluctuation $\Lambda = S_N - S_\rho$ on the optimal scale and the previous bound $\sqrt{\frac{\sqrt{E}}{N\eta}}$ gets improved to $\frac{\sqrt{E}}{N\eta}$ without any logarithmic corrections. Our methods and results apply to the "bulk" as well as to the "soft" and "hard" edges. The main objective of this work is to extend the results and the methodology of [15] to a "hard edge setup". This is the case

when the limiting measure has a square root singularity near 0 with typical distance between the eigenvalues of the order of $\frac{1}{N^2}$. We were able to simplify the proof of Theorem 4.2.1 in [15] by avoiding different cases for the "bulk" and the "edges". Unlike in the Wigner case, where both edges are "soft", the presence of the "hard" edge at 0 allows us to extend the bounds on the real part of the Stieltjes transform to the negative real line, thus also yielding a fluctuation for the individual eigenvalue near the "hard" edge that is decreasing with the eigenvalue number. In summary, this paper improves on [14] by removing the logarithmic corrections and improving the fluctuation bounds while keeping the same optimal scale of convergence. We also extended the proofs in [35, 62] on fluctuations of quadratic forms to a "soft edge setup" by improving a factor of $|S_\rho|$ to a factor of $\text{Im}(S_\rho)$.

To show our results, we used the Stieltjes transform method which involves a parameter $z \in \mathbb{C}$ that depends on N. We define:

$$S_N(z) = \int_\mathbb{R} \frac{1}{x-z} \rho_N(x) dx$$

and

$$S(z) = \int_\mathbb{R} \frac{1}{x-z} \rho(x) dx,$$

where $z = E + i\eta$, with $E \in \mathbb{R}$ representing the position of an eigenvalue and $0 < \eta(N) \leq 1$ representing the local scale of the convergence for the two distributions. It is proven that for:

$$\frac{N\eta}{|\sqrt{z}|} \geq M, \tag{1.1}$$

for a suitable large constant M and for each power q with $1 \leq q \leq c_0 \left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/8}$ and for each $z \in Z_{E,\eta} = \left\{E, \eta \in \mathbb{R} : E^2 + \eta^2 - 4|\eta| \leq 4E\right\}$ or $E < 0$:

$$\mathbb{P}\left(|\sqrt{z}(S_N(z) - S_\rho(z))| \geq K \frac{|\sqrt{z}|}{N\eta}\right) \leq \frac{(Cq)^{cq^2}}{K^q}, \tag{1.2}$$

for some constants C, c_0 and for each $K > 0$.

Relation (1.1) gives the optimal scaling according to the position of the eigenvalues,

namely we have that for E close to 0 :

$\eta \gtrsim \frac{1}{N^2}$ while for E away from the "hard" edge: $\eta \gtrsim \frac{1}{N}$.

Relation (1.2) gives the optimal bound on the convergence of the Stieltjes transforms. In the "hard" edge the term $\sqrt{|z|}$ corrects this convergence, as we have that $|\sqrt{z}(S_N(z) - S_\rho(z))| \to 0$ in probability in the optimal scaling.

This theorem has as a consequence a bound on the convergence rate of the empirical spectral distribution to the Marchenko-Pastur distribution. Letting:

$$P(E) = \int_0^E \rho(x)dx,$$

we compare it to ρ_N. With the same assumption as before, we have that:

$$\mathbb{P}\left(|\rho_N(E) - P(E)| \geq K \min\left\{\sqrt{E}, \frac{\log N}{N}\right\}\right) \leq \frac{(Cq)^{cq^2}}{K^q}, \tag{1.3}$$

for all $E \in \mathbb{R}$, $K > 0$, $N > N_0$ and for each $q \in \mathbb{N}$ with $1 \leq q \leq c_0 \left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/8}$.

Relation (1.3) is not optimal since it is known that the fluctuations between the number of eigenvalues predicted by the counting measure and the M-P distribution are of the order of $\frac{\sqrt{\log N}}{N}$ but the technique used here couldn't provide such a result.

Finally the rigidity of the eigenvalues was proven. There exist constants $C, c, N_0, \epsilon > 0$ such that:

$$\mathbb{P}\left(|\lambda_i - \gamma_i| \geq K \frac{\log N}{N}\left(\frac{i}{N}\right)\right) \leq \frac{(Cq)^{cq^2}}{K^q}, \tag{1.4}$$

for $i = 1, ..., \lceil N/2 \rceil$, $N > N_0$, $K > 0$, and $q \in \mathbb{N}$ with $q \leq N^\epsilon$.

Furthermore, for $i \leq \log N$ we have that:

$$\mathbb{P}\left(|\lambda_i - \gamma_i| \geq K \frac{i^2}{N^2}\right) \leq \frac{(Cq)^{cq^2}}{K^{q/2}}. \tag{1.5}$$

In this theorem the factor $\frac{i}{N}$ accounts for the higher density at the "hard" edge. Here we focus on hard-edge rigidity, since proofs of soft-edge rigidity require control of the

largest eigenvalue which, to our knowledge, is not currently available in the case of truncated entries with four moments, in either the Wigner or the Sample Covariance case.

The terms γ_i are defined as the classical locations of the eigenvalues predicted by the M-P distribution, that is:

$$\int_0^{\gamma_i} \rho(E)dE = \frac{i}{N}.$$

In particular, we obtain the fluctuation of the eigenvalues near the "hard" edge to be of the order of $\frac{\log N}{N^2}$. The fluctuations of eigenvalues in the "bulk" and "soft" edges of both the Gaussian Unitary Ensemble and the Wishart Ensemble are known to be respectively of the order $\frac{\sqrt{\log N}}{N}$ in the bulk and $\frac{\sqrt{\log k}}{k^{1/3} N^{2/3}}$ for the k-th eigenvalue from the edge, $k \to \infty$ (see [36, 60]). To our knowledge similar results are not yet available for the "hard" edge.

2. Universality of correlated covariance matrices.

In the second part of this thesis, we study the spectrum of matrices of the form $X_1 X_2^*$, where the matrices X_1 and X_2 are correlated according to a parameter $\tau \in [0, 1]$. This topic has a particularly wide application in Quantum Chromodynamics.

For $\tau = 0$ we have two completely uncorrelated matrices whereas for $\tau = 1$ we have two completely correlated matrices ($X_1 = X_2$) and we can recover the Marchenko-Pastur distribution in the limit $N \to \infty$.

The goal of this analysis is to determine the distribution (law and support) of the complex eigenvalues of $X_1 X_2^*$ when the entries of each matrix are general i.i.d. random variables. This means that we are proving universality results, so that we can extend a recent work by Akemann, Byun and Kang in [5], where the entries there were assumed to be Gaussian. Due to this assumption, the techniques that were used there could not be adapted to our analysis, so we followed the techniques used in [7].

Specifically, the model we are examining are matrices of the form $X_1 X_2^*$ where

$$X_1 = \sqrt{1+\tau}P + \sqrt{1-\tau}Q, \quad X_2 = \sqrt{1+\tau}P - \sqrt{1-\tau}Q$$

and P, Q are $N \times N$ random matrices with i.i.d. entries and τ is a parameter that models the correlation between the matrices X_1 and X_2.

In [5] it was shown that the eigenvalue counting measure of the matrix $X = X_1 X_2^*$ (for random Gaussian entries of the matrices P, Q) converges to the measure $\hat{\mu}$, where

$$d\hat{\mu}(\zeta) = \frac{1}{1-\tau^2} \frac{1}{2\pi|\zeta|} \mathbf{1}_{\hat{S}_\tau}(\zeta) d^2\zeta \tag{1.6}$$

and $d^2\zeta$ is the Lebesgue measure on the complex plane. The support $\widehat{S}_\tau$ of this measure is given by:

$$\hat{S}_\tau := \left\{ \zeta = x + iy : \left(\frac{x-2\tau}{1+\tau^2}\right)^2 + \left(\frac{y}{1-\tau^2}\right)^2 \leq 1 \right\}. \tag{1.7}$$

It represents a transition from the circular law to an ellipsoid and finally to the Marchenko-Pastur law according to the values of the correlation parameter τ.

The universal result we prove here is the convergence of the empirical spectral density to the measure $d\hat{\mu}$ independently of the choice of the distributions for P, Q and extend it on local scales.

The techniques used are based on the recent publication "Local elliptic law" [7], where the Dyson equation method is used for non-Hermitian matrices with correlated entries after they get "hermitized" through the Hermitization technique. The Dyson equation is a powerful technique and except for universality, it can also provide local laws for the limiting spectral measure.

The main idea in our analysis is to firstly prove a universal uniform limiting local elliptic law for the distribution of the eigenvalues of the auxiliary *Dirac* matrix:

$$D := \begin{pmatrix} 0 & X_1 \\ X_2^* & 0 \end{pmatrix} \tag{1.8}$$

and then recover the eigenvalues of the correlated covariance matrix $X_1 X_2^*$. The Hermitization matrix, resolvent matrix and then the Dyson equation are defined for this auxiliary matrix to provide the uniform elliptic distribution for its eigenvalues depending on the non-hermiticity parameter τ, in a similar way to [7]. Specifically, for the Dirac matrix, we have a local law given by the distribution:

$$d\mu(\zeta) := \frac{1}{1 - \tau^2} \frac{1}{\pi} \mathbf{1}_{S_\tau}(\zeta) d^2 \zeta. \tag{1.9}$$

with its support S_τ given by:

$$S_\tau := \left\{ \zeta = x + iy : \left(\frac{x}{1 + \tau} \right)^2 + \left(\frac{y}{1 - \tau} \right)^2 \leq 1 \right\}. \tag{1.10}$$

Much different analysis for the stability of the Dyson equation was now used to establish this local law as the equation is now given by 4×4 matrices instead of 2×2 as in [7]. The least singular value is controlled after a bounded density assumption for the entries of P, Q as in [6].

After establishing the local law for the Dirac matrix, the linearization technique as used for example in [51], together with a careful change of variables is then enough to provide us with the limiting local spectral law for the initial correlated covariance matrix.

Chapter 2

Preliminaries

Before going to the analysis of our results, we present here some preliminaries about random matrices and their spectral properties.

2.1 Random matrices

In this section, the notion of a random matrix is defined rigorously. We will be working with matrices which have their entries in the complex plane $\mathbb{C}$. We denote by $\mathbb{C}^{N \times N}$ the space of $N \times N$ matrices X with complex entries $(X_{ij})_{i,j=1}^N$, which can be regarded as vectors in $\mathbb{C}^{N^2}$. We can equip this space with a norm such as the operator norm $\| \cdot \|_{op}$ or the Hilbert-Schmidt norm $\| \cdot \|_{HS}$. The operator norm is defined with respect to the Euclidean norm $\| \cdot \|_2$ in $\mathbb{C}^N$, as follows:

$$\|X\|_{op} := \sup \left\{ \|Xu\|_2 : \ u \in \mathbb{C}^N \ , \ \|u\|_2 = 1 \right\},$$

while the Hilbert-Schmidt norm $\| \cdot \|_{HS}$ is defined as:

$$\|X\|_{HS} := \sqrt{\mathrm{Tr}(XX^*)},$$

where X^* is the conjugate transpose of X and we also defined the *trace* functional $\text{Tr} : \mathbb{C}^{N \times N} \mapsto \mathbb{C}$ of a matrix X as:

$$\text{Tr}(X) := \sum_{i=1}^{N} X_{ii}.$$

Some remarks about these norms are that they are equivalent to each other for any finite dimension $N \in \mathbb{N}$ and that the operator norm is in some sense similar to the maximum norm $\| \cdot \|_\infty$ in $\mathbb{C}^{N^2}$ while the Hilbert-Schmidt norm similar to the Euclidean norm $\| \cdot \|_2$ in $\mathbb{C}^{N^2}$. The operator norm transforms $(\mathbb{C}^{N \times N}, \| \cdot \|_{op})$ into a Banach space while the Hilbert-Schmidt norm transforms $(\mathbb{C}^{N \times N}, \| \cdot \|_{HS})$ into a Hilbert space with corresponding inner product:

$$\langle X, Y \rangle := \text{Tr}(XY^*).$$

Another remark is that the convergence of a sequence of matrices $\left\{ X^{(k)} \right\}_{k=1}^\infty$, for $k \to \infty$ under any of these two norms in $\mathbb{C}^{N \times N}$ is equivalent to the convergence of each matrix entry $\left\{ X_{ij}^{(k)} \right\}_{k=1}^\infty$ in $\mathbb{C}$ for $i, j = 1, ..., N$, under the usual complex norm.

A matrix $X \in \mathbb{C}^{N \times N}$ is called *Hermitian* if the equality $X = X^*$ holds. An *eigenvalue* $\lambda \in \mathbb{C}$ of a matrix $X \in \mathbb{C}^{N \times N}$ is a complex number such that $Xu = \lambda u$ for some $u \in \mathbb{C}^N$ different than the zero vector. Counting multiplicities, we can say that each complex matrix in $\mathbb{C}^{N \times N}$ has exactly N eigenvalues. The set $\text{Spec}(X) = \{\lambda_1, ..., \lambda_N\}$ of the eigenvalues of a matrix X is called its *spectrum.* We say that a matrix $X \in \mathbb{C}^{N \times N}$ is *invertible* if there exists a matrix $Y \in \mathbb{C}^{N \times N}$ such that $XY = YX = I$ and we write then $Y = X^{-1}$. We also define the *determinant* functional $\det : \mathbb{C}^{N \times N} \mapsto \mathbb{C}$ as $\det(X) = \prod_{i=1}^{N} \lambda_i$, where $\lambda_1, ..., \lambda_N$ are the N eigenvalues of the matrix X.

Some properties of the trace functional are the following:

- It is a continuous linear functional from $(\mathbb{C}^{N \times N}, \| \cdot \|_{op})$ to $(\mathbb{C}, \| \cdot \|)$.

- For each matrix $X \in \mathbb{C}^{N \times N}$ with corresponding eigenvalues $\lambda_1, ..., \lambda_N$ the equality $\text{Tr}(X) = \sum_{i=1}^{N} \lambda_i$ holds.

- For each $S, M \in \mathbb{C}^{N \times N}$ we have that $\mathrm{Tr}(SM) = \mathrm{Tr}(MS)$.

In the following we present two important theorems about the eigenvalue structure of a matrix. We note that if a matrix X is Hermitian then all of its eigenvalues are real. We denote by $\mathrm{diag}(a_1, ..., a_N)$ the *diagonal* matrix $D \in \mathbb{C}^{N \times N}$ such that $D_{ii} = a_i$ for $i = 1, ..., N$ and $D_{ij} = 0$ for $i \neq j \in \{1, ..., N\}$. A *triangular* matrix is a matrix $T \in \mathbb{C}^{N \times N}$ such that $T_{ij} = 0$ for all $i > j$. A *unitary* matrix $U \in \mathbb{C}^{N \times N}$ is an invertible complex matrix such that $U^{-1} = U^*$.

Theorem 2.1.1 (Spectral theorem). *For any Hermitian matrix $X \in \mathbb{C}^{N \times N}$ there exists a unitary matrix $U \in \mathbb{C}^{N \times N}$ such that $U^{-1}XU = \mathrm{diag}(\lambda_1, ..., \lambda_N)$, where $\lambda_1, ..., \lambda_N$ are the N real eigenvalues of the Hermitian matrix X.*

Theorem 2.1.2 (Schur decomposition). *For any matrix $X \in \mathbb{C}^{N \times N}$ there exists a unitary matrix $U \in \mathbb{C}^{N \times N}$ such that $U^{-1}XU = T$, where T is a triangular matrix, satisfying $T_{ii} = \lambda_i$, for $i = 1, ..., N$ where λ_i are the eigenvalues of X.*

For a specific $i \in \{1, ..., N\}$, the i-th eigenvalue λ_X^i of a matrix $X \in \mathbb{C}^{N \times N}$ is a continuous function in the topological sense. This means that if $X_k \to X$ for a sequence of matrices $\{X_k\}_{k=1}^{\infty}$ under the operator or Hilbert-Schmidt norm, then the eigenvalues $\{\lambda_{X_k}^i\}_{i=1}^{N}$ can be numbered in such a way that $\lambda_{X_k}^i \to \lambda_X^i$. See [9], pages 138-139.

This is a consequence of the fact that the roots of a polynomial are continuous functions of the polynomial coefficients under a suitable norm, and this fact is applied to the *characteristic polynomial* of a matrix, $p(x) := \det(xI - X)$, see again [9].

We now turn our attention to the definition of a random matrix.

Definition 2.1.1 (Random matrix). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $N \in \mathbb{N}$. An $N \times N$ random matrix is a measurable map $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto (\mathbb{C}^{N \times N}, \mathcal{B}^{N^2})$, where $\mathcal{B}^{N^2}$ denotes the Borel σ-algebra on $\mathbb{C}^{N \times N}$.*

It is clear that X is a measurable function if and only if all entries $X(i, j)$ are measurable functions in $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ where $\mathcal{B}_{\mathbb{C}}$ denotes the Borel σ-algebra in $\mathbb{C}$. We show next that the

eigenvalues of a random matrix are also measurable functions. Specifically, we show that the eigenvalue function:

$$\omega \mapsto \lambda^i_{X(\omega)} : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{C},$$

where $\lambda^i_{X(\omega)}$ denotes the i-th eigenvalue of the random matrix $X(\omega)$ is measurable, which makes it a complex-valued random variable.

We know that the eigenvalue function:

$$\mathrm{eig}^i : \mathbb{C}^{N \times N} \mapsto \mathbb{C},$$

$$X \mapsto \lambda^i_X$$

is continuous so it is measurable. The map $X : \Omega \mapsto \mathbb{C}^{N \times N}$ is also measurable by definition, hence the composition $\lambda^i_X := \mathrm{eig}^i \circ X$ is also measurable.

This property of the eigenvalue function allows us to study the eigenvalues of a random matrix in the context of probability theory. For this, we firstly need to define the concept of a random measure.

2.2 Random probability measures

We denote the set of all measures on $(\mathbb{C}, \mathcal{B}_\mathbb{C})$ by $\mathcal{M}(\mathbb{C})$, the set of probability measures by $\mathcal{M}_1(\mathbb{C})$ and the set of sub-probability measures by $\mathcal{M}_{\leq 1}(\mathbb{C})$. If $f : \mathbb{C} \to \mathbb{C}$ is a measurable function, we write:

$$\langle \mu, f \rangle := \int_\mathbb{C} f d\mu$$

We are interested in studying the convergence behaviour of sequences of measures in $\mathcal{M}_1(\mathbb{C})$ where the limit may generally lie in $\mathcal{M}_{\leq 1}(\mathbb{C})$. For this, we define the sets of test functions as follows:

1. $C_b(\mathbb{C}) := \{f : \mathbb{C} \mapsto \mathbb{C} \mid f \text{ continuous and bounded}\}$

2. $C_0(\mathbb{C}) := \{f : \mathbb{C} \mapsto \mathbb{C} \mid f \text{ continuous and vanishes at infinity}\}$

3. $C_c(\mathbb{C}) := \{f : \mathbb{C} \mapsto \mathbb{C} \mid f \text{ continuous with compact support}\}$

Definition 2.2.1. *Let $\{\mu_N\}_{N=1}^{\infty}$ be a sequence of probability measures in $\mathcal{M}_1(\mathbb{C})$.*

- *The sequence $\{\mu_N\}_{N=1}^{\infty}$ is said to converge weakly to an element $\mu \in \mathcal{M}_1(\mathbb{C})$, if*

$$\forall f \in C_b(\mathbb{C}) : \lim_{N \to \infty} \langle \mu_N, f \rangle = \langle \mu, f \rangle$$

- *The sequence $\{\mu_N\}_{N=1}^{\infty}$ is said to converge vaguely to an element $\mu \in \mathcal{M}_{\leq 1}(\mathbb{C})$, if*

$$\forall f \in C_c(\mathbb{C}) : \lim_{N \to \infty} \langle \mu_N, f \rangle = \langle \mu, f \rangle$$

Some remarks about the definition are that weak convergence of measures implies vague convergence, what's more weak and vague limits are unique. The target measures of weak convergence of probability measures are again probability measures. This is because $1_{\mathbb{C}} \in C_b(\mathbb{C})$ and so we must have that $\mu(\mathbb{C}) = 1$. The target measures of vague convergence are exactly sub-probability measures, see for example [39], page 143.

We will transform $\mathcal{M}_1(\mathbb{C})$ into a metric space by defining an appropriate distance function on it which metrizes the weak convergence of measures. We will define it according to test functions $f \in \mathbb{C}_b$. We then have that

- The following metric:

$$d(\mu, \nu) := \sup \left\{ \left| \int_{\mathbb{C}} f \, d\mu - \int_{\mathbb{C}} f \, d\nu \right| : \|f\|_{\infty} \leq 1 \right\}$$

metrizes weak convergence of measures, that is $d(\mu_N, \mu) \to 0 \Leftrightarrow \mu_N \to \mu$ weakly.

- Equipped with this metric $(\mathcal{M}_1(\mathbb{C}), d)$ becomes a separable but not complete metric space.

The metric above is called the *Prokhorov metric* which can be found for example in [17], pages 394-395.

Since $\mathcal{M}_1(\mathbb{C})$ is a metric space with a corresponding Borel algebra it is also a measurable space, so we can then study $\mathcal{M}_1(\mathbb{C})-$valued random variables which we will call random measures.

Definition 2.2.2 (Random probability measure). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A random probability measure on $(\mathbb{C}, \mathcal{B}_\mathbb{C})$ is a measurable map $\mu : \Omega \mapsto \mathcal{M}_1(\mathbb{C})$.*

We remark that if μ is a random probability measure $\mu : \Omega \mapsto \mathcal{M}_1(\mathbb{C})$ and $f : \mathbb{C} \mapsto \mathbb{C}$ is a measurable bounded function, then the map $\omega \mapsto \langle \mu(\omega), f \rangle$ is also measurable and bounded by $\|f\|_\infty$.

Based on a random probability measure, we can also define its "expected" deterministic measure. If μ is a random probability measure $\mu : \Omega \mapsto \mathcal{M}_1(\mathbb{C})$ and $B \in \mathcal{B}$ then $\mu(\omega, B)$ is a random variable, which means that we can consider its expectation $\mathbb{E}\mu(B)$ as the expected mass that μ would prescribe to the set B. The map $B \mapsto \mathbb{E}\mu(B)$ is then the deterministic "expected" measure on $(\mathbb{C}, \mathcal{B}_\mathbb{C})$.

Definition 2.2.3 (Expected measure). *Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and μ a random probability measure on $(\mathbb{C}, \mathcal{B})$. Then the map:*

$$\overline{\mu} : \mathcal{B} \mapsto [0, 1]$$

with

$$\overline{\mu}(B) = \int_\Omega \mu(\omega, B) d\mathbb{P}(\omega) = \mathbb{E}\mu(B)$$

is an element of $\mathcal{M}_1(\mathbb{C})$ and is called the expected measure of μ.

We will define three notions of convergence of random probability measures on $(\mathbb{C}, \mathcal{B}_\mathbb{C})$, weak convergence in expectation, weak convergence in probability and weak convergence almost surely.

Definition 2.2.4 (Weak convergence in expectation). *Let $(\mu_n)_{n=1}^\infty$ be a sequence of random*

probability measures on $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ and μ a random probability measure on the same space. We say that $(\mu_N)_{n=1}^{\infty}$ converges weakly in expectation to μ, if the sequence of expected measures $(\mathbb{E}\mu_n)_{n=1}^{\infty}$ converges weakly to the expected measure $\mathbb{E}\mu$.

Definition 2.2.5 (Weak convergence in probability). *We say that a sequence of random measures $(\mu_N)_{N=1}^{\infty}$ converges weakly in probability to μ, if $\langle \mu_N, f \rangle$ converges to $\langle \mu, f \rangle$ in probability, for all $f \in C_b$.*

Definition 2.2.6 (Almost sure weak convergence). *We say that a sequence of random measures $(\mu_N)_{N=1}^{\infty}$ converges weakly to μ almost surely, if $\langle \mu_N, f \rangle$ converges to $\langle \mu, f \rangle$ almost surely, for all $f \in C_b$.*

We recall here that the definition of $\langle \cdot, \cdot \rangle$ contains a random measure integral which makes it a random variable. So we can use the definitions of almost sure convergence and convergence of probability for random variables. See for example [17], pages 516-517.

2.3 The empirical spectral distribution

Let X be an $N \times N$ random matrix on $(\Omega, \mathcal{A}, \mathbb{P})$, then the *empirical spectral distribution* (*ESD*) σ_N of X is the random probability measure on $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ given by:

$$\sigma_N : \Omega \times \mathcal{B}_{\mathbb{C}} \mapsto [0, 1]$$

$$(\omega, B) \mapsto \sigma_N(\omega, B) = \frac{1}{N} \sum_{k=1}^{N} \delta_{\lambda_{X(\omega)}^k}(B),$$

where $\delta_{\lambda_X^k}(\cdot)$ is the Dirac measure for the $k-$th eigenvalue of the random matrix X, i.e.

$$\delta_{\lambda_X^k}(B) = \begin{cases} 1, \text{ if } \lambda_X^k \in B \\ 0, \text{ if } \lambda_X^k \notin B \end{cases}$$

Note that σ_N is indeed a random probability measure on $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$. This is because $\sigma_N(\omega)$

is a convex combination of probability measures and thus again a probability measure. On the other hand, if $B \in \mathcal{B}_\mathbb{C}$ is arbitrary, then $\sigma_N(B)$ is clearly measurable, making it a random measure.

For any measurable set $B \subset \mathbb{C}$, $\sigma_N(B)$ gives the proportion of the N eigenvalues that fall into the set B. Thus σ_N carries information about the location of the eigenvalues and we are particularly interested in the case $N \to \infty$.

Chapter 3

The Stieltjes transform method

In order to analyze properties of the empirical spectral distribution, it is useful to use a tool that relates this distribution with entries of the random matrix or another matrix related to it (see section 4.1). For this, we can use a suitable transform of the measure.

Definition 3.0.1 (Stieltjes transform). *Let μ be a finite measure on some measurable linear space Y. The Stieltjes transform S_μ of μ is the map:*

$$S_\mu : Y \setminus \mathrm{supp}(\mu) \mapsto Y$$

$$y \mapsto \int_{\mathrm{supp}(\mu)} \frac{1}{w-y} d\mu(w)$$

The Stieltjes transform is defined via an integral involving our measure of interest.

3.1 Distributions on the real line

In the case that the measure is supported on the real line, we define the Stieltjes transform map as:

$$S_\mu : \mathbb{C} \setminus \mathrm{supp}(\mu) \mapsto \mathbb{C}$$

$$z \mapsto \int_{\mathbb{R}} \frac{1}{x - z} d\mu(x)$$

The Stieltjes transform is now defined via an integral of a complex function with respect to our measure of interest. We remark some properties of the Stieltjes transform:

- $\text{Im}(z) \geq 0 \Leftrightarrow \text{Im} S_\mu(z) \geq 0$.

- $S_\mu(\overline{z}) = \overline{S_\mu(z)}$.

- S_μ is uniquely determined by its restriction $S_\mu : \mathbb{C}_+ \to \mathbb{C}_+$.

- $|S_\mu(z)| \leq \frac{1}{|\text{Im}(z)|}$.

- S_μ is holomorphic and in particular, S_μ is continuous and can be represented by a power series around $z_0 \in \mathbb{C} \setminus \text{supp}(\mu)$ and is infinitely differentiable.

- The derivatives of the Stieltjes transform are given by

$$S_\mu^{(k)}(z) = \int_{\text{supp}(\mu)} \frac{k!}{(w - z)^{k+1}} d\mu(w),$$

  where $S_\mu^{(k)}$ denotes the k-th derivative of S_μ.

These properties can be found for example in [61], pages 169-171. We now remark an important measure-retrieval formula.

Proposition 3.1.1 (Retrieval of measure). *For any bounded interval $I \subset \mathbb{R}$ with endpoints $a < b$, we have the following formula:*

$$\mu(a, b) + \frac{1}{2}(\mu(\{a\}) + \mu(\{b\})) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_I \text{Im} S_\mu(E + i\eta) dE, \qquad (3.1)$$

Proof. Let I be an interval with endpoints $a < b$ and $\eta > 0$. By Fubini's theorem we get that:

$$\frac{1}{\pi} \int_I \text{Im} \, S_\mu(E + i\eta) dE = \frac{1}{\pi} \int_I \int_{\mathbb{R}} \frac{\eta}{(x - E)^2 + \eta^2} d\mu(x) dE$$

$$= \frac{1}{\pi} \int_{\mathbb{R}} \int_I \frac{\eta}{(x - E)^2 + \eta^2} dE d\mu(x)$$

$$= \frac{1}{\pi} \int_{\mathbb{R}} \int_a^b \frac{\eta}{(x - E)^2 + \eta^2} \, dE d\mu(x)$$

Now since

$$\int_a^b \frac{\eta}{(x-E)^2 + \eta^2} \, \mathrm{d}E = \frac{1}{\eta} \int_a^b \frac{1}{\left(\frac{E-x}{\eta}\right)^2 + 1} \, \mathrm{d}E$$

$$= \int_{\frac{a-x}{\eta}}^{\frac{b-x}{\eta}} \frac{1}{E^2 + 1} \, \mathrm{d}E$$

$$= \arctan\left(\frac{b-x}{\eta}\right) - \arctan\left(\frac{a-x}{\eta}\right),$$

and $\arctan : \mathbb{R} \mapsto \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is strictly increasing with $\lim\limits_{x\to\pm\infty} \arctan(x) = \pm\frac{\pi}{2}$, we obtain

$$\lim_{\eta\to0^+} \left[\arctan\left(\frac{b-x}{\eta}\right) - \arctan\left(\frac{a-x}{\eta}\right)\right] = \begin{cases} \pi, & \text{if } x \in (a,b) \\ 0, & \text{if } x \notin [a,b] \\ \frac{\pi}{2}, & \text{if } x = a \text{ or } x = b. \end{cases}$$

Thus, by dominated convergence we find

$$\lim_{\eta\to0^+} \frac{1}{\pi} \int_I \operatorname{Im} S_\mu(E + i\eta) \mathrm{d}E = \lim_{\eta\to0^+} \frac{1}{\pi} \int_{\mathbb{R}} \arctan\left(\frac{b-x}{\eta}\right) - \arctan\left(\frac{a-x}{\eta}\right) \mathrm{d}\mu(x)$$

$$= \int_{\mathbb{R}} \mathbf{1}_{(a,b)}(x) + \frac{1}{2}\mathbf{1}_{\{a,b\}}(x)\mathrm{d}\mu(x)$$

$$= \mu((a,b)) + \frac{1}{2}(\mu(\{a\}) + \mu(\{b\})).$$

$\square$

Corollary 3.1.2. *For any bounded interval $I \subset \mathbb{R}$ with $\mu(\partial I) = 0$, we find that:*

$$\mu(I) = \lim_{\eta\to0^+} \frac{1}{\pi} \int_I \operatorname{Im} S_\mu(E + i\eta) dE$$

The previous proposition and corollary suggest that any finite measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is uniquely determined by S_μ or equivalently the map $\mu \mapsto S_\mu$ is injective.

3.2   Distributions on the complex plane

In the case that the measure is supported on the complex plane, we define the Stieltjes transform map as:

$$S_\mu : \mathbb{C} \setminus \mathrm{supp}(\mu) \mapsto \mathbb{C}$$

$$z \mapsto \int_{\mathbb{C}} \frac{1}{w - z} d\mu(w)$$

We remark that all the previous properties of the Stieltjes transform in the real line still hold in the complex plane. Specifically, we have that:

- $\mathrm{Im}(z) \geq 0 \Leftrightarrow \mathrm{Im}S_\mu(z) \geq 0$.

- $S_\mu(\bar{z}) = \overline{S_\mu(z)}$.

- S_μ is uniquely determined by its restriction $S_\mu : \mathbb{C}_+ \to \mathbb{C}_+$.

- $|S_\mu(z)| \leq \frac{1}{|\mathrm{Im}(z)|}$.

- S_μ is holomorphic and in particular, S_μ is continuous and can be represented by a power series around $z_0 \in \mathbb{C} \setminus \mathrm{supp}(\mu)$ and is infinitely differentiable.

- The derivatives of the Stieltjes transform are given by

$$S_\mu^{(k)}(z) = \int_{\mathrm{supp}(\mu)} \frac{k!}{(w - z)^{k+1}} d\mu(w),$$

  where $S_\mu^{(k)}$ denotes the k-th derivative of S_μ.

The proofs are similar to the ones for the real-case distributions, see [61], pages 169-171.

The measure-retrieval formula is different in the complex case. We have the following proposition:

Proposition 3.2.1 (Retrieval of measure). *Let μ be a finite measure on $(\mathbb{C}, \mathcal{B}_\mathbb{C})$ with corresponding Stieltjes transform S_μ, such that $d\mu = f(z)d^2z$, we then get that:*

$$f(z) = \frac{1}{\pi}\partial_{\bar{z}}S_\mu(z). \tag{3.2}$$

Some remarks about this proposition are the following.

Here, d^2z denotes the Lebesgue measure applied on the complex domain, while $f : \mathbb{C} \mapsto [0, 1]$ is a complex density function, i.e. $f(z) = f_{\mathrm{Re}(Z),\mathrm{Im}(Z)}(\mathrm{Re}(z), \mathrm{Im}(z))$ is the joint density of the real and imaginary part of a complex random variable Z evaluated at the point $z \in \mathbb{C}$.

The symbol $\partial_{\bar{z}}$ denotes the *Wirtinger* derivative of a complex function $g(z)$, i.e.

$$\partial_{\bar{z}}g = \frac{\partial g}{\partial \bar{z}} := \frac{1}{2}\left(\frac{\partial g}{\partial x} + i\frac{\partial g}{\partial y}\right), \quad \text{whenever } z = x + iy.$$

For the proof, see section 3.5.

3.3   Stieltjes transform and weak convergence of measures

Since S_μ carries all the information about μ, S_μ can also be used to analyze weak convergence of probability measures.

Theorem 3.3.1 (Convergence theorem). *Let $(Y, \mathcal{B}_Y)$ be a measurable linear normed topological space (either $\mathbb{C}$ or $\mathbb{R}$) with corresponding Borel algebra $\mathcal{B}_Y$ and let μ be a probability measure on that space. If $Z \subset Y \setminus \mathrm{supp}(\mu)$ is a set that has at least one accumulation point in $Y \setminus \mathrm{supp}(\mu)$, then we get the following statement:*

For a sequences of measures $(\mu_n)_{n=1}^\infty \in \mathcal{M}_1(Y)$ and a measure $\mu \in \mathcal{M}_1(Y)$, we have that

$$\mu_n \to \mu \text{ weakly} \Leftrightarrow \forall y \in Z : S_{\mu_n}(y) \to S_\mu(y),$$

where we remind that $S_\mu(y) = \int_{\mathrm{supp}(\mu)} \frac{1}{w-y}d\mu(w)$.

In order to prove theorem 3.3.1 we need the following three Lemmas:

Lemma 3.3.2 (Helly's selection theorem). *Let $(\mu_n)_{n=1}^\infty$ be a sequence in $\mathcal{M}_1(Y)$, then there exists a subsequence $(\mu_{n_k})_{k=1}^\infty$ and a sub-probability measure $\mu \in \mathcal{M}_{\leq 1}(Y)$ such that $\mu_{n_k} \to \mu$ vaguely.*

Proof. The Lemma and its proof can be found for example in [18], page 108. □

Lemma 3.3.3. *Let $(\mu_n)_{n=1}^\infty \in \mathcal{M}_1(Y)$ and $\mu \in \mathcal{M}_{\leq 1}(Y)$. Then $(\mu_n)_{n=1}^\infty$ converges weakly (respectively vaguely) to μ if every subsequence $(\mu_n)_{n \in J}, J \subset \mathbb{N}$ has a further subsequence $(\mu_n)_{n \in I}, I \subset J$ that converges weakly (respectively vaguely) to μ.*

Proof. We prove this by contradiction. Suppose that $(\mu_n)_{n=1}^\infty$ doesn't converge weakly (respectively vaguely) to μ. Then we can find a continuous and bounded (respectively with compact support) function $f : Y \mapsto Y$ and an $\epsilon > 0$ such that $|\langle \mu_n, f \rangle - \langle \mu, f \rangle| \geq \epsilon$, for all n in an infinite set $J \subset \mathbb{N}$. But now we can find a subsequence $(\mu_n)_{n \in I}, I \subset J$ that converges weakly (respectively vaguely) to μ. This means that we can find $n \in I \subset J$ such that $|\langle \mu_n, f \rangle - \langle \mu, f \rangle| < \epsilon$, which is a contradiction. □

Lemma 3.3.4. *If we have a sequence of probability measures that converges vaguely to a probability measure on $\mathcal{M}_1(Y)$, then vague convergence can be strengthened to weak convergence.*

Proof. For this Lemma and its proof, check for example [16], page 93. □

We now continue with the proof of our main theorem in this section.

Proof of Theorem 3.3.1. The " $\Rightarrow$ " is obvious since the function $y \mapsto \frac{1}{w-y}$ is continuous and bounded. To show " $\Leftarrow$ ", we will use Lemma 3.3.3. Let $(\mu_n)_{n \in J}, J \subset \mathbb{N}$ be a subsequence of $(\mu_n)_{n=1}^\infty$. Then, by Lemma 3.3.2 there exists a further subsequence $(\mu_n)_{n \in I}, I \subset J$ such that $(\mu_n)_{n \in I} \to \nu$ vaguely for some $\nu \in \mathcal{M}_{\leq 1}(Y)$. Since $y \mapsto \frac{1}{w-y}$ vanishes at infinity, it follows that $S_{(\mu_n)_{n \in I}}(y) \to S_\nu(y)$ and therefore $S_\nu(y) = S_\mu(y)$, for all $y \in Z$. Since the functions are holomorphic, we establish that $S_\mu = S_\nu$. By the retrieval of measure identities, (3.1) and (3.2) we conclude that $\mu = \nu$. For any

other subsequence $(\mu_n)_{n \in J}$, $J \subset \mathbb{N}$ we get by the same arguments, that every further subsequence should converge to μ vaguely. Therefore, $\mu_n \to \mu$, vaguely. Because all measures involved are probability measures, by using Lemma 3.3.4, we conclude that $\mu_n \to \mu$ weakly, as we want. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proposition 3.3.5 (Convergence theorem for random measures)**.** *Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of random probability measures on $\mathcal{M}_1(Y)$ and μ a deterministic probability measure on the same space. Then we have the following equivalences:*

- *$\mu_n \to \mu$ weakly in expectation $\Leftrightarrow \mathbb{E}S_{\mu_n}(y) \to S_\mu(y)$ for all $y \in Z$.*

- *$\mu_n \to \mu$ weakly in probability $\Leftrightarrow S_{\mu_n}(y) \to S_\mu(y)$ in probability for all $y \in Z$.*

- *$\mu_n \to \mu$ weakly almost surely $\Leftrightarrow S_{\mu_n}(y) \to S_\mu(y)$ almost surely for all $y \in Z$.*

Proof. The first statement comes from our main theorem in this section, considering that:

$$\mathbb{E}S_{\mu_n}(y) = \mathbb{E} \int_{\text{supp}(\mu)} \frac{1}{w-y} d\mu(w) \overset{(*)}{=} \int_{\text{supp}(\mu)} \frac{1}{w-y} d\mathbb{E}\mu(w) = S_{\mathbb{E}\mu_n}(y),$$

where $d\mathbb{E}\mu(w)$ denotes the "expected measure" as defined in 2.2.3. The $(*)$ equality is justified because it holds generally that:

$$\mathbb{E} \int f d\mu = \int f d\mathbb{E}\mu,$$

where μ is a random probability measure, $\mathbb{E}d\mu$ its expected measure and $f : Y \mapsto Y$ any bounded measurable function. This can be found in [40], page 53.

For the second statement, the " $\Rightarrow$ " direction is obvious since the function $y \mapsto \frac{1}{y-w}$ is continuous and bounded. The proof of the " $\Leftarrow$ " direction can be found in [8], page 45.

For the third statement we work as follows. If $\mu_n \to \mu$ weakly in a measurable set $A \subset \Omega$ such that $\mathbb{P}(A) = 1$, then on A we also have that $S_{\mu_n}(y) \to S_\mu(y)$ for all $y \in Z$, by our main theorem. This proves the " $\Rightarrow$ " direction. For the " $\Leftarrow$ " direction, we fix

a sequence $(z_k)_{k=1}^\infty$ in Z that converges to some $z \in Y \setminus \text{supp}(\mu)$. For each $k \in \mathbb{N}$ we find a measurable set A_k with $\mathbb{P}(A_k) = 1$ and $S_{\mu_n}(z_k) \to S_\mu(z_k)$. Then $A := \bigcap_{k \in \mathbb{N}} A_k$ is measurable with $\mathbb{P}(A) = 1$ and for all $y \in \{z_k : k \in \mathbb{N}\}$ we have $S_{\mu_n}(y) \to S_\mu(y)$. Since this set has an accumulation point, we find by our main theorem that $\mu_n \to \mu$ weakly on A, as we want. $\qquad\square$

3.4 The imaginary part of the Stieltjes transform

In this section we analyze further the retrieval of measure identity from Corollary 3.1.2 for measures on the real line. According to this identity, the function $E \mapsto \frac{1}{\pi}\text{Im}S_\mu(E+i\eta)$ should represent a density that approximates μ well as $\eta \to 0^+$.

To analyze this property we introduce the concept of the convolution of probability measures as well as the kernel density estimators.

Definition 3.4.1 (Convolution). *Let μ and ν be two probability measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and f, g be two probability density functions.*

- *(convolution of measures). The convolution of the two measures is defined as $\mu * \nu := (\mu \otimes \nu)^+$, where $\mu \otimes \nu$ is the product measure and the addition in the exponent represents the push-forward of the product measure over the addition map.*

- *(convolution of density with measure). The convolution of the density f and the probability measure ν is defined as the function $f * \nu : \mathbb{R} \to \mathbb{R}$ with:*

$$(f * \nu)(x) := \int_\mathbb{R} f(x - y)d\nu(y)$$

- *(convolution of densities). The convolution of the densities f and g is the function $f * g : \mathbb{R} \to \mathbb{R}$ with:*

$$(f * g)(x) := \int_\mathbb{R} f(x - y)g(y)dy$$

Some remarks about the definitions are the following. For the first definition, for an

arbitrary set $B \in \mathcal{B}_{\mathbb{R}}$, we have that:

$$(\mu * \nu)(B) = (\mu \otimes \nu)(\{(x, y) \in \mathbb{R}^2 : x + y \in B\}).$$

If $f : \mathbb{R} \to \mathbb{R}$ is $\mu \otimes \nu$-integrable, we then obtain that:

$$\int_{\mathbb{R}} f d(\mu * \nu) = \int_{\mathbb{R}^2} (f \circ +) d(\mu \otimes \nu) = \int_{\mathbb{R}^2} f(x+y) d(\mu \otimes \nu)(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x+y) d\mu(x) d\nu(y).$$

Particularly, for an indicator function $f = \mathbf{1}_B$ for some $B \in \mathcal{B}_{\mathbb{R}}$ we get that:

$$(\mu * \nu)(B) = \int_{\mathbb{R}} \mathbf{1}_B d(\mu * \nu) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_B(x + y) d\mu(x) d\nu(y) = \int_{\mathbb{R}} \mu(B - y) d\nu(y).$$

The equality of the first with the third term shows that the convolution function is commutative in the space of probability measures. We denote by $f dx$ the probability measure with density f, where f is a probability density function. We then have the following properties of the convolution:

- The convolution is a commutative binary operation in the space of probability measures where the neutral element is given by the Dirac probability measure δ_0.

- The function $f * \nu$ is a probability density for the convolution measure $(f dx * \nu)$, that is $f dx * \nu = (f * \nu) dx$.

- The function $f * g$ is a probability density for the convolution measure $(f dx * g dx)$, that is $f dx * g dx = (f * g) dx$, where g is another probability density function.

- The convolution of probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is continuous with respect to weak convergence of measures. That is, if μ, ν, $(\mu_n)_{n=1}^{\infty}$, $(\nu_n)_{n=1}^{\infty}$ are probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ with $\mu_n \to \mu$ weakly and $\nu_n \to \nu$ weakly, then $\mu_n * \nu_n \to \mu * \nu$ weakly.

We introduce here the kernel density estimators and particularly focus on the Cauchy kernel.

Definition 3.4.2 (Cauchy kernel). *For all $\eta > 0$, we define the Cauchy kernel as the function $P_\eta : \mathbb{R} \to \mathbb{R}$ with*

$$P_\eta(x) = \frac{1}{\pi} \frac{\eta}{x^2 + \eta^2},$$

which is the density function of the Cauchy distribution with scale parameter η.

For the Cauchy kernel, the following important identity holds:

$$\lim_{\eta \to 0^+} (P_\eta dx) = \delta_0,$$

where δ_0 denotes the Dirac measure centered at 0. That is, for every $B \in \mathcal{B}_\mathbb{R}$,

$$\delta_0(B) = \begin{cases} 1 \text{ , if } 0 \in B \\ 0 \text{ , if } 0 \notin B. \end{cases}$$

This is because the characteristic function of the Cauchy distribution with parameter η is given by $t \mapsto e^{-\eta|t|}$ and so letting $\eta \to 0^+$ yields the identity.

For the imaginary part of the Stieltjes transform we have the following fundamental property for any real bounded measure μ:

$$\frac{1}{\pi} \text{Im} S_\mu(E + i\eta) = \int_\mathbb{R} \frac{1}{\pi} \frac{\eta}{(E - x)^2 + \eta^2} d\mu(x) = (P_\eta * \mu)(E)$$

This property means that the function $\frac{1}{\pi}\text{Im}S_\mu(\cdot + i\eta)$ is the convolution of the kernel density P_η with μ and thus a density for the measure $(P_\eta dx) * \mu$. As $\eta \to 0^+$, we have that:

$$\frac{1}{\pi} \text{Im} S_\mu(\cdot + i\eta) dx = (P_\eta dx) * \mu \to \delta_0 * \mu = \mu.$$

Assume now that $\{\sigma_n\}_{n=1}^\infty$ is a sequence of empirical spectral distributions of *Hermitian* random matrices so that σ_n converges to a *real* deterministic measure σ. We now take the convolution of σ_n with the Cauchy kernel P_η. Since $\sigma_n \to \sigma$, we get by the previous analysis that $(P_{\eta_n} * \sigma_n)dx \to \sigma$, for any sequence $\eta_n \to 0$. However, if σ has some density $\sigma = f_\sigma dx$, this is not enough to deduce that $P_{\eta_n} * \sigma_n \to f_\sigma$ for example in the

supremum norm. This would allow *local estimation* by σ_n through the Cauchy kernel about the density f_σ. If $\eta = \eta_n$ drops too quickly to zero as $n \to \infty$ then $(P_{\eta_n} * \sigma_n)$ will have steep peaks at each eigenvalue and thus will not approximate the density of σ. This phenomenon is typical for kernel density estimators. We introduce them next in their generality.

Definition 3.4.3 (Kernel). *A kernel K is a real probability density function. If K is a kernel and $h > 0$ a real parameter, we define K_h as the kernel with $K_h(x) = \frac{1}{h} K(\frac{x}{h})$ for all $x \in \mathbb{R}$ and call K_h "the kernel K at bandwidth h."*

It is clear that K_h is a kernel if $h > 0$ and K is a kernel. An example of a kernel is the previously defined Cauchy kernel P_η. We have for all $x \in \mathbb{R}$ and $\eta > 0$:

$$P(x) = \frac{1}{\pi} \frac{1}{x^2 + 1} \ \text{ and } \ P_\eta(x) = \frac{1}{\pi \eta} \frac{1}{\left(\frac{x}{\eta}\right)^2 + 1} = \frac{1}{\pi} \frac{\eta}{x^2 + \eta^2}.$$

We are interested in constructing a density function f that describes the experiment of drawing at random from the real-valued observations $u = (u_1, ..., u_N)$, in other words that approximates the empirical probability measure:

$$f \, dx \approx \rho_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{u_i}$$

This can be done with the help of a kernel K_h, which is usually chosen to be unimodal and symmetric around zero, see for example [52].

Definition 3.4.4 (Kernel density estimator). *The kernel density estimator with kernel K and bandwidth $h > 0$ for an empirical measure ρ_N is the density function given by the convolution $K_h * \rho_N$.*

We therefore have the density function estimator $K_h * \rho_N : \mathbb{R} \to \mathbb{R}$ with

$$x \mapsto (K_h * \rho_N)(x) = \frac{1}{N} \sum_{i=1}^{N} K_h(x - u_i) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - u_i}{h}\right).$$

The center of the kernel is placed upon each observation, whose influence is smoothed

out over its neighbourhood. The size of this neighborhood is governed by the bandwidth. A small bandwidth will retain the probability mass of $1/N$ to be closer to its observation, whereas a larger bandwidth will result in a wider spread of probability mass. Therefore, a smaller bandwidth will result in a peaky density function (with steep peaks at each observation), whereas a larger one will result in a smoother density function.

Assume now we are given an empirical spectral distribution ρ_N which comes from a Hermitian $N \times N$ matrix X_N. The kernel density estimator at the location $E \in \mathbb{R}$ for ρ_N with Cauchy kernel P at bandwidth $\eta > 0$ is then given by:

$$
\begin{aligned}
(P_\eta * \rho_N)(E) &= \frac{1}{N\eta} \sum_{i=1}^{N} \frac{1}{\pi} \frac{1}{\left(\frac{E - \lambda_i^{X_N}}{\eta}\right)^2 + 1} \\
&= \frac{1}{\pi N} \sum_{i=1}^{N} \frac{\eta}{(E - \lambda_i^{X_N})^2 + \eta^2} \\
&= \frac{1}{\pi} \mathrm{Im} S_{\rho_N}(E + i\eta).
\end{aligned}
$$

This gives the imaginary part of the Stieltjes transform the role of the kernel density estimator, see for example [10] pages 7-8. Choosing different values of η we get different estimations of the empirical spectral distribution. For a very small parameter of η, we may not obtain a useful approximation of the density, whereas for a bigger one the estimation would be better. In the next chapter, we will find the optimal value of η so that the estimation is still useful for the approximation of the empirical eigenvalue density of the Marchenko-Pastur law for covariance matrices.

3.5   The Wirtinger derivative of the Stieltjes transform

In this section we analyze further the retrieval of measure identity from Corollary 3.2.1 for measures on the complex plane. According to this identity, we have a density given by the function $z \mapsto \frac{1}{\pi} \partial_{\overline{z}} S_\mu(z)$ which matches the density of μ on the complex plane.

We will analyze the generalized Cauchy theorem as well as convolutions with the Cauchy kernel, see for example [43], pages 109-110.

Theorem 3.5.1 (Generalized Cauchy theorem). *Let D be a disk on $\mathbb{C}$ and f a complex-valued C^1 function on the closure of D. Then:*

$$\forall \zeta \in D : f(\zeta) = \frac{1}{2\pi i} \int_{\partial D} \frac{f(z)}{z - \zeta} dz - \frac{1}{\pi} \iint_D \frac{\partial}{\partial \overline{z}} \frac{f(z)}{z - \zeta} d^2 z.$$

Notice that if f is complex-differentiable (holomorphic) then the identity reduces to the well-known Cauchy formula. By saying that f is a C^1 function we just mean that the derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist and they are continuous functions.

Notice also that if we pick a function $f \in C_c(\mathbb{C})$ with compact support, then the identity reduces to:

$$\forall \zeta \in \text{supp}(f) : f(\zeta) = -\frac{1}{\pi} \iint_{\mathbb{C}} \frac{\partial}{\partial \overline{z}} \frac{f(z)}{z - \zeta} d^2 z = \frac{1}{2\pi i} \iint_{\mathbb{C}} \frac{\partial}{\partial \overline{z}} \frac{f(z)}{z - \zeta} dz \wedge d\overline{z},$$

where we defined the *wedge product* as $dz \wedge d\overline{z} := -2i \, d^2 z$. This is a *complex measure*, the space of which is a linear space over the complex numbers and also a Banach space, see [57], pages 116-119.

We now turn our attention to convolutions of complex functions and measures on the complex plane.

Definition 3.5.1 (Convolution of complex measures and densities). *Let μ and ν be two probability measures on $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ and f, g two complex probability density functions. We can define again the following convolutions:*

- *(convolution of measures). The convolution of the two complex measures is defined as in the real case in which we recall that we defined them in a general setting.*

- *(convolution of density with measure). The convolution of the complex density f and probability measure ν is defined as the complex function $f * \nu : \mathbb{C} \to \mathbb{C}$ with:*

$$\forall z \in \mathbb{C} : (f * \nu)(z) := \int_{\mathbb{C}} f(z - w) d\nu(w).$$

- *(convolution of densities). The convolution of the complex densities f and g is the function $f * g : \mathbb{C} \to \mathbb{C}$ with:*

$$\forall z \in \mathbb{C} : (f * g)(z) := \int_{\mathbb{C}} f(z - w) g(w) dw.$$

The same remarks that hold for the real measures and densities also hold for the complex ones.

We introduce now the Cauchy kernel in complex analysis, see for example [43], page 10.

Definition 3.5.2 (Cauchy kernel). *We define the complex Cauchy kernel as the function $k : \mathbb{C} \setminus \{0\} \to \mathbb{C}$ with:*

$$k(z) := \frac{1}{2\pi i} \frac{1}{z}.$$

For the Stieltjes transform of a complex measure μ we now have the following expression for it as a convolution with the complex Cauchy kernel:

$$S_\mu(z) = \int_{\mathbb{C}} \frac{1}{w - z} d\mu(w) = -2\pi i \int_{\mathbb{C}} k(z - w) d\mu(w) = -2\pi i \cdot (k * \mu)(z).$$

Suppose now that the measure μ has a complex density f such that $d\mu = f(z) \, d^2 z$. Then through the Stieltjes transform of μ we can retrieve the density f by the Cauchy-Riemman equation:

$$\frac{1}{\pi} \frac{\partial}{\partial \bar{z}} (S_\mu) = f.$$

In a more general setting, we have that the convolution of a measure μ with complex

density $\phi(z,\overline{z})$ such that $d\mu = \phi(z,\overline{z})\,dz \wedge d\overline{z}$ with the Cauchy kernel solves the equation:

$$\frac{\partial}{\partial\overline{z}}(k*\mu) = \phi.$$

This means that we have the following retrieval of measure identity which we prove through the generalized Cauchy theorem:

$$d\mu = \left(\frac{1}{\pi}\frac{\partial}{\partial\overline{z}}S_\mu\right)d^2z.$$

Proof. We pick a complex test function f with compact support. By the analysis of the generalized Cauchy formula, we recall that:

$$\forall\zeta \in \mathrm{supp}(f) : f(\zeta) = \frac{1}{2\pi i}\iint_{\mathbb{C}}\frac{\partial}{\partial\overline{z}}\frac{f(z)}{z-\zeta}dz \wedge d\overline{z}.$$

Suppose now we have the C-R equation:

$$\frac{\partial}{\partial\overline{z}}(v) = g,$$

where g is a complex continuous function with compact support and v is an unknown complex function. Suppose that there exists a complex measure μ such that $d\mu = g(z,\overline{z})\,dz \wedge d\overline{z}$. The existence of a solution v for the C-R problem is established for example in [63], pages 35-36. We show that if such a solution exists then we must have $v = k*\mu$. Indeed, by the generalized Cauchy formula, we get that:

$$(k*\mu)(z,\overline{z}) = \frac{1}{2\pi i}\int_{\mathbb{C}}\frac{1}{w-z}d\mu(w) = \frac{1}{2\pi i}\int_{\mathbb{C}}\frac{g(w,\overline{w})}{w-z}dw \wedge d\overline{w}$$

$$= \frac{1}{2\pi i}\int_{\mathbb{C}}\frac{\partial}{\partial\overline{w}}\frac{v(w,\overline{w})}{w-z}dw \wedge d\overline{w} = v(z,\overline{z}). \qquad \square$$

In particular, for the Stieltjes transform, we get that:

$$\frac{1}{\pi}\frac{\partial}{\partial\overline{z}}S_\mu = -2i\frac{\partial}{\partial\overline{z}}(k*\mu) = -2i\frac{d\mu}{dz \wedge d\overline{z}} = \frac{d\mu}{d^2z}.$$

Assume now that $\{\sigma_n\}_{n=1}^{\infty}$ is a sequence of empirical spectral distributions of *non-Hermitian* random matrices, so that σ_n converges to a *complex* deterministic measure σ. If we take the convolution of σ_n with the complex Cauchy kernel k, then to show the convergence $\sigma_n \to \sigma$, it is enough to establish that $\sigma_n * k \to \sigma * k$.

Chapter 4

Local Marchenko-Pastur law

We let X be an $M \times N$ matrix with complex components $x_{ij} = \text{Re}(x_{ij}) + i\text{Im}(x_{ij})$, for $i = 1, ..., M$ and $j = 1, ..., N$. Assume that $\text{Re}(x_{ij})$ and $\text{Im}(x_{ij})$ are independent and identically distributed real random variables with mean zero and variance $\frac{1}{2}$ so that:

$$\mathbb{E}(x_{ij}) = 0 \quad \text{and} \quad \mathbb{E}|x_{ij}|^2 = 1, \ \text{for} \ i = 1, ..., M, \ j = 1, \dots, N. \tag{4.1}$$

We denote by X_N the scaled matrix:

$$X_N := X/\sqrt{N}. \tag{4.2}$$

In this chapter, we will analyze asymptotics of the empirical spectral measure (see section 2.3) of the matrix $X_N^* X_N$ for $N \to \infty$, when $M = N$. We define:

$$K_{XX} := X_N^* X_N \in \mathbb{C}^{N \times N} \tag{4.3}$$

For general M and N we can consider the collection of the complex random variables $\{x_{ij}\}$ as M *observations* of $\mathbb{C}^N$-valued normalized random variables, where we can consider N as the number of *features* of each observation. Hence we have the name *Sample Coviariance Matrix* for the random matrix $X_N^* X_N$, as its matrix-expectation gives the covariances of the sample of M observations of the $\mathbb{C}^N$-valued random variables.

The Sample Covariance Matrix K_{XX} is always Hermitian and positive semi-definite. This is because:

$$K_{XX}^* = (X_N^* X_N)^* = X_N^* X_N = K_{XX},$$

where we used the property $(AB)^* = B^* A^*$, valid for any complex matrices A, B such that their multiplication makes sense. We also have that:

$$\mathbf{x}^* K_{XX} \mathbf{x} = \mathbf{x}^* X_N^* X_N \mathbf{x} = (X_N \mathbf{x})^* X_N \mathbf{x} = \|X_N \mathbf{x}\|^2 \geq 0,$$

where we used the usual Euclidean norm for the vector $X_N \mathbf{x}$ and $\mathbf{x}$ is any vector in $\mathbb{C}^N$.

We remind that a matrix $A \in \mathbb{C}^{N \times N}$ is called *positive semi-definite* when

$$\forall \mathbf{x} \in \mathbb{C}^N : \mathbf{x}^* A \mathbf{x} \geq 0.$$

For a Hermitian matrix $H \in \mathbb{C}^{N \times N}$ the following relation holds:

$$H \text{ is positive semi-definite} \Leftrightarrow \text{All eigenvalues of } H \text{ are non-negative}.$$

We denote by λ_i, $i = 1, ... N$ the eigenvalues of the matrix K_{XX}. Since K_{XX} is Hermitian and positive semi-definite, we can assume that:

$$0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N.$$

We denote by ρ_N the empirical spectral cumulative density of these eigenvalues, according to the definition in section 2.3, applied to the Borel sets $B_x := (-\infty, x] \subset \mathbb{R}$:

$$\rho_N(x) := \frac{1}{N} \#\{i \leq N : \lambda_i \leq x\} \tag{4.4}$$

It was first established in 1967 in [44] that the limiting measure of $\rho_N dx$ is the following measure $\rho^d dx$ which is defined according to the parameter:

$$d := \lim_{M,N \to \infty} \frac{M}{N} \in (0, 1]. \tag{4.5}$$

Specifically, we have that $\rho_N dx \to \rho^d dx$ weakly almost surely, where:

$$\rho^d(x) := \frac{1}{2\pi d}\frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{x}, \quad x \in [\lambda_-, \lambda_+] \tag{4.6}$$

and $\lambda_- := (1 - \sqrt{d})^2, \quad \lambda_+ := (1 + \sqrt{d})^2$.

This has been called the Marchenko-Pastur distribution. In the case of a square matrix X in which the number of observations is equal to the number of features, we have that $M = N \Rightarrow d = 1$ and so the limiting density becomes:

$$\rho(x) = \begin{cases} \dfrac{1}{2\pi}\sqrt{\dfrac{4}{x} - 1}, & 0 < x \leq 4 \\ 0, & \text{otherwise} \end{cases} \tag{4.7}$$

This is the case when the limiting measure has a square root singularity near 0. We call this point the *hard edge* of the density, while the point 4 on the real line will be called the *soft edge* of the limiting density. The space between 0 and 4 will be called the *bulk* of the density, as there will be a "bulk amount" of eigenvalues there.

4.1 Resolvent identities for Sample Covariance Matrices

The matrix transformation that is closely related to the Stieltjes transform of the empirical spectral measure is called the *resolvent matrix*.

For a matrix X we define its resolvent transformation as the matrix:

$$G_z^X := (X - zI)^{-1},$$

where $z \in \mathbb{C}$ is a complex parameter related to the complex parameter of the Stieltjes transformation.

We have the following identity relating the Stieltjes transform of the empirical spectral

measure with the resolvent matrix:

$$S_N(z) := S_{\mu_N}(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mu_N(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} = \frac{1}{N} \text{Tr}(G_z^X), \qquad (4.8)$$

for any $z \in \mathbb{C} \setminus \mathbb{R}$, where μ_N is the empirical spectral measure of the matrix $X \in \mathbb{C}^{N \times N}$.

We recall from section 3.4 that the imaginary part of the Stieltjes transform of ρ_N can provide local information about the convergence $\rho_N dx \to \rho dx$. If we set $z = E + i\eta$, then $\frac{1}{\pi} \text{Im} S_{\rho_N}(E)$ will approximate the cumulative density $\rho_N(E)$ while the parameter η depending on N will model the scale of this convergence around each individual eigenvalue.

In this chapter we will prove this convergence in the optimal scale for η while also improving bounds for the rate of convergence of the corresponding Stieltjes transforms of the empirical spectral measure and the limiting Marchenko-Pastur measure:

$$S_N \to S_\rho.$$

This will provide us with some new results about the rate of convergence of the cumulative density ρ_N to the cumulative density provided by the measure ρ. This means that we will quantify the convergence:

$$\rho_N(E) \to \int_0^E \rho(x) dx, \qquad (4.9)$$

in a new way. We will also quantify the rigidity of each individual eigenvalue of the Sample Covariance Matrix in a new way, that is:

$$\lambda_i \approx \gamma_i, \qquad (4.10)$$

where γ_i is the $i-$th quantile of the Marchenko-Pastur distribution, i.e.

$$\int_0^{\gamma_i} \rho(x) dx = \frac{i}{N}. \qquad (4.11)$$

We collect now for the sake of the proofs some useful resolvent identities for the Sample Covariance Matrix. The resolvent transformation of the matrix K_{XX} will be denoted by G, omitting the dependence on z and K_{XX}.

Firstly, we remark some properties of the Stieltjes transform of the Marchenko-Pastur measure, see also [14] section 2.2.

Lemma 4.1.1. *Let S_ρ be the Stieltjes transfrom of the measure ρdx. We then have that:*

$$S_\rho(z) = -\frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{4}{z}}, \quad z \in \mathbb{C} \setminus [0, 4] \tag{4.12}$$

and the following algebraic identity:

$$S_\rho(z) = -\frac{1}{z(S_\rho(z) + 1)}. \tag{4.13}$$

We note that the square root is chosen so that $\mathrm{Im}\left(\sqrt{1 - 4/z}\right) \geq 0$. It is a simple calculus exercise to compute the integral in the definition of S_ρ to derive (4.12) and then derive the algebraic equation (4.13).

The algebraic identity (4.13) will be the basis of our proof. We will show that S_N satisfies a similar algebraic identity with high probability so that it has to be "close" to S_ρ.

<u>Resolvent identities</u>.

From now on X will denote the scaled version of our matrix. We define the two resolvent matrices G and $\mathcal{G}$ related to K_{XX} as:

$$G := (XX^* - zI_N)^{-1} \quad \text{and} \quad \mathcal{G} := (X^*X - zI_N)^{-1}. \tag{4.14}$$

For $j_1, j_2 \in \{0, 1, ..., N-1\}$ we denote by $X^{(j_1)}$ the submatrix of the $N \times N$ matrix X with the first j_1 columns removed. We denote by $X_{(j_2)}$ the submatrix of X with the first j_2 rows removed.

We now define the resolvents of the covariance of the "stripped" matrix $X_{(j_2)}^{(j_1)}$ as follows:

$$G_{(j_2)}^{(j_1)} := \left((X_{(j_2)}^{(j_1)})^* X_{(j_2)}^{(j_1)} - z \right)^{-1} \quad \text{and} \quad \mathcal{G}_{(j_2)}^{(j_1)} := \left(X_{(j_2)}^{(j_1)} (X_{(j_2)}^{(j_1)})^* - z \right)^{-1} \tag{4.15}$$

We denote by $G_{(j_2),kl}^{(j_1)}$ the (k,l) element of the resolvent matrix $G_{(j_2)}^{(j_1)}$. We remark that $(X^{(m)})^* X^{(m)}$ is the minor matrix of K_{XX} with the first m rows and the first m columns removed.

Lemma 4.1.2. *The following trace equality holds:*

$$\text{Tr}\left[\mathcal{G}_{(j_2)}^{(j_1)} \right] = \frac{j_1 - j_2}{z} + \text{Tr}\left[G_{(j_2)}^{(j_1)} \right]. \tag{4.16}$$

Proof. We have that $(X_{(j_2)}^{(j_1)})^* X_{(j_2)}^{(j_1)}$ and $X_{(j_2)}^{(j_1)} (X_{(j_2)}^{(j_1)})^*$ have the same non-zero eigenvalues and a total of $N - j_1$ eigenvalues for the first and a total of $N - j_2$ eigenvalues for the second. If $j_1 < j_2$ then the first matrix has $j_2 - j_1$ extra zero eigenvalues, while if $j_1 > j_2$ the second one has $j_1 - j_2$ extra zero eigenvalues. Hence, we have the identity. $\square$

We now define the following quantities

$$S_N(z)_{(j_2)}^{(j_1)} := \frac{1}{N} \text{Tr}\left[G_{(j_2)}^{(j_1)} \right] \quad \text{and} \quad \Lambda(z)_{(j_2)}^{(j_1)} := S_N(z)_{(j_2)}^{(j_1)} - S_\rho(z) \tag{4.17}$$

and we use the notation S_N and Λ when $j_1 = j_2 = 0$. We will generally omit denoting the dependence on the z-variable.

We now state some identities for resolvent entries.

We denote by $\left[G_{(j_2)}^{(j_1)} \right]^k$ the resolvent which occurs after an extra removal of the k−th column of the scaled matrix X, where $k \in \{ j_1 + 1, ..., N \}$.

Lemma 4.1.3 (Stripping lemma). *With $G_{(j_2)}^{(j_1)}$ and k as before, we have that for $i, j \neq k$:*

$$G_{(j_2),ij}^{(j_1)} = \left[G_{(j_2)}^{(j_1)} \right]_{ij}^k + \frac{G_{(j_2),ik}^{(j_1)} G_{(j_2),kj}^{(j_1)}}{G_{(j_2),kk}^{(j_1)}}. \tag{4.18}$$

Proof. For the proof we will use Woodbury's matrix identity:

$$(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

This identity can be found for example in [38] page 258.

We can assume that $j_1 = j_2 = 0$, otherwise the proof will be similar. We denote by M the matrix $X^*X - zI$. Let $\mathbf{e_k}$ be the standard $k-$th column unit basis vector and define the matrices:

$$M_{ij}^{[k]} := M_{ij}\mathbf{1}(i \neq k)\mathbf{1}(j \neq k), \quad U := \begin{pmatrix} M\mathbf{e}_k & \mathbf{e}_k \end{pmatrix}, \quad V := \begin{pmatrix} \mathbf{e}_k^* \\ \mathbf{e}_k^*M \end{pmatrix}$$

so that we have

$$M = M^{[k]} + UV - M_{kk}\mathbf{e}_k\mathbf{e}_k^*.$$

Then by the Woodbury matrix identity we have

$$\left(M^{[k]} - M_{kk}\mathbf{e}_k\mathbf{e}_k^*\right)^{-1} = G + GU(I - VGU)^{-1}VG.$$

A straightforward calculation yields

$$I - VGU = -\begin{pmatrix} 0 & G_{kk} \\ M_{kk} & 0 \end{pmatrix}, \quad (I - VGU)^{-1} = -\frac{1}{M_{kk}G_{kk}}\begin{pmatrix} 0 & G_{kk} \\ M_{kk} & 0 \end{pmatrix},$$

from which we deduce, after a short calculation,

$$\left(M^{[k]} - M_{kk}\mathbf{e}_k\mathbf{e}_k\right)^{-1} = G - \frac{1}{M_{kk}}\mathbf{e}_k\mathbf{e}_k^* - \frac{1}{G_{kk}}G\mathbf{e}_k\mathbf{e}_k^*G.$$

Using the formula for the block inversion of a matrix, we conclude

$$[G]_{ij}^k = G_{ij} - \frac{1}{G_{kk}}G\mathbf{e}_k\mathbf{e}_k^*G,$$

from which (4.18) follows. $\qquad\square$

Next, we have the following relationship between the (i, i) element of $|G|^2 := GG^*$ and $(\text{Im}G)_{ii}$, and the same holds for $G^{(j_1)}_{(j_2)}, \mathcal{G}^{(j_1)}_{(j_2)}$ instead of G:

Lemma 4.1.4 (Ward identity). *If $z = E + i\eta$, then*

$$|G(z)|^2_{ii} = \frac{(\text{Im}G(z))_{ii}}{\eta}. \tag{4.19}$$

Proof. We apply spectral decomposition to G as follows:

$$G = \sum_{j=1}^{N} \frac{u_j u_j^*}{\lambda_j - z},$$

where u_j is the $j-$th eigenvector of K_{XX} and λ_j is the $j-$th eigenvalue of K_{XX}. We then have that:

$$(GG^*)_{ii} = \sum_{j=1}^{N} \frac{|u_j(i)|^2}{|\lambda_j - z|^2} = \sum_{j=1}^{N} \frac{1}{\eta} \text{Im} \frac{|u_j(i)|^2}{\lambda_j - z} = \frac{(\text{Im}G_{ii})}{\eta}. \qquad \square$$

The Lemma above yields the following inequality:

$$|(G^2)_{ii}| \leq \frac{\text{Im}G_{ii}}{\eta}. \tag{4.20}$$

This is because

$$|(G^2)_{ii}| = |\langle e_i, G^2 e_i \rangle| \leq \|G^* e_i\| \, \|G e_i\| = |G|^2_{ii}.$$

Furthermore, we obtain the following result for the resolvent of the Sample Covariance ensemble, and the proof works also for $G^{(j_1)}_{(j_2)}$ and $\mathcal{G}^{(j_1)}_{(j_2)}$ instead of G, see [15], equation (3.10):

Lemma 4.1.5. *With G and $\mathcal{G}$ as before and $s \in \mathbb{R}$, we have that*

$$G_{ii}(E + i\eta/s) \leq sG_{ii}(E + i\eta). \tag{4.21}$$

Proof. The proof can be found in [15], page 18 and is based on the Ward property of the resolvent of a Hermitian matrix. $\qquad \square$

The following identity holds exclusively for Sample Covariance matrices and it will be the starting point of our analysis. It can be found in [14], equation 2.1.

We denote by $[X^{(j_1)}_{(j_2)}]^k$ the matrix that occurs after the extra removal of the $k-$th column, where $k \in \{j_1 + 1, ..., N\}$. We denote by $[X^{(j_1)}_{(j_2)}]_l$ the matrix that occurs after the extra removal of the $l-$th row, where $l \in \{j_2 + 1, ..., N\}$.

Lemma 4.1.6 (Sample Covariance stripping lemma). *We let $\mathbf{x}^k_{(j)}$ denote the $k-$th column of the scaled matrix $X_{(j)}$. We then have the following identity regarding the resolvent diagonal elements, for $j_1, j_2 \in \{0, 1, ..., N - 1\}$ and k as before:*

$$G^{(j_1)}_{(j_2),kk} = \frac{1}{\|\mathbf{x}^k_{(j_2)}\|^2 - z - (\mathbf{x}^k_{(j_2)})^* [X^{(j_1)}_{(j_2)}]^k \left(([X^{(j_1)}_{(j_2)}]^k)^* [X^{(j_1)}_{(j_2)}]^k - z \right)^{-1} ([X^{(j_1)}_{(j_2)}]^k)^* \mathbf{x}^k_{(j_2)}}$$

$$\tag{4.22}$$

$$= -\frac{1}{z \left(1 + (\mathbf{x}^k_{(j_2)})^* \mathcal{G}^{(k)}_{(j_2)} \mathbf{x}^k_{(j_2)} \right)} \tag{4.23}$$

Proof. The proof is based on the block inversion formula (Schur's inversion). To ease the notation, we can assume that $j_1 = j_2 = 0$, otherwise the proof is similar. We can also asume that $k = 1$. We strip the X matrix by the column $\mathbf{x}^1$, giving that:

$$G = (X^*X - z)^{-1} = M^{-1},$$

where M is the matrix:

$$M = X^*X - zI = \begin{pmatrix} \mathbf{x}^1 & X^{(1)} \end{pmatrix}^* \begin{pmatrix} \mathbf{x}^1 & X^{(1)} \end{pmatrix} - zI$$

$$= \begin{pmatrix} (\mathbf{x}^1)^* \\ (X^{(1)})^* \end{pmatrix} \begin{pmatrix} \mathbf{x}^1 & X^{(1)} \end{pmatrix} - zI$$

$$= \begin{pmatrix} \|\mathbf{x}^1\|^2 & (\mathbf{x}^1)^* X^{(1)} \\ (X^{(1)})^* \mathbf{x}^1 & (X^{(1)})^* X^{(1)} \end{pmatrix} - zI$$

$$= \begin{pmatrix} \|\mathbf{x}^1\|^2 - z & (\mathbf{x}^1)^* X^{(1)} \\ (X^{(1)})^* \mathbf{x}^1 & (X^{(1)})^* X^{(1)} - zI \end{pmatrix}.$$

By the matrix inversion formula, we get that:

$$G_{11} = (M_{11} - M_{12}M_{22}^{-1}M_{21})^{-1},$$

for the corresponding sub-blocks of M. This gives (4.22).

To derive (4.23), we use the matrix identity:

$$X^{(1)}\left[(X^{(1)})^*X^{(1)} - zI\right]^{-1}(X^{(1)})^* = X^{(1)}(X^{(1)})^*\left[X^{(1)}(X^{(1)})^* - zI\right]^{-1},$$

which can be proved as follows:

$$LHS = RHS \Leftrightarrow$$
$$X^{(1)}\left[(X^{(1)})^*X^{(1)} - zI\right]^{-1}(X^{(1)})^*\left[X^{(1)}(X^{(1)})^* - zI\right] = X^{(1)}(X^{(1)})^* \Leftrightarrow$$
$$X^{(1)}G^{(1)}(X^{(1)})^*X^{(1)}(X^{(1)})^* - zX^{(1)}G^{(1)}(X^{(1)})^* = X^{(1)}(X^{(1)})^* \Leftrightarrow$$
$$X^{(1)}\left[G^{(1)}(X^{(1)})^*X^{(1)} - zG^{(1)}\right](X^{(1)})^* = X^{(1)}(X^{(1)})^* \Leftrightarrow$$
$$X^{(1)}(X^{(1)})^* = X^{(1)}(X^{(1)})^*,$$

after which (4.23) is straight-forward. $\qquad\square$

Furthermore we have the following asymptotic behaviour for the Stieltjes transform of the Marchenko-Pastur distribution near the "soft" edge, $E \to 4$ and $\eta \to 0^+$. Notice that near the "soft" edge we have that $S_\rho + \frac{1}{2} \approx 0$ and $\mathrm{Im}(S_\rho) \approx 0$, according to (4.12).

Lemma 4.1.7. *For $E > 0$ we set $\kappa := |E - 4|$. Then for any fixed $E_0, E_1 > 0$ and $\eta_0 > 0$ there exist constants $c, C > 0$ such that*

$$\left|S_\rho + \frac{1}{2}\right| \geq C(\kappa^2 + \eta^2)^{\frac{1}{4}} \geq C\sqrt{\kappa + \eta}, \tag{4.24}$$

and

$$c\frac{\eta}{\sqrt{\kappa + \eta}} \leq \mathrm{Im}(S_\rho) \leq C\frac{\eta}{\sqrt{\kappa + \eta}}, \tag{4.25}$$

for all $E \in [E_0, E_1]$, and $\eta \in (0, \eta_0]$, with $\kappa \geq \eta$.

Proof. For (4.24) we notice that

$$S_\rho + \frac{1}{2} = \sqrt{1 - \frac{4}{z}} = \sqrt{\frac{z-4}{z}} = \sqrt{\frac{E - 4 + i\eta}{E + i\eta}},$$

and so

$$\left|S_\rho + \frac{1}{2}\right| = \sqrt{\frac{|\kappa + i\eta|}{|E + i\eta|}} \geq C\sqrt{|\kappa + i\eta|} = C(\kappa^2 + \eta^2)^{\frac{1}{4}}.$$

The result now follows from the elementary inequality:

$$\sqrt{\frac{a^2 + b^2}{2}} \geq \frac{a + b}{2}.$$

For (4.25), we notice that

$$\mathrm{Im}(S_\rho) = \frac{1}{2}\mathrm{Im}\sqrt{1 - \frac{4}{z}} \overset{(z \to 4)}{\sim} \mathrm{Im}\sqrt{z - 4}.$$

If we set $w := z - 4$, with $|w| = |\kappa + i\eta|$, we then get that:

$$\mathrm{Im}(S_\rho) \overset{(z \to 4)}{\sim} \mathrm{Im}(\sqrt{w}) = \sqrt{|w|}\sin\left(\frac{\theta}{2}\right),$$

where

$$\theta := \tan^{-1}\left(\frac{\eta}{\kappa}\right).$$

By the asymptotics $\tan^{-1}(x) \overset{x \to 0}{\sim} x$ and $\sin(x) \overset{x \to 0}{\sim} x$, we get that:

$$\mathrm{Im}(S_\rho) \overset{(z \to 4)}{\sim} (\kappa^2 + \eta^2)^{1/4}\frac{\eta}{\kappa}$$

$$\overset{(\kappa, \eta \to 0)}{\sim} \sqrt{\kappa + \eta}\frac{\eta}{\kappa}$$

$$= \frac{\eta}{\sqrt{\kappa + \eta}}\left(1 + \frac{\eta}{\kappa}\right)$$

$$\overset{(\kappa, \eta \to 0)}{\sim} \frac{\eta}{\sqrt{\kappa + \eta}},$$

where we used the fact that $\eta = \mathcal{O}(\kappa)$. This concludes the proof of (4.25). $\square$

4.2 Main theorems

To state our theorems we define the domain for our parameters where we obtain our results:

$$Z_{E,\eta} := \left\{ 4|\eta| \geq c(E^2 + \eta^2 - 4E) \mid E, \eta \in \mathbb{R} \right\} \tag{4.26}$$

for some $c > 0$. This domain is chosen so that $|\mathrm{Im}(S_\rho + 1/2)^2| \geq c|\mathrm{Re}(S_\rho + 1/2)^2|$ which we need for the proof of Proposition 4.4.1. While all the proofs work for all $c > 0$ not dependent on N, we will specifically work with $c = 1$ to allow us the opportunity to illustrate it in the following picture, Figure 4.1.
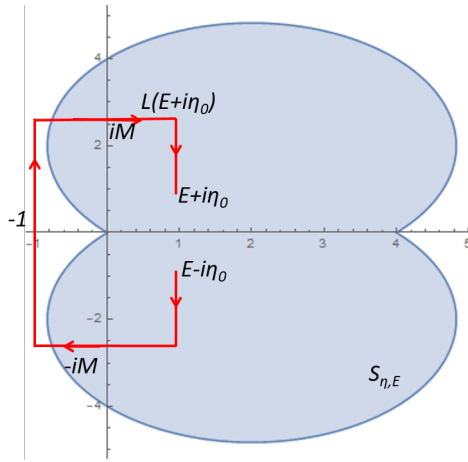
Figure 4.1: The set $Z_{E,\eta}$ shaded in blue and the integration contour $L(z_0)$ from (4.126) in red.

We assume that

$$\mathbb{E}|x_{11}|^4 =: \mu_4 < \infty, \tag{4.27}$$

and that there exists a constant $D > 0$ such that for all $N \in \mathbb{N}$ and $q \geq 1$:

$$\mathbb{E}|x_{11}|^{4q} \leq D^{4(q-1)} N^{q-1} \mu_4. \tag{4.28}$$

These assumptions are inspired by the papers of Götze-Tikhomirov [35, 62], where they assume four-moment bounds *just for independent random matrix entries* and not necessarily identically distributed. Notice that by the identical-distribution assumption, x_{11} can be replaced by any x_{ij}, $i, j = 1, ..., N$.

We are now ready to state our first theorem.

Theorem 4.2.1. *Let X_N be a $N \times N$ matrix as described in equations* (4.1) *and* (4.2) *and assume* (4.27) *and* (4.28) *for its matrix entries. Let S_N and S_ρ be the Stieltjes transforms as defined in equations* (4.8) *and* (4.12)*. Moreover set $z = E + i\eta$, with $\frac{N\eta}{|\sqrt{z}|} \geq M$ for some suitably large M. Then there exist positive constants c_0, C such that for each $K > 0$ and $1 \leq q \leq c_0 \left(\frac{N\eta}{|\sqrt{z}|} \right)^{1/8}$ and $z \in Z_{E,\eta}$ or $E < 0$:*

$$\mathbb{P}\left(|S_N(z) - S_\rho(z)| \geq \frac{K}{N\eta} \right) \leq \frac{(Cq)^{cq^2}}{K^q}. \tag{4.29}$$

Furthermore, for any $E \in \mathbb{R}$ and $\eta > 0$ such that $\frac{N\eta}{|\sqrt{z}|} \geq M$ we have that:

$$\mathbb{P}\left(|\mathrm{Im}S_N(z) - \mathrm{Im}S_\rho(z))| \geq \frac{K}{N\eta} \right) \leq \frac{(Cq)^{cq^2}}{K^q}. \tag{4.30}$$

Some remarks about this theorem are the following. Firstly, if $\eta \sim 1$, we have the so-called "global" law. Notice that in this case we also have that $|\sqrt{z}| \sim 1$. Then, for a fixed and large enough N we can pick a large enough $K(N) > 0$ such that $\frac{K}{N} \sim N^{-c}$, for some $c > 0$ and also choose $q(N)$ such that $\frac{(Cq)^{cq^2}}{K^q} \sim N^{-d}$ for some $d > 0$. This gives the standard convergence $S_N \to S_\rho$ on "global" scales $\eta \sim 1$.

For "local" scales, $\eta \to 0^+$, we have to consider two cases, whether we are near the "hard" edge, $E \to 0^+$, or somewhere else, $E \sim 1$. Notice that in the second case, we once more have that $|\sqrt{z}| \sim 1$ and we must have $N\eta$ to be large enough, so that we can choose $K(N)$ such that $\frac{K}{N\eta} \sim N^{-c}$ for some $c > 0$. This means that we can optimally have $\eta \sim \frac{1}{N^{1-\epsilon}}$, for any small $\epsilon > 0$. Choosing also a suitable $q(N)$ gives that $\frac{(Cq)^{cq^2}}{K^q} \sim N^{-d}$, for some $d > 0$. Thus, we have achieved the "local" convergence $S_N \to S_\rho$ up to the scale $\eta \sim \frac{1}{N}$ away from the "hard" edge.

What is more important in this theorem, is the case when $E, \eta \to 0^+$, which means that $|\sqrt{z}| \to 0^+$ and we get close to the "hard" edge of the Marchenko-Pastur distribution. Notice that in this case, the quantity $\frac{N\eta}{|\sqrt{z}|}$ needs to be large enough. Now, because of the singularity on the "hard" edge, the convergence $S_N \to S_\rho$ is still valid up to scales

$\eta \sim \frac{1}{N}$ but if we want to go further down to scales $\eta \sim \frac{1}{N^2}$, we have to correct it as $|\sqrt{z}||S_N - S_\rho| \to 0$. Notice that:

$$|S_N(z) - S_\rho(z)| \geq \frac{K}{N\eta} \Leftrightarrow |\sqrt{z}|(S_N - S_\rho) \geq K \frac{|\sqrt{z}|}{N\eta}.$$

We now want $\frac{N\eta}{|\sqrt{z}|} \geq M$, for some large enough M. This means that:

$$\frac{N\eta}{(E^2 + \eta^2)^{1/4}} \geq M \Leftrightarrow$$

$$\frac{\eta}{\sqrt{E + \eta}} \gtrsim \frac{1}{N} \Leftrightarrow$$

$$E + \eta \lesssim (N\eta)^2.$$

If now $\eta \sim N^{-a}$, for some $a > 0$, we must then have that:

$$-a \leq 2 - 2a \Leftrightarrow a \leq 2,$$

so that we can have this convergence up to the scale $\eta \sim \frac{1}{N^2}$. Notice that we must also have that $E = \mathcal{O}\left(\frac{1}{N^2}\right)$ for these η - scales. Thus, we have proved the convergence $|\sqrt{z}||S_N - S_\rho| \to 0$, optimally for $\eta \sim \frac{1}{N^{2-\epsilon}}$ for any small $\epsilon > 0$ since we can choose $K(N) > 0$ large enough such that $\frac{K}{\frac{N\eta}{|\sqrt{z}|}} \sim N^{-c}$, for some $c > 0$ and choose $q(N)$ such that $\frac{(Cq)^{cq^2}}{K^q} \sim N^{-d}$, for some $d > 0$. To summarize, we have proved the convergence $|\sqrt{z}||S_N - S_\rho| \to 0$, up to the scale $\eta \sim \frac{1}{N^2}$ near the "hard" edge for $E = \mathcal{O}\left(\frac{1}{N^2}\right)$.

We now state our second theorem which is a result of theorem 4.2.1. We obtain fluctuation estimates on the counting cumulative function as stated next. Letting:

$$P(E) := \int_0^E \rho(x)dx, \tag{4.31}$$

we compare it to $\rho_N(E)$, as defined in (4.4).

Theorem 4.2.2. *With assumptions as in Theorem 4.2.1, there exist constants $M_0, N_0, C, c > 0$ such that for any $K > 0$ and $E \geq \frac{M_0}{N^2}$:*

$$\mathbb{P}\left(|\rho_N(E) - P(E)| \geq K \min\left\{\sqrt{E}, \frac{\log N}{N}\right\}\right) \leq \frac{(Cq)^{cq^2}}{K^q}, \tag{4.32}$$

for all $E \in \mathbb{R}$, $K > 0$, $N > N_0$, $q \in \mathbb{N}$.

We use the above theorem 4.2.2 to obtain rigidity estimates, that is how far each eigenvalue can fluctuate away from its "classical" location.

We define the "classical" locations of the eigenvalues, predicted by the Marchenko-Pastur distribution, as the quantiles γ_i, for $i = 1, ..., N$, such that:

$$\int_0^{\gamma_i} \rho(x)dx = \frac{i}{N}.$$

Theorem 4.2.3. *With assumptions as in Theorem 4.2.1, there exist constants $C, c, N_0, \epsilon > 0$ such that:*

$$\mathbb{P}\left(|\lambda_i - \gamma_i| \geq K \frac{\log N}{N}\left(\frac{i}{N}\right)\right) \leq \frac{(Cq)^{cq^2}}{K^q} \tag{4.33}$$

for $i = 1, ..., \lceil N/2 \rceil$, $N > N_0$, $K > 0$, and $q \in \mathbb{N}$ with $q \leq N^\epsilon$. Furthermore, for $i \leq \log N$ we have that

$$\mathbb{P}\left(|\lambda_i - \gamma_i| \geq K \left(\frac{i}{N}\right)^2\right) \leq \frac{(Cq)^{cq^2}}{K^{q/2}}. \tag{4.34}$$

4.3   A quadratic formula for the difference of the Stieltjes transforms

We have defined the difference of the Stieltjes tranform $S_\rho(z)$ from its approximation $S_N(z)$ as follows:

$$\Lambda_{(j_2)}^{(j_1)} := S_N(z)_{(j_2)}^{(j_1)} - S_\rho(z). \tag{4.35}$$

The fixed-point relation for $S_\rho(z)$ is:

$$S_\rho(z) = -\frac{1}{z(S_\rho(z) + 1)}.$$ (4.36)

We will try to deduce something similar for $S_N(z)$. We denote by $\mathbf{x}_k^{(j)}$ the k-th row of the scaled matrix $X^{(j)}$ and by $\mathbf{x}_{(j)}^k$ the k-th column of the scaled matrix $X_{(j)}$. We start with the following lemma, after we define the following quantities:

$$\left[\mathcal{T}_{(j_2)}^{(j_1)}\right]^k := \frac{1}{N}\text{Tr}\left[\mathcal{G}_{(j_2)}^{(j_1)}\right]^k - \frac{1}{N}\text{Tr}\left[G_{(j_2)}^{(j_1)}\right] \quad , \quad \left[T_{(j_2)}^{(j_1)}\right]_l := \frac{1}{N}\text{Tr}\left[G_{(j_2)}^{(j_1)}\right]_l - \frac{1}{N}\text{Tr}\left[\mathcal{G}_{(j_2)}^{(j_1)}\right]_l$$ (4.37)

$$\left[\Upsilon_{(j_2)}^{(j_1)}\right]^k := (\mathbb{I}-\mathbb{E}_{\mathbf{x}_{(j_2)}^k})(\mathbf{x}_{(j_2)}^k)^*\left[\mathcal{G}_{(j_2)}^{(j_1)}\right]^k \mathbf{x}_{(j_2)}^k \quad , \quad \left[Y_{(j_2)}^{(j_1)}\right]_l := (\mathbb{I}-\mathbb{E}_{\mathbf{x}_l^{(j_1)}})(\mathbf{x}_l^{(j_1)})^*\left[G_{(j_2)}^{(j_1)}\right]_l \mathbf{x}_l^{(j_1)}$$ (4.38)

where $\mathbb{E}_{\mathbf{x}_{(j)}^k}$ denotes the expectation with respect only to the randomness of the $k-$th column of $X_{(j)}$ while $\mathbb{E}_{\mathbf{x}_l^{(j)}}$ denotes the expectation with respect only to the randomness of the $l-$th row of $X^{(j)}$. The term $\mathbb{I}$ denotes the identity operator applied to a scalar random variable, i.e $\mathbb{I}(a) := a$. The superscipt k in the resolvents denotes an extra removal of the $k-$th column of the scaled matrix X, where $k \in \{j_1 + 1, ..., N\}$. The subscript l in the resolvents denotes an extra removal of the $l-$th row of the scaled matrix X, where $l \in \{j_2 + 1, ...N\}$.

Lemma 4.3.1 (Probabilistic fixed-point relation). *For any $z \in Z_{E,\eta}$ and $N \in \mathbb{N}$ we have the following almost fixed point equation for S_N :*

$$S_N{}_{(j_2)}^{(j_1)} = -\frac{1}{N}\sum_{k=j_1+1}^{N}\frac{1}{z\left(1 + S_N{}_{(j_2)}^{(j_1)} + [\mathcal{T}_{(j_2)}^{(j_1)}]^k + [\Upsilon_{(j_2)}^{(j_1)}]^k\right)}$$ (4.39)

$$= -\frac{1}{N}\sum_{l=j_2+1}^{N}\frac{1}{z\left(1 + S_N{}_{(j_2)}^{(j_1)} + [T_{(j_2)}^{(j_1)}]_l + [Y_{(j_2)}^{(j_1)}]_l\right)},$$ (4.40)

where the quantities $[T_{(j_2)}^{(j_1)}]_l, [\Upsilon_{(j_2)}^{(j_1)}]^k, [\mathcal{T}_{(j_2)}^{(j_1)}]^k, [Y_{(j_2)}^{(j_1)}]_l,$ are the previous probabilistic error terms.

Proof. To derive the first equality (4.39) we use (4.23):

$$G^{(j_1)}_{(j_2),kk} = -\frac{1}{z\left(1 + (\mathbf{x}^k_{(j_2)})^*[\mathcal{G}^{(j_1)}_{(j_2)}]^k\mathbf{x}^k_{(j_2)}\right)}.$$

It is now enough to show that (for $j_1 = j_2 = 0$):

$$\mathbb{E}_{\mathbf{x}^k}\left[(\mathbf{x}^k)^*\mathcal{G}^k\mathbf{x}^k\right] = \frac{1}{N}\mathrm{Tr}\left[\mathcal{G}^k\right]. \tag{4.41}$$

The first identity will then follow, because we can take summation in (4.23) and then divide by N.

To prove (4.41), we notice that:

$$\mathbb{E}_{\mathbf{x}^k}\left[(\mathbf{x}^k)^*\mathcal{G}^k\mathbf{x}^k\right] = \mathbb{E}_{\mathbf{x}^k}\left[\sum_{i,j=1}^N \overline{x_{ki}}\mathcal{G}^k_{ij}x_{jk}\right] = \frac{1}{N}\sum_{i=1}^N \mathcal{G}^k_{ii},$$

where we used the fact that the entries of X are independent with expectation 0 and variance $\frac{1}{N}$. We also used the fact that the matrix $\mathcal{G}^k$ is independent of the column $\mathbf{x}^k$.

The second equality (4.40) comes from the following modified stripping identity, the proof of which is similar to the original:

$$\mathcal{G}^{(j_1)}_{(j_2),ll} = -\frac{1}{z\left(1 + (\mathbf{x}^{(j_1)}_l)\left(([X^{(j_1)}_{(j_2)}]_l)^*[X^{(j_1)}_{(j_2)}]_l - z\right)^{-1}(\mathbf{x}^{(j_1)}_l)^*\right)}, \tag{4.42}$$

where l is defined as before. It similarly yields the second part (4.40), recalling (4.16). □

We have an easy deterministic bound for the quantities $[\mathcal{T}^{(j_1)}_{(j_2)}]^k$ and $[T^{(j_1)}_{(j_2)}]_l$ in the following lemma:

Lemma 4.3.2.

$$\left|[T^{(j_1)}_{(j_2)}]_l\right|, \left|[\mathcal{T}^{(j_1)}_{(j_2)}]^k\right| \le \frac{|j_1 - j_2|+1}{N\eta}. \tag{4.43}$$

Proof. We prove the bound for $[\mathcal{T}^{(j_1)}_{(j_2)}]^k$. We use the stripping lemma 4.1.3 for the proof.

We use the notation $G = G_{(j_2)}^{(j_1)}$.

$$
\begin{aligned}
[\mathcal{T}_{(j_2)}^{(j_1)}]^k &= -\frac{j_1 - j_2}{Nz} + \frac{1}{N} \left(\sum_{i \neq k} G_{ii}^k - \sum_{i=1}^{N} G_{ii} \right) \\
&= -\frac{j_1 - j_2}{Nz} + \frac{1}{N} \left(\sum_{i \neq k} \left(G_{ii} - \frac{G_{ik} G_{ki}}{G_{kk}} \right) - \sum_{i=1}^{N} G_{ii} \right) \\
&= -\frac{j_1 - j_2}{Nz} - \frac{1}{N} \left(\sum_{i \neq k} \frac{G_{ik} G_{ki}}{G_{kk}} - G_{kk} \right) = -\frac{j_1 - j_2}{Nz} - \frac{1}{N} \frac{1}{G_{kk}} \sum_{i=1}^{N} G_{ik} G_{ki} \\
&= -\frac{j_1 - j_2}{Nz} - \frac{(G^2)_{kk}}{N G_{kk}}.
\end{aligned}
$$

We now use (4.20) to obtain

$$
\left| [T_{(j_2)}^{(j_1)}]^k \right| \leq \frac{|j_1 - j_2|}{N|z|} + \frac{\mathrm{Im} G_{kk}}{|G_{kk}| N \eta}
$$

yielding that

$$
\left| [T_{(j_2)}^{(j_1)}]^k \right| \leq \frac{|j_1 - j_2| + 1}{N \eta}. \tag{4.44}
$$

A similar argument also works for the quantity $[T_{(j_2)}^{(j_1)}]_l$, replacing G with $\mathcal{G}$ and using the modified stripping lemma for $\mathcal{G}$:

$$
\mathcal{G}_{(j_2),ij}^{(j_1)} = \left[\mathcal{G}_{(j_2)}^{(j_1)} \right]_{l,ij} + \frac{\mathcal{G}_{(j_2),il}^{(j_1)} \mathcal{G}_{(j_2),li}^{(j_1)}}{\mathcal{G}_{(j_2),ll}^{(j_1)}}, \tag{4.45}
$$

which holds for $l \in \{j_2 + 1, ..., N\}$ and $i, j \neq l$. $\qquad \square$

It is highly non-trivial to bound the terms $[\Upsilon_{(j_2)}^{(j_1)}]^k$ and $[Y_{(j_2)}^{(j_1)}]_l$ if we only have a four moment bound as an assumption for the matrix entries of X. We will bound them in section 4.5. In the case of a more powerful assumption like a sub-gaussian decay of the entries, bounds on these quantities would be almost trivial.

We now connect the two similar fixed-point equations yielding a quadratic formula for the difference $\Lambda = S_N - S_\rho$ involving a new probabilistic error term R which combines the previous ones and prevents the equation from being trivial with a zero solution.

We start with the equation (4.39) and modify it by including in it the information provided by equation (4.13) to deduce an equation for Λ.

We will use the identity

$$\frac{1}{A+\epsilon} = \frac{1}{A} - \frac{\epsilon}{A(A+\epsilon)},$$

where ϵ will represent the sum of the two probabilistic error terms plus the quantity Λ, all multiplied by z.

We obtain that (also for any j_1, j_2 and so we suppress here the notation):

$$
\begin{aligned}
S_N &= -\frac{1}{N}\sum_{k=1}^{N}\frac{1}{z(1+S_\rho)+z\Lambda+z(T^k+\Upsilon^k)} \\
&= -\frac{1}{z(1+S_\rho)}+\frac{1}{N}\sum_{k=1}^{N}\frac{1}{z(1+S_\rho)}\frac{z\Lambda+z(T^k+\Upsilon^k)}{z\left(1+S_N+T^k+\Upsilon^k\right)} \\
&= S_\rho+\frac{S_\rho}{N}\sum_{k=1}^{N}z\Lambda G_{kk}-\frac{S_\rho}{N}\sum_{k=1}^{N}G_{kk}z(T^k+\Upsilon^k) \\
&= S_\rho+S_\rho z\Lambda S_N-\frac{S_\rho}{N}\sum_{k=1}^{N}G_{kk}z(T^k+\Upsilon^k).
\end{aligned}
\tag{4.46}
$$

This yields that

$$\Lambda = zS_\rho\Lambda(S_\rho+\Lambda)-\frac{S_\rho}{N}\sum_{k=1}^{N}G_{kk}z(T^k+\Upsilon^k). \tag{4.47}$$

We can now define the error term R we were talking about, as:

$$R := \frac{1}{N}\sum_{k=1}^{N}G_{kk}(T^k+\Upsilon^k). \tag{4.48}$$

Of course we can similarly define the error term $R^{(j_1)}_{(j_2)}$. Now (4.47) together with the definition of R yields the following quadratic equation for Λ:

$$zS_\rho\Lambda^2 + (zS_\rho^2 - 1)\Lambda + zS_\rho R = 0. \tag{4.49}$$

Dividing by zS_ρ, using that $zS_\rho = -\frac{1}{1+S_\rho}$ and the quadratic formula, yields

$$-(S_\rho+1/2) \pm \sqrt{(S_\rho+1/2)^2 - R} \tag{4.50}$$

as two solutions for our estimated difference $S_N - S_\rho$. From the definition of Λ in (4.17), it follows that $\text{Im}(\Lambda) \geq -\text{Im}(S_\rho)$. Thus, if we take the branch cut of the square root to be on the positive reals so that the imaginary part of the square root is always positive, we obtain that:

$$\Lambda = -(S_\rho + 1/2) + \sqrt{(S_\rho + 1/2)^2 - R} \tag{4.51}$$

We also notice that the second solution, call it $\tilde{\Lambda}$, to (4.49) is given by:

$$\tilde{\Lambda} = -\Lambda - 2S_\rho - 1. \tag{4.52}$$

4.4 A first look at the error term R

We explain here how exactly the probabilistic error term R quantifies the distance of Λ from 0 according to our quadratic equation. We get the following estimates:

Proposition 4.4.1. *Let $z = E + i\eta$. There exists a constant $C > 0$, such that:*

$$|\Lambda| \leq C \min\left\{ \frac{|R|}{|S_\rho + \frac{1}{2}|}, \sqrt{|R|} \right\}, \tag{4.53}$$

for all $(E, \eta) \in Z_{E,\eta}$ or $E < 0$. Furthermore, for any $E \in \mathbb{R}$ and $\eta > 0$ we have that:

$$|\text{Im}\Lambda| \leq C \min\left\{ \frac{|R|}{|S_\rho + \frac{1}{2}|}, \sqrt{|R|} \right\} \tag{4.54}$$

and

$$\min\{|\Lambda|, |\tilde{\Lambda}|\} \leq C\sqrt{|R|}. \tag{4.55}$$

Analogous statements hold for $\Lambda_{(j_2)}^{(j_1)}$ with $R_{(j_2)}^{(j_1)}$.

We will use the first bound near the "hard" edge, because $|S_\rho + \frac{1}{2}| \to \infty$ when $z \to 0$ and so it will be most useful there. The $\sqrt{|R|}-$ bound will be most useful in the "bulk" and near the "soft" edge, where $z \sim 1$.

Proof. We apply the following Lemma which can be found in [15], page 12:

Lemma 4.4.2. *We denote by $\sqrt{w}$ the square root of w with $\mathrm{Im}\sqrt{w} \geq 0$.*

- *For any fixed $c > 0$, there exists a constant $C > 0$ such that:*

$$\left| \sqrt{a+b} - \sqrt{a} \right| \leq C \frac{|b|}{\sqrt{|a|+|b|}}, \tag{4.56}$$

  for all $a, b \in \mathbb{C}$ with $|\mathrm{Im}(a)| \geq c\mathrm{Re}(a)$.

- *There exists a constant $C > 0$ such that:*

$$\left| \mathrm{Im}\left(\sqrt{a+b} \right) - \mathrm{Im}(\sqrt{a}) \right| \leq C \frac{|b|}{\sqrt{|a|+|b|}}, \tag{4.57}$$

  for all $a, b \in \mathbb{C}$.

We apply this Lemma with $a = (S_\rho + \frac{1}{2})^2$ and $b = -R$.

Since $\mathrm{Im}(S_\rho + \frac{1}{2}) \geq 0$, with our choice of branch cut we have that $\sqrt{(S_\rho + 1/2)^2} = S_\rho + 1/2$, and we recall that we defined $Z_{E,\eta}$ in (4.26) to be exactly the set where $\left| \mathrm{Im}\left[(S_\rho + 1/2)^2 \right] \right| \geq c \left| \mathrm{Re}\left[(S_\rho + 1/2)^2 \right] \right|$ for some $c > 0$. These observations are enough for the proof of (4.53) according to the proof of Lemma 4.4.2 in [15] on page 13.

Analogously, (4.54) follows directly from (4.57) and again the fact that $\sqrt{(S_\rho + 1/2)^2} = S_\rho + 1/2$. For the $\sqrt{R}-$ bound, it is again enough the follow the proof of 4.4.2 in [15] on page 13.

The proof of (4.55) follows from Vieta's formula:

$$|\Lambda||\tilde{\Lambda}| = |R|,$$

so that we must have:

$$\min\{|\Lambda|, |\tilde{\Lambda}|\} \leq \sqrt{|R|}.$$

We now extend (4.53) for $E < 0$. Recalling that $S_N(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - E - i\eta}$ and noting that for $E < 0$ the real part of each summand is positive we conclude that $\mathrm{Re}(S_N) > 0$ for $E < 0$, and similarly to our argument about the imaginary part of Λ, we see from (4.17) that $\mathrm{Re}(\Lambda) > -\mathrm{Re}(S_\rho)$ while from (4.52) we see that $\mathrm{Re}(\tilde{\Lambda}) < -\mathrm{Re}(S_\rho) - 1$. Since we have that:

$$\mathrm{Re}(\Lambda) = -\mathrm{Re}(S_\rho + 1/2) + \mathrm{Re}\left(\sqrt{(S_\rho + 1/2)^2 - R}\right)$$

$$\mathrm{Re}(\tilde{\Lambda}) = -\mathrm{Re}(S_\rho + 1/2) - \mathrm{Re}\left(\sqrt{(S_\rho + 1/2)^2 - R}\right),$$

we see that $\mathrm{Re}\left(\sqrt{(S_\rho + 1/2)^2 - R}\right) > 0$ and thus $|\mathrm{Re}(\Lambda)| < |\mathrm{Re}(\tilde{\Lambda})|$ and thus one part of (4.53) follows from (4.55). For the other part of (4.53), we estimate that:

$$|\Lambda| = \left| \frac{R}{\sqrt{(S_\rho + 1/2)^2 - R} + (S_\rho + 1/2)} \right| \leq \left| \frac{R}{S_\rho + 1/2} \right|, \tag{4.58}$$

where the last inequality follows since both the real and imaginary parts of both summands in the denominator are positive. The fact that $\mathrm{Re}(S_\rho + \frac{1}{2}) \geq 0$ comes from the definition of the Stieltjes transform which integrates a measure that has a non-negative support. $\qquad\square$

What remains now is to focus on the error term

$$R := \frac{1}{N} \sum_{k=1}^{N} G_{kk}(T^k + \Upsilon^k) \tag{4.59}$$

and deduce some bounds for it (also for $R_{(j_2)}^{(j_1)}$). To obtain our optimal bounds, the deterministic bounds on T^k are enough, Υ^k will be regarded as a quadratic form and will get bounded in section 4.5. The resolvent entry G_{kk} will get bounded in section 4.6.

4.5 The four-moments condition and bounds for quadratic forms

Here we obtain bounds for the quadratic form $\Upsilon := \frac{1}{N}(\mathbb{I} - \mathbb{E}_{\mathbf{x}})\mathbf{x}^*\mathcal{G}\mathbf{x}$, where $\mathbf{x}$ is a column vector of X.

We remark that we only have a four-moment condition as an assumption for the entries of X. If instead, we had a sub-gaussian decay assumption for the entries, a bound for Υ would be much easier to deduce by using the Hanson-Wright inequality for quadratic forms as for example in [14], page 5. The proof here heavily relies on the Ward identity (4.1.4) which is true both for G and $\mathcal{G}$.

We simplify the notation by taking $j_1 = j_2 = 0$, but everything would work out for a different case as well.

We prove the following Lemma:

Lemma 4.5.1. *Let $\mathcal{G} = \mathcal{G}$ or G. Let $\Upsilon := \frac{1}{N}(\mathbb{I} - \mathbb{E}_{\mathbf{x}})\mathbf{x}^*\mathcal{G}\mathbf{x}$, assuming (4.27) and (4.28) for the elements of $\mathbf{x}$. Then, for any $q \geq 1$, we have that*

$$\mathbb{E}|\Upsilon|^{2q} \leq (Cq)^{cq}\left(\frac{\mathbb{E}\left(\mathrm{Im}\mathrm{Tr}\mathcal{G}\right)^q}{N^q(N\eta)^q} + \frac{\mathbb{E}|\mathcal{G}_{11}|^{2q}}{N^q} + \frac{\mathbb{E}|\mathcal{G}_{11}|^q}{(N\eta)^q}\right). \tag{4.60}$$

Moreover, we have the more precise inequality:

$$\mathbb{E}|\Upsilon|^{2q} \leq \frac{(Cq)^{cq}}{(N\eta)^q}\left(\mathbb{E}\left(\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N}\right)^q + \mathbb{E}\left|\frac{\mathcal{G}_{11}}{\sqrt{N}}\right|^q\right) + (Cq)^{cq}\frac{\mathbb{E}|\mathcal{G}_{11}|^{2q}}{N^q}. \tag{4.61}$$

Proof. We start by the decomposition:

$$\Upsilon = \frac{1}{N}\sum_{j\neq l}\overline{x_j}x_l\mathcal{G}_{jl} + \frac{1}{N}\sum_j(|x_j|^2 - 1)\mathcal{G}_{jj} = \epsilon_2 + \epsilon_1,$$

where

$$\epsilon_2 := \frac{1}{N}\sum_{j\neq l}\overline{x_j}x_l\mathcal{G}_{jl} \quad \text{and} \quad \epsilon_1 := \frac{1}{N}\sum_j(|x_j|^2 - 1)\mathcal{G}_{jj}. \tag{4.62}$$

We use Rosenthal's inequality (Lemma 4.5.2) to obtain for any $q \geq 1$, that:

$$\mathbb{E}|\epsilon_1|^{2q} \leq (Cq)^{2q} N^{-2q} \left[\left(\mu_4 \sum_{j=1}^{N} \mathbb{E}|\mathcal{G}_{jj}|^2 \right)^q + \sum_{j=1}^{N} \mathbb{E}|x_j|^{4q} \mathbb{E}|\mathcal{G}_{jj}|^{2q} \right]. \tag{4.63}$$

We use our four-moment assumption (4.28):

$$\mathbb{E}|x_j|^{4q} \leq D^{4q-4} N^{q-1} \mu_4,$$

which yields that

$$\mathbb{E}|\epsilon_1|^{2q} \leq (Cq)^{2q} N^{-q} \mathbb{E}|\mathcal{G}_{11}|^{2q}. \tag{4.64}$$

For ϵ_2 we will systematically use both Burkholder's and Rosenthal's inequalities, expressed for complex random variables in the Lemmas 4.5.2 and 4.5.3. Using Burkholder's inequality we obtain:

$$\mathbb{E}|\epsilon_2|^{2q} \leq N^{-2q}(C_1 q)^{2q} \left[\mathbb{E} \left(\sum_{j=2}^{N} \left| \sum_{k=1}^{j-1} x_k \mathcal{G}_{jk} \right|^2 \right)^q + \max_k \mathbb{E}|x_k|^{2q} \sum_{j=2}^{N} \mathbb{E} \left| \sum_{k=1}^{j-1} x_k \mathcal{G}_{jk} \right|^{2q} \right]$$

$$+ N^{-2q}(C_1 q)^{2q} \left[\mathbb{E} \left(\sum_{j=2}^{N} \left| \sum_{1 \leq k \leq j-1} x_k \mathcal{G}_{kj} \right|^2 \right)^q + \max_k \mathbb{E}|x_k|^{2q} \sum_{j=2}^{N} \mathbb{E} \left| \sum_{k=1}^{j-1} x_k \mathcal{G}_{kj} \right|^{2q} \right] \tag{4.65}$$

We define the quantities:

$$Q_0 := \sum_{j=2}^{N} \left| \sum_{k=1}^{j-1} x_k \frac{\mathcal{G}_{jk}}{\sqrt{N}} \right|^2 \quad \text{and} \quad \widehat{Q}_0 := \sum_{j=2}^{N} \left| \sum_{k=1}^{j-1} x_k \frac{\mathcal{G}_{kj}}{\sqrt{N}} \right|^2 \tag{4.66}$$

The difficult part of the proof will be to bound expectations of powers of these quantities. For the other terms we apply Rosenthal's inequality and (4.27) getting that:

$$\mathbb{E}|\epsilon_2|^{2q} \leq (Cq)^{2q} N^{-q} (\mathbb{E}|Q_0|^q + \mathbb{E}|\widehat{Q}_0|^q) + (Cq)^{4q} N^{-\frac{3q}{2}-1} \times \tag{4.67}$$

$$\left[\sum_{j=2}^{N} \mathbb{E} \left\{ \left(\sum_{k=1}^{j-1} |\mathcal{G}_{jk}|^2 \right)^q + \left(\sum_{k=1}^{j-1} |\mathcal{G}_{kj}|^2 \right)^q \right\} + \sum_{j=2}^{N} \sum_{k=1}^{j-1} \mathbb{E}|x_k|^{2q} (\mathbb{E}|\mathcal{G}_{jk}|^{2q} + \mathbb{E}|\mathcal{G}_{kj}|^{2q}) \right]$$

For the last terms we observe that

$$|\mathcal{G}_{jk}| \leq \frac{1}{2}\sqrt{\frac{\mathrm{Im}\mathcal{G}_{jj}}{\eta}} + \frac{1}{2}\sqrt{\frac{\mathrm{Im}\mathcal{G}_{kk}}{\eta}}, \tag{4.68}$$

which can be obtained as follows. Let $(u_i)_{i=1}^N$ be the normalized eigenvectors of $\mathcal{G}$ and $\lambda_0 = 0$. Then

$$|\mathcal{G}_{lj}| = \left|\sum_{q=0}^{N} \frac{u_{lq}u_{qj}}{\lambda_q - z}\right| \leq \sum_{q=0}^{N} \frac{|u_{lq}u_{qj}|}{|\lambda_q - z|} \leq \frac{1}{2}\sum_{q=0}^{N} \frac{|u_{lq}|^2 + |u_{qj}|^2}{|\lambda_q - z|}$$

$$\leq \frac{1}{2}\sqrt{\sum_{q=0}^{N} \frac{|u_{lq}|^2}{|\lambda_q - z|^2}} + \frac{1}{2}\sqrt{\sum_{q=0}^{N} \frac{|u_{qj}|^2}{|\lambda_q - z|^2}},$$

where in the last step we recall that the eigenvectors are normalized and use Jensen's inequality. Then (4.68) follows. Using for the first terms the Ward identity (4.1.4), $\sum_{l=1}^{N} |\mathcal{G}_{jl}|^2 \leq \eta^{-1}\mathrm{Im}\, \mathcal{G}_{jj}$, we now get that

$$\mathbb{E}|\epsilon_2|^{2q} \leq (Cq)^{2q}N^{-q}\left[\mathbb{E}|Q_0|^q + \mathbb{E}|\widehat{Q}_0|^q\right] + \frac{(Cq)^{4q}}{(N\eta)^q}\mathbb{E}|\mathrm{Im}\mathcal{G}_{11}|^q. \tag{4.69}$$

We will now bound the quantity $\mathbb{E}|Q_0|^q$, and note that $\mathbb{E}|\widehat{Q}_0|^q$ is similar. We will implement an induction scheme on the quantity $\mathbb{E}|Q_0|^q$ to gradually decrease its exponent q and finally remove it. The technique is similar to the one in [35, 62] but we extend Rosenthal and Burkholder inequalities to complex entries and improve the bound near the "soft" edge, where $\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N} = \mathrm{Im}S_N \approx \mathrm{Im}S_\rho \to 0$, see (4.60) and (4.61). Notice that this improvement is captured only on (4.61).

Similarly to [35] we define the following quantities. This analysis will let us exploit the resolvent identities for the resolvent entries that appear on the quadratic forms Q_0 and $\widehat{Q}_0$. Also, notice the decomposition (4.73).

$$Q_\nu := \sum_{j=2}^N \left| \sum_{k=1}^{j-1} x_j a_{jk}^{(\nu)} \right|^2, \quad Q_{\nu 1} := \sum_{l=1}^N a_{ll}^{(\nu+1)},$$

$$Q_{\nu 2} := \sum_{l=1}^N \left(|x_l|^2 - 1 \right) a_{ll}^{(\nu+1)}, \quad \text{and} \quad Q_{\nu 3} := \sum_{l \neq j} x_l \overline{x_j} a_{lj}^{(\nu+1)}, \quad (4.70)$$

where $a_{jk}^{(\nu)}$ are defined recursively via

$$a_{jk}^{(0)} := \frac{\mathcal{G}_{jk}}{\sqrt{N}} \quad \text{and} \quad a_{jk}^{(\nu+1)} := \sum_{l=\max\{j,k\}+1}^N a_{jl}^{(\nu)} \overline{a_{kl}^{(\nu)}} \qquad (4.71)$$

for $\nu = 0, 1, ..., L-1, L$ where L is an integer such that $2^{L-1} < q \leq 2^L$.

From [35] Lemma 5.1 and Corollaries 5.2 and 5.3, we have the following bounds:

$$\max \left\{ |a_{rr}^{(\nu+1)}|, \sum_j |a_{jr}^{(\nu)}|^2 \right\} \leq \left(\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N\eta} \right)^{2^\nu - 1} \frac{\mathrm{Im}\mathcal{G}_{rr}}{N\eta} \qquad (4.72)$$

To set up our induction scheme we expand the absolute value square and interchange the order of summations in Q_ν:

$$Q_\nu = \sum_{j=2}^N \sum_{1 \leq k_1, k_2 \leq j-1} x_{k_1} \overline{x_{k_2}} a_{k_1 j}^{(\nu)} \overline{a_{k_2 j}^{(\nu)}} = \sum_{1 \leq k_1, k_2 \leq N-1} x_{k_1} \overline{x_{k_2}} \sum_{j=\max\{k_1, k_2\}+1}^N a_{k_1 j}^{(\nu)} \overline{a_{k_2 j}^{(\nu)}}$$

$$= \sum_{1 \leq j_1, j_2 \leq N} x_{j_1} \overline{x_{j_2}} a_{j_1 j_2}^{(\nu+1)} = Q_{\nu 1} + Q_{\nu 2} + Q_{\nu 3}. \qquad (4.73)$$

We now take the power $2^{L-\nu}$ so as to match the case $\mathbb{E}|Q_0|^{2^L}$, for $\nu = 0$, after we gradually increase the powers under an induction scheme from $\nu = L$ to $\nu = 0$. The base case will be $\mathbb{E}|Q_L|$.

Taking power $2^{L-\nu}$ and expectation we obtain that:

$$\mathbb{E}|Q_\nu|^{2^{L-\nu}} \leq 3^{2^{L-\nu}} \left(\mathbb{E}|Q_{\nu 1}|^{2^{L-\nu}} + \mathbb{E}|Q_{\nu 2}|^{2^{L-\nu}} + \mathbb{E}|Q_{\nu 3}|^{2^{L-\nu}} \right). \qquad (4.74)$$

Firstly, by the definition of $Q_{\nu 1}$ and the definition of the coefficients $a_{ll}^{(\nu+1)}$ and (4.72)

we can check that:

$$\mathbb{E}|Q_{\nu 1}|^{2^{L-\nu}} \le \mathbb{E}\left[\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N\eta}\right]^{2^L}. \tag{4.75}$$

We apply Rosenthal's inequality for the quantity $Q_{\nu 2}$ getting that:

$$
\mathbb{E}|Q_{\nu 2}|^{2^{L-\nu}} \le (Cq)^q \left[\mathbb{E}\left(\sum_{l=1}^{N}|a_{ll}^{(\nu+1)}|^2\right)^{2^{L-(\nu+1)}} + \sum_{l=1}^{N}\mathbb{E}||x_l|^2-1|^{2^{L-\nu}}|a_{ll}^{(\nu+1)}|^{2^{L-\nu}}\right]
$$

$$
\le (Cq)^q \left[\mathbb{E}\left(\sum_{l=1}^{N}|a_{ll}^{(\nu+1)}|^2\right)^{2^{L-(\nu+1)}} + N^{2^{L-(\nu+1)}}\frac{1}{N}\sum_{l=1}^{N}\mathbb{E}|a_{ll}^{(\nu+1)}|^{2^{L-\nu}}\right], \tag{4.76}
$$

where we used (4.28) in the last line. We notice that the first term is bounded above by the second term by Jensen's inequality, so we obtain that:

$$\mathbb{E}|Q_{\nu 2}|^{2^{L-\nu}} \le (Cq)^q N^{2^{L-(\nu+1)}}\frac{1}{N}\sum_{l=1}^{N}\mathbb{E}|a_{ll}^{(\nu+1)}|^{2^{L-\nu}} \le (Cq)^q J_\nu, \tag{4.77}$$

where we used (4.72) and introduced the notation

$$J_\nu := N^{-2^{L-(\nu+1)}}\mathbb{E}\left[\left|\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N\eta}\right|^{2^L-2^{L-\nu}}\left|\frac{\mathrm{Im}\mathcal{G}_{11}(z)}{\eta}\right|^{2^{L-\nu}}\right]. \tag{4.78}$$

Now we apply Burkholder's inequality to $\mathbb{E}|Q_{\nu 3}|^q$ to obtain a bound which involves $\mathbb{E}|Q_{\nu+1}|^{q/2}$ and $\mathbb{E}|\widehat{Q}_{\nu+1}|^{q/2}$ as in (4.65). We use Rosenthal's inequality to bound the other term arising from the application of Burkholder's inequality:

$$
\mathbb{E}|Q_{\nu 3}|^{2^{L-\nu}} \le \mathbb{E}\left|\sum_{j_1 \ne j_2} x_{j_1}\overline{x_{j_2}}a_{j_1 j_2}^{(\nu+1)}\right|^{2^{L-\nu}}
$$

$$
\le (Cq)^q \left(\mathbb{E}|Q_{\nu+1}|^{2^{L-(\nu+1)}} + \mathbb{E}|\widehat{Q}_{\nu+1}|^{2^{L-(\nu+1)}}\right)
$$

$$
+ (Cq)^q \mathbb{E}|x_1|^{2^{L-\nu}}\sum_{j=2}^{n}\mathbb{E}\left(\left|\sum_{k=1}^{j-1}a_{jk}^{(\nu+1)}x_k\right|^{2^{L-\nu}} + \left|\sum_{k=1}^{j-1}a_{kj}^{(\nu+1)}x_k\right|^{2^{L-\nu}}\right)
$$

$$
\le (Cq)^q \left(\mathbb{E}|Q_{\nu+1}|^{2^{L-(\nu+1)}} + \mathbb{E}|\widehat{Q}_{\nu+1}|^{2^{L-(\nu+1)}}\right)
$$

$$
+ (Cq)^{2q}N^{2^{L-(\nu+2)}}\frac{1}{N}\sum_{j=2}^{N}\mathbb{E}\left(\sum_{1\le l\le N, l\ne j}|a_{jl}^{(\nu+1)}|^2\right)^{2^{L-(\nu+1)}}
$$

$$+ (Cq)^q N^{2^{L-(\nu+1)}} \frac{1}{N^2} \sum_{j=2}^{N} \sum_{1 \leq l \leq N, l \neq j} \mathbb{E}|a_{jl}^{(\nu+1)}|^{2^{L-\nu}} \tag{4.79}$$

The resulting terms are bounded by (4.72) and the following argument. By Hölder inequality and the definition (4.71), we obtain that:

$$\frac{1}{N^2} \sum_{j=2}^{N} \sum_{1 \leq r \leq N, r \neq j} |a_{rj}^{(\nu+1)}|^{2^{L-\nu}} \leq \frac{1}{N^2} \sum_{j=2}^{N} \sum_{1 \leq r \leq N, r \neq j} \left| \sum_{l_1} |a_{rl_1}^{(\nu+1)}|^2 \sum_{l_2} |a_{l_2 j}^{(\nu+1)}|^2 \right|^{2^{L-(\nu+1)}}$$

$$\leq \frac{1}{N^2} \sum_{j=2}^{N} \sum_{1 \leq r \leq N, r \neq j} \left| a_{rr}^{(\nu)} a_{jj}^{(\nu)} \right|^{2^{L-(\nu+1)}} \leq \left(\frac{1}{N} \sum_{j} |a_{jj}^{(\nu)}|^{2^{L-(\nu+1)}} \right)^2$$

$$\leq \frac{1}{N} \sum_{j} |a_{jj}^{(\nu)}|^{2^{L-\nu}}, \tag{4.80}$$

where in the last step we used Jensen's inequality. This yields that:

$$\mathbb{E}|Q_{\nu 3}|^{2^{L-\nu}} \leq (Cq)^q \left(\mathbb{E}|Q_{\nu+1}|^{2^{L-(\nu+1)}} + \mathbb{E}|\widehat{Q}_{\nu+1}|^{2^{L-(\nu+1)}} \right)$$

$$+ (Cq)^{2q} N^{2^{L-(\nu+2)}} \mathbb{E} \left| \left(\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N\eta} \right)^{2^{\nu+1}-1} \frac{\mathrm{Im}\mathcal{G}_{11}}{N\eta} \right|^{2^{L-(\nu+1)}}$$

$$+ (Cq)^{q} N^{2^{L-(\nu+1)}} \mathbb{E} \left| \left(\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N\eta} \right)^{2^{\nu}-1} \frac{\mathrm{Im}\mathcal{G}_{11}}{N\eta} \right|^{2^{L-\nu}}$$

$$\leq (Cq)^q \left(\mathbb{E}|Q_{\nu+1}|^{2^{L-(\nu+1)}} + \mathbb{E}|\widehat{Q}_{\nu+1}|^{2^{L-(\nu+1)}} \right) + (Cq)^{2q}(J_{\nu+1} + J_{\nu}). \tag{4.81}$$

Same bounds hold for $\widehat{Q}_\nu$ with $\widehat{Q}_{\nu 1}$, $\widehat{Q}_{\nu 2}$ and $\widehat{Q}_{\nu 3}$ defined analogously. Thus, by (4.74), we have proved that:

$$\mathbb{E}|Q_\nu|^{2^{L-\nu}} \leq 3^{2^{L-\nu}} \left(\mathbb{E}|Q_{\nu 1}|^{2^{L-\nu}} + \mathbb{E}|Q_{\nu 2}|^{2^{L-\nu}} + \mathbb{E}|Q_{\nu 3}|^{2^{L-\nu}} \right) \leq$$

$$(Cq)^{2q} \left(\mathbb{E} \left[\frac{\mathrm{Im}\mathrm{Tr}\mathcal{G}}{N\eta} \right]^{2^L} + J_\nu + \mathbb{E}|Q_{\nu+1}|^{2^{L-(\nu+1)}} + \mathbb{E}|\widehat{Q}_{\nu+1}|^{2^{L-(\nu+1)}} + (J_{\nu+1} + J_\nu) \right),$$

for any $\nu = 0, 1, ..., L$. By an induction argument we obtain that:

$$\mathbb{E}(|Q_0|^{2^L} + |\widehat{Q}_0|^{2^L}) \leq$$

$$\leq (Cq)^{cq} \left(\mathbb{E}(|Q_L| + |\widehat{Q}_L|) + \sum_{\nu=0}^{L-1} \left(\mathbb{E}|Q_{\nu 1}|^{2^{L-\nu}} + \mathbb{E}|\widehat{Q}_{\nu 1}|^{2^{L-\nu}} + \mathbb{E}|Q_{\nu 2}|^{2^{L-\nu}} + \mathbb{E}|\widehat{Q}_{\nu 2}|^{2^{L-\nu}} + J_\nu + J_{\nu+1} \right) \right)$$

$$\leq (Cq)^{cq} \left(\mathbb{E}|Q_L| + \mathbb{E}|\widehat{Q}_L| + \sum_{\nu=0}^{L-1} \mathbb{E}\left(|Q_{\nu 1}|^{2^{L-\nu}} + |\widehat{Q}_{\nu 1}|^{2^{L-\nu}} \right) + \sum_{\nu=0}^{L} J_\nu \right). \tag{4.82}$$

Now, we bound the three terms on the RHS. For $0 \leq \nu \leq L$ we have that

$$J_\nu = \eta^{-2^L} \mathbb{E}\left[\left(\frac{\operatorname{Im}\operatorname{Tr}\mathcal{G}}{N} \right)^{2^L} \left| \frac{\operatorname{Im}\mathcal{G}_{11}}{\sqrt{N}} \frac{N}{\operatorname{Im}\operatorname{Tr}\mathcal{G}} \right|^{2^{L-\nu}} \right]$$

$$\leq \eta^{-2^L} \mathbb{E}\left[\left(\frac{\operatorname{Im}\operatorname{Tr}\mathcal{G}}{N} \right)^{2^L} + \left| \frac{\operatorname{Im}\mathcal{G}_{11}}{\sqrt{N}} \right|^{2^L} \right], \tag{4.83}$$

where in the last step we used Young's inequality and $0 \leq \nu \leq L$.

Using (4.73) and (4.72) we obtain

$$\mathbb{E}|Q_L| = \mathbb{E}\left| \sum_{1 \leq j_1, j_2 \leq N} x_{j_1} \overline{x_{j_2}} a_{j_1 j_2}^{(L+1)} \right| = \sum_{j=1}^{N} \mathbb{E}|a_{jj}^{(L+1)}| \leq \mathbb{E}\left(\frac{\operatorname{Im}\operatorname{Tr}\mathcal{G}}{N\eta} \right)^{2^L}, \tag{4.84}$$

and a similar bound for $\mathbb{E}|\widehat{Q}_L|$.

Upon substitution of (4.75), (4.83) and (4.84) into (4.82), we get that:

$$\mathbb{E}|Q_0|^q + \mathbb{E}|\widehat{Q}_0|^q \leq (Cq)^{cq} \eta^{-q} \left(\mathbb{E}\left(\frac{\operatorname{Im}\operatorname{Tr}\mathcal{G}}{N} \right)^q + \mathbb{E}\left| \frac{\operatorname{Im}\mathcal{G}_{11}}{\sqrt{N}} \right|^q \right), \tag{4.85}$$

We can now go back to (4.69) and replace this bound on the RHS. The desired bounds (4.60) and (4.61) on $\mathbb{E}|\Upsilon|^{2q}$ will follow from (4.64) and (4.69). $\qquad\square$

Rosenthal's and Burkholder's inequalities.

Here we state Rosenthal's and Burkholder's inequalities adapted to complex variables and non-Hermitian bilinear forms (useful for $\mathcal{G}$). We are given complex random variables $x_1, ..., x_N$ with i.i.d. real and imaginary parts and $\mathbb{E}\left[\mathrm{Re}(x_j)\right] = \mathbb{E}\left[\mathrm{Im}(x_j)\right] = 0$ and $\mathbb{E}\left|\mathrm{Re}(x_j)\right|^2 = \mathbb{E}\left|\mathrm{Im}(x_j)\right|^2 = 1/2$, for $j = 1, ..., N$ as in our setup. We assume that for $j = 1, ..., N : \mathbb{E}|x_j|^p \leq \mu_p$ for any $p \geq 1$, so that all the moments exist and are bounded by these quantities. In our setup, they may also depend on N.

The following Lemma is our version of Rosenthal's inequality. Here we have a vector $\mathbf{a} = (a_1, ..., a_N)$ of complex scalars. It is easy to prove by separating real and imaginary parts of the random variables and using Lemma 7.1 of [35]:

Lemma 4.5.2 (Rosenthal's inequality). *For any $p \geq 1$, there exists a constant C_1 such that*

$$\mathbb{E}\left|\sum_{j=1}^{N} a_j x_j\right|^p \leq (C_1 p)^p \left[\left(\sum_{j=1}^{N}|a_j|^2\right)^{p/2} + \mu_p \sum_{j=1}^{N}|a_j|^p\right]. \tag{4.86}$$

Proof. We let $x_j = \mathrm{Re}(x_j) + i\mathrm{Im}(x_j)$ for $j = 1, .., N$ and then use Lemma 7.1 of [35]. $\square$

The following Lemma is our version of Burkholder's inequality. Here, we have a family of complex scalars $(a_{ij})_{i,j=1}^{N}$. This is an extension of Lemma 7.3 of [35] for complex entries and non-Hermitian quadratic forms. Let

$$Q := \sum_{j \neq k} a_{jk} x_j \overline{x_k}.$$

Lemma 4.5.3 (Burkholder's Inequality). *For any $q \geq 1$, there exist absolute constants C_1, C_2 such that:*

$$\mathbb{E}|Q|^q \leq (C_1 q)^q \left(\mathbb{E}\left[\sum_{j=2}^{n}\left|\sum_{k=1}^{j-1} a_{jk} x_k\right|^2\right]^{q/2} + \mu_q \sum_{j=2}^{n}\mathbb{E}\left|\sum_{k=1}^{j-1} a_{jk} x_k\right|^q\right)$$

$$+ (C_2 q)^q \left(\mathbb{E}\left[\sum_{j=2}^{n}\left|\sum_{k=1}^{j-1} a_{kj} x_k\right|^2\right]^{q/2} + \mu_q \sum_{j=2}^{n}\mathbb{E}\left|\sum_{k=1}^{j-1} a_{kj} x_k\right|^q\right).$$

Proof. We introduce for $j = 1, 2, ..., N$ the random variables

$$\xi_j := x_j \sum_{k=1}^{j-1} a_{jk} x_k \ , \ \widehat{\xi}_j := x_j \sum_{k=1}^{j-1} a_{kj} x_k \tag{4.87}$$

We let $\mathcal{R}_j := \sigma(\xi_1, ..., \xi_j)$ be the sigma-algebra generated by the first j random variables $\xi_1, ..., \xi_j$. We observe that the ξ_j and $\widehat{\xi}_j$ are $\mathcal{R}_j-$measurable with $\mathbb{E}[\xi_j | \mathcal{R}_{j-1}] = 0$ and $\mathbb{E}[\widehat{\xi}_j | \mathcal{R}_{j-1}] = 0$, which means that they form martingale differences. Next, we write Q as

$$Q = \sum_{j=2}^{n} \xi_j + \sum_{j=2}^{n} \widehat{\xi}_j \tag{4.88}$$

and so,

$$\mathbb{E}|Q|^q \leq C^q \mathbb{E} \left| \sum_{j=2}^{n} \xi_j \right|^q + C^q \mathbb{E} \left| \sum_{j=2}^{n} \widehat{\xi}_j \right|^q.$$

We now apply a general Burkholder-Rosenthal Inequality as seen in [50], analogous to Lemma 7.2 from [35], to the martingale-difference sequences $\xi_1, ..., \xi_n$ and $\widehat{\xi}_1, ..., \widehat{\xi}_n$. We just have to evaluate $\mathbb{E}[\xi_j^2 | \mathcal{R}_{j-1}]$ and $\mathbb{E}|\xi_j|^q$, for $j = 1, 2, ..., N$. The case is similar for the $\widehat{\xi}_1, ..., \widehat{\xi}_n$ random variables. We therefore observe that:

$$\mathbb{E}\left[|\xi_j|^2 \Big| \mathcal{R}_{j-1} \right] = \mathbb{E}|x_j|^2 \left| \sum_{k=1}^{j-1} a_{jk} x_k \right|^2 = \left| \sum_{k=1}^{j-1} a_{jk} x_k \right|^2 ,$$

$$\mathbb{E}|\xi_j|^q = \mathbb{E}|\zeta_j|^q \mathbb{E} \left| \sum_{k=1}^{j-1} a_{jk} x_k \right|^q \leq \mu_q \, \mathbb{E} \left| \sum_{k=1}^{j-1} a_{jk} x_k \right|^q ,$$

and the lemma follows after applying the inequality seen in [50]. $\qquad \square$

4.6 The resolvent bounds

In this section we will bound the quantities $\mathbb{E}|G_{kk}|^q$ that appear in the main error term R defined in (4.59).

We remind that our strategy is to bound high powers of the expected value of this term (i.e. bound $\mathbb{E}|R|^q$) and use Markov's inequality to show that it is small in probability.

Let

$$\lambda_{(j_2)}^{(j_1)} := \max \left\{ \left|\Lambda_{(j_2)}^{(j_1)}\right| \chi_{S_{E,\eta}}, \min\left\{ \left|\Lambda_{(j_2)}^{(j_1)}\right|, \left|\tilde\Lambda_{(j_2)}^{(j_1)}\right| \right\}, \left|\mathrm{Im}\Lambda_{(j_2)}^{(j_1)}\right| \right\} \tag{4.89}$$

By Proposition 4.4.1, $\mathbb{E}\left|\lambda_{(j_2)}^{(j_1)}\right|^{2q} \le C^{2q}\mathbb{E}\left|R_{(j_2)}^{(j_1)}\right|^{q}$. Taking expectation of a power $q \ge 1$ of $\left|R_{(j_2)}^{(j_1)}\right|$ we obtain (as in [15]), using (4.44), (4.60), (4.16), Jensen's and Cauchy-Schwartz inequalities:

$$\mathbb{E}\left|R_{(j_2)}^{(j_1)}\right|^{q} \le \frac{1}{N}\sum_{k=j_1+1}^{N} \mathbb{E}\left|\left([T_{(j_2)}^{(j_1)}]^k + [\Upsilon_{(j_2)}^{(j_1)}]^k\right) G_{(j_2)kk}^{(j_1)}\right|^{q} \le$$

$$\mathbb{E}\left|\left([T_{(j_2)}^{(j_1+1)}] + [\Upsilon_{(j_2)}^{(j_1+1)}]\right) G_{(j_2),11}^{(j_1)}\right|^{q}$$

$$\le \frac{\mathbb{E}\left|C(|j_1-j_2|+1)G_{(j_2),11}^{(j_1)}\right|^{q}}{(N\eta)^q} + |C|^q\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}\mathbb{E}\left|[\Upsilon_{(j_2)}^{(j_1+1)}]\right|^{2q}}$$

$$\le \frac{\mathbb{E}\left|C(|j_1-j_2|+1)G_{(j_2),11}^{(j_1)}\right|^{q}}{(N\eta)^q} + |Cq|^{cq}\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}\frac{\mathbb{E}|\mathcal{G}_{(j_2),11}^{(j_1+1)}|^{2q}}{N^q}}$$

$$+ |Cq|^{cq}\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}\left(\frac{\mathbb{E}(\mathrm{Im}\mathrm{Tr}\mathcal{G}_{(j_2)}^{(j_1+1)})^q}{(N\eta)^q N^q} + \frac{\mathbb{E}|\mathcal{G}_{(j_2),11}^{(j_1+1)}|^q}{(N\eta)^q}\right)}$$

$$\le \frac{\mathbb{E}\left|C(|j_1-j_2|+1)G_{(j_2),11}^{(j_1)}\right|^{q}}{(N\eta)^q} + |Cq|^{cq}\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}\frac{\mathbb{E}|\mathcal{G}_{(j_2),11}^{(j_1+1)}|^{2q}}{N^q}}$$

$$+ C^q\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}\frac{(Cq)^{cq}}{(N\eta)^q}\left(\left(\frac{|j_1+1-j_2|}{N\eta}\right)^q + \mathbb{E}\left(\mathrm{Im}S_\rho + \mathrm{Im}\Lambda_{(j_2)}^{(j_1+1)}\right)^q + \mathbb{E}|\mathcal{G}_{(j_2),11}^{(j_1+1)}|^q\right)}$$

$$\le \frac{|C(|j_1-j_2|+1)|^q\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}}}{(N\eta)^q} + |Cq|^{cq}\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}}\sqrt{\frac{\mathbb{E}|\mathcal{G}_{(j_2),11}^{(j_1+1)}|^{2q}}{N^q}} \tag{4.90}$$

$$+ |Cq|^{cq}\frac{\sqrt{\mathbb{E}|G_{(j_2),11}^{(j_1)}|^{2q}}}{(N\eta)^{\frac{q}{2}}}\left[\left(\frac{|j_1+1-j_2|}{N\eta}\right)^{q/2} + \sqrt{\left(\frac{C}{\sqrt{|z|}}\right)^q + \mathbb{E}|\lambda_{(j_2)}^{(j_1)}|^q} + \sqrt{\mathbb{E}|\mathcal{G}_{(j_2),11}^{(j_1+1)}|^q}\right].$$

In the second to last line, the term $\frac{|j_1+1-j_2|}{N|z|}$ arises from equation (4.16) and $|\lambda_{(j_2)}^{(j_1)} - \lambda_{(j_2)}^{(j_1+1)}| = \mathcal{O}\left(\frac{1}{N\eta}\right)$ similarly with the proof of (4.44). We notice here that in order to get finite bounds near the "hard" edge, we have to extinguish the $\frac{1}{\sqrt{|z|}}$ term, that came from the imaginary part of S_ρ. This suggests that we should multiply everything with $|z|^q$ and actually get bounds for the quantity $\mathbb{E}|zR|^q$, which in turn suggests bounds for the resolvent entries of the form $\mathbb{E}|\sqrt{z}G_{11}|^q$ and $\mathbb{E}|\sqrt{z}\mathcal{G}_{11}|^q$.

Multiplying by $|z|^q$, (4.90) becomes:

$$\mathbb{E}\left|zR^{(j_1)}_{(j_2)}\right|^q \leq$$

$$\frac{|C\sqrt{z}(|j_1-j_2|+1)|^q}{(N\eta)^q}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}} + |Cq|^{cq}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}\sqrt{\frac{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^{2q}}{N^q}}$$

$$+ |Cq|^{cq}\frac{|z|^{q/4}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}}{(N\eta)^{\frac{q}{2}}}\left[|j_1+1-j_2|^{q/2} + \sqrt{\mathbb{E}|\sqrt{z}\lambda^{(j_1)}_{(j_2)}|^q} + \sqrt{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^q}\right]$$

$$\leq C^q\frac{|z|^{q/2}(|j_1-j_2|+1)^q}{(N\eta)^q}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}} + (Cq)^{cq}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}\frac{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^{2q}}{N^q}}$$

$$(Cq)^{cq}\frac{|z|^{q/4}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}}{(N\eta)^{\frac{q}{2}}}\left[|j_1-j_2+1|^{q/2} + \sqrt{\mathbb{E}|\sqrt{z}\lambda^{(j_1)}_{(j_2)}|^q} + \sqrt{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^q}\right].$$

$$(4.91)$$

We now try to deduce a bound for $\mathbb{E}|\sqrt{z}\lambda^{(j_1)}_{(j_2)}|^{2q}$. Since for any $x, \delta > 0$,

$$x^{1/4} \leq \delta x + \delta^{-1/3},$$

for $x = \mathbb{E}|\sqrt{z}\lambda^{(j_1)}_{(j_2)}|^{2q}$ by Cauchy-Schwartz we obtain that:

$$x \leq (Cq)^q\frac{|z|^{q/2}(|j_1-j_2|+1)^q}{(N\eta)^q}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}} + (Cq)^{cq}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}\sqrt{\frac{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^{2q}}{N^q}}$$

$$+ (Cq)^{cq}\frac{|z|^{q/4}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}}{(N\eta)^{\frac{q}{2}}}\left[\delta x + \delta^{-1/3} + \sqrt{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^q} + |j_1-j_2+1|^{q/2}\right].$$

Writing this as

$$x \leq A + B\left[\delta x + \delta^{-1/3} + \Gamma\right] \Leftrightarrow$$

$$(1 - B\delta)x \leq A + B\delta^{-1/3} + B\Gamma,$$

we see that we can set

$$\delta = \frac{B^{-1}}{2} = \left[2(Cq)^{cq}\frac{\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}}{(N\eta)^{\frac{q}{2}}}\right]^{-1},$$

to get that:

$$
\mathbb{E}|\sqrt{z}\lambda^{(j_1)}_{(j_2)}|^{2q} \leq (Cq)^{cq}\frac{|z|^{q/2}(|j_1-j_2|+1)^{cq}}{(N\eta)^q}\left((\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q})^{1/2}\right)
$$
$$
+ (Cq)^{cq}\frac{|z|^{q/3}\left(\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}\right)^{2/3}}{(N\eta)^{2q/3}} + (Cq)^{cq}\sqrt{\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}\sqrt{\frac{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^{2q}}{N^q}}
$$
$$
+ (Cq)^{cq}\sqrt{\frac{|z|^{q/2}\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^{2q}}{(N\eta)^q}}\left[\sqrt{\mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^q} + |j_1-j_2+1|^{q/2}\right].
$$

$$(4.92)$$

Lemma 4.6.1 (Resolvent bounds). *Let $E \leq 4, \eta \leq \eta_0$ and $q \leq \left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/4}$ with $\frac{N\eta}{|\sqrt{z}|} \geq M$ for some suitable large constant $M > 0$. Assume that $j_1, j_2 \in \{0, ..., N-1\}$ are such that $0 \leq |j_1 - j_2| \leq C_1 q$ for a uniform constant C_1. Then with definitions as before,*

$$
\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}|^q \leq C^q \quad \text{and} \quad \mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}|^q \leq C^q,
$$

for some constant C.

Proof. We will implement an induction argument similar to [15, 35]. The induction hypothesis will be that for $\eta_i = \eta_0/16^i$ for some constant η_0 and any j_1, j_2 with $|j_1 - j_2| \leq L_i := C_1\left(\frac{N\eta_i}{|\sqrt{E}|}\right)^{1/4}$.

$$
\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}(\eta_i)|^q \leq C_0^q \quad \text{and} \quad \mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}(\eta_i)|^q \leq C_0^q \tag{4.93}
$$

for $q \leq \left(\frac{N\eta_i}{|\sqrt{E}|}\right)^{1/4}$ for a universal constant C_0.

<u>Induction basis:</u> We notice that this holds to initiate our induction for η_0 as a constant. This can be proved for example by the inequality $|S_N(z)| \leq \frac{1}{\eta}$ as seen in section 3.1, point 4.

<u>Induction step:</u> Letting $\eta_{i+1} = \eta_i/16$ and $L_{i+1} = L_i/2$ we will show that inequality (4.93) taken at η_i implies the same inequality with the same constant C_0 for η_{i+1}.

From the induction hypothesis and Lemma 4.1.5 we see that

$$\mathbb{E}|\sqrt{z}G^{(j_1)}_{(j_2),11}(\eta_{i+1})|^q \leq (16C_0)^q \quad \text{and} \quad \mathbb{E}|\sqrt{z}\mathcal{G}^{(j_1+1)}_{(j_2),11}(\eta_{i+1})|^q \leq (16C_0)^q \tag{4.94}$$

for any j_1, j_2 with $|j_1 - j_2| \leq L_i$. This will need to be improved to the bound C_0^q for any k_1, k_2 with $|k_1 - k_2| \leq 2^{-1}L_i$.

We will use the inequality $|k_1 - k_2| \leq L_{i+1}$ in equation (4.101).

From (4.23), (4.37), (4.38), (4.13) and (4.52) we obtain that (for η_{i+1}):

$$G^{(k_1)}_{(k_2),11} = S_\rho + zS_\rho \left(\Lambda^{(k_1)}_{(k_2)} + [\mathcal{T}^{(k_1)}_{(k_2)}]^N + [\Upsilon^{(k_1)}_{(k_2)}]^N \right) G^{(k_1)}_{(k_2),11} \tag{4.95}$$

$$G^{(k_1)}_{(k_2),11} = S_\rho - zS_\rho \left(\tilde{\Lambda}^{(k_1)}_{(k_2)} - [\mathcal{T}^{(k_1)}_{(k_2)}]^N - [\Upsilon^{(k_1)}_{(k_2)}]^N \right) G^{(k_1)}_{(k_2),11}$$
$$+ zS_\rho(2S_\rho + 1)G^{(k_1)}_{(k_2),11}. \tag{4.96}$$

The analogous statements for $\mathcal{G}^{(k_1)}_{(k_2),11}$ follow similarly from (4.42), (4.37), (4.38), (4.13) and (4.52) (for η_{i+1}):

$$\mathcal{G}^{(k_1)}_{(k_2),11} = S_\rho + zS_\rho \left(\Lambda^{(k_1)}_{(k_2)} + [T^{(k_1)}_{(k_2)}]_N + [Y^{(k_1)}_{(k_2)}]_N \right) \mathcal{G}^{(k_1)}_{(k_2),11} \tag{4.97}$$

$$\mathcal{G}^{(k_1)}_{(k_2),11} = S_\rho - zS_\rho \left(\tilde{\Lambda}^{(k_1)}_{(k_2)} - [T^{(k_1)}_{(k_2)}]_N - [Y^{(k_1)}_{(k_2)}]_N \right) \mathcal{G}^{(k_1)}_{(k_2),11}$$
$$+ zS_\rho(2S_\rho + 1)\mathcal{G}^{(k_1)}_{(k_2),11}. \tag{4.98}$$

This yields that:

$$|G^{(k_1)}_{(k_2),11}| \leq |S_\rho| + \left|G^{(k_1)}_{(k_2),11}\right| \left(\left|\sqrt{z}\Lambda^{(k_1)}_{(k_2)}\right| + \left|\sqrt{z}[\mathcal{T}^{(k_1)}_{(k_2)}]^N\right| + \left|\sqrt{z}[\Upsilon^{(k_1)}_{(k_2)}]^N\right| \right) |\sqrt{z}S_\rho|$$

$$|G^{(k_1)}_{(k_2),11}| \leq \left|\frac{S_\rho}{1 - zS_\rho(2S_\rho + 1)}\right| + \left|G^{(k_1)}_{(k_2),11}\right| \left(\left|\sqrt{z}\tilde{\Lambda}^{(k_1)}_{(k_2)}\right| + \left|\sqrt{z}[T^{(k_1)}_{(k_2)}]^N\right| + \left|\sqrt{z}[\Upsilon^{(k_1)}_{(k_2)}]^N\right| \right) \times$$
$$\left|\frac{\sqrt{z}S_\rho}{1 - zS_\rho(2S_\rho + 1)}\right|,$$

and using (4.13) we see that $1 - zS_\rho(2S_\rho + 1) = 2 - zS_\rho^2$ and thus $\frac{S_\rho}{1 - zS_\rho(2S_\rho + 1)} = \frac{S_\rho}{2 - zS_\rho^2}$.

We use the bounds $|zS_\rho^2 - 2| \geq c_1$ and $|\sqrt{z}S_\rho| \leq C_2$, valid in our domain, and let $C = \max\{C_2, C_2/c_1\}$.

We then have that:

$$|\sqrt{z}G^{(k_1)}_{(k_2),11}|\leq C\left[1+\left|\sqrt{z}G^{(k_1)}_{(k_2),11}\right|\left(|\sqrt{z}|\min\left\{|\Lambda^{(k_1)}_{(k_2)}|,|\tilde{\Lambda}^{(k_1)}_{(k_2)}|\right\}+\left|\sqrt{z}[T^{(k_1)}_{(k_2)}]^N\right|+\left|\sqrt{z}[\Upsilon^{(k_1)}_{(k_2)}]^N\right|\right)\right]$$

and taking power q, expectation, and using Jensen, Cauchy-Schwarz and (4.44) we get at η_{i+1}:

$$\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^q\leq C^q\left[1+\sqrt{\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^{2q}}\sqrt{\mathbb{E}|\sqrt{z}\lambda^{(k_1)}_{(k_2)}|^{2q}}\right.$$

$$\left.+\frac{|\sqrt{z}|^q(|k_1-k_2|+1)^q}{(N\eta_{i+1})^q}\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^q+\sqrt{\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^{2q}}\sqrt{\mathbb{E}|\sqrt{z}[\Upsilon^{(k_1)}_{(k_2)}]^N|^{2q}}\right]\quad(4.99)$$

Using the above, Lemma 4.5.1 and a calculation similar to (4.90) we obtain again at η_{i+1}:

$$\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^q\leq C^{cq}\left[1+\sqrt{\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^{2q}}\sqrt{\mathbb{E}|\sqrt{z}\lambda^{(k_1)}_{(k_2)}|^{2q}}\right.$$

$$+\frac{|\sqrt{z}|^q(|k_1-k_2|+1)^q}{(N\eta_{i+1})^q}\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^q$$

$$+(Cq)^{cq}\sqrt{\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^{2q}}\frac{|\sqrt{z}|^{q/4}}{(N\eta_{i+1})^{q/2}}\sqrt{1+\mathbb{E}|\sqrt{z}\lambda^{(k_1)}_{(k_2)}|^q+\mathbb{E}|\sqrt{z}\mathcal{G}^{(k_1+1)}_{(k_2),11}|^q}$$

$$\left.+\sqrt{\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^{2q}\frac{\mathbb{E}|\sqrt{z}\mathcal{G}^{(k_1+1)}_{(k_2),11}|^{2q}}{N^q}}\right]\quad(4.100)$$

We use (4.94) to bound the terms $\mathbb{E}|\sqrt{z}G^{(k_1)}_{(k_2),11}|^{2q}$ and $\mathbb{E}|\sqrt{z}\mathcal{G}^{(k_1+1)}_{(k_2),11}|^{2q}$ in the above inequality, noting that $|k_1-k_2|\leq L_{i+1}\leq L_i$. To use (4.94) we need $2q\leq\left(\frac{N\eta_{i+1}}{|\sqrt{E}|}\right)^{1/4}$, which gives us $q\leq\left(\frac{N\eta_i}{16|\sqrt{E}|}\right)^{1/4}$, which is what we need. Then using (4.94) on equation (4.92) at η_{i+1} and recalling that $|k_1-k_2|\leq C_1q$ we obtain:

$$\mathbb{E}|\sqrt{z}\lambda^{(k_1)}_{(k_2)}|^q\leq(Cq)^{cq}\frac{|z|^{q/2}}{(N\eta_{i+1})^q}(16C_0)^{q/2}+(Cq)^{cq}\frac{|z|^{q/3}}{(N\eta_{i+1})^{2q/3}}(16C_0)^{2q/3}$$

$$+(Cq)^{cq}\frac{|z|^{q/4}}{(N\eta_{i+1})^{q/2}}(16C_0)^{3q/4}+(Cq)^{cq}\frac{(16C_0)^q}{N^{q/2}}.\quad(4.101)$$

Since we will choose a large enough $M > 0$ such that $\frac{N\eta_{i+1}}{|\sqrt{z}|} \geq M$ to work the argument, we can assume that $\frac{N\eta_{i+1}}{|\sqrt{z}|} > 1$, so that (4.101) becomes:

$$\mathbb{E}|\sqrt{z}\lambda_{(k_2)}^{(k_1)}|^q \leq (Cq)^{cq}\frac{|z|^{q/4}}{(N\eta_{i+1})^{q/2}}(16C_0)^{3q/4} + (Cq)^{cq}\frac{(16C_0)^q}{N^{q/2}}. \qquad (4.102)$$

Substituting this into (4.100), we obtain that at η_{i+1}:

$$\mathbb{E}|\sqrt{z}G_{(k_2),11}^{(k_1)}|^q \leq C^{cq}\Bigg\{ 1 + (16C_0)^q(Cq)^{cq}\left[\frac{|\sqrt{z}|^{q/2}}{(N\eta_{i+1})^{q/2}}(16C_0)^{q/4} + \frac{(16C_0)^q}{N^{q/2}}\right] +$$

$$+ (16C_0)^q(Cq)^{cq}\frac{|\sqrt{z}|^{q/2}}{(N\eta_{i+1})^{q/2}}\sqrt{1 + \frac{|\sqrt{z}|^{q/2}}{(N\eta_{i+1})^{q/2}}(16C_0)^{q/4} + \frac{(16C_0)^q}{N^{q/2}} + (16C_0)^q + \frac{(16C_0)^{2q}}{N^{q/2}}}\Bigg\}$$

$$\leq C^{cq}\left[1 + K^q\left(\frac{|\sqrt{z}|}{N\eta_{i+1}}\right)^{q/4}\right],$$

where we used the bound for q. Here we have a constant $K > 0$ depending on C_0 and C. We can choose $C_0 > 2C^c$ and $\frac{N\eta_{i+1}}{|\sqrt{z}|} > M > K^4$, so that $K^q\left(\frac{|\sqrt{z}|}{N\eta_{i+1}}\right)^{q/4} < 1$ and therefore $\mathbb{E}|\sqrt{z}G_{(k_2),11}^{(k_1)}(\eta_{i+1})]^q \leq C_0^q$ as required.

We notice that all the steps are identical for $\mathcal{G}_{(k_2),11}^{(k_1+1)}$ using (4.97) and exactly one row gets stripped as well as exactly one column so that $|k_1 + 1 - (k_2 + 1)| = |k_1 - k_2| \leq L_{i+1}$. $\qquad \square$

Combining 4.6.1 with (4.92), we can deduce a first bound for Λ, which will be optimized in the next section.

Corollary 4.6.2 (Weak local law). *Let $z \in Z_{E,\eta}$ such that $\frac{N\eta}{|\sqrt{z}|} \geq M$, for some suitably large constant M. Then, there exist $C, C_1 > 0$ such that:*

$$\mathbb{E}|\Lambda|^{2q} \leq \frac{(Cq)^{cq}}{(N\eta)^{q/2}},$$

for any $0 \leq q \leq C_1\left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/4}$.

4.7 Optimal bound for the Stieltjes transform

In this section we prove Theorem 4.2.1. We will use the matrix expansion algorithm from [15], which carries over directly as it is based entirely on linear algebra of resolvents. We will make a note of the important modifications. We note, importantly, that as we expand resolvent entries, we will be removing columns of X_N and we never need to remove rows. The expansion algorithm yields results in terms of high moments of the following quantities:

$$|\sqrt{z}G_{kk}^{(\mathbb{J})}|, \left|\frac{1}{\sqrt{z}G_{kk}^{(\mathbb{J})}}\right|, \left|(\mathbb{I} - \mathbb{E}_k)\frac{1}{\sqrt{z}G_{kk}^{(\mathbb{J})}}\right|, |\sqrt{z}G_{kl}^{(\mathbb{J})}|, \tag{4.103}$$

where $\mathbb{J}$ is any subset of $\{1, .., N\}$ and $G^{(\mathbb{J})}$ means that we remove all columns of the scaled matrix X that belong on that subset $\mathbb{J}$ and then take the resolvent of K_{XX}. Analogously, $G_{(\mathbb{J})}$ means that we remove all rows of the scaled matrix X that belong on that subset $\mathbb{J}$ and then take the resolvent of K_{XX}.

To obtain optimal bounds on Λ, we will use the precise inequality in (4.61).

We begin this section by estimating high moments of the quantities in (4.103). We will use (4.60) to obtain bounds on $\mathbb{E}|\sqrt{z}G_{(j_2),kl}^{(j_1)}|$ as well as $|\mathbb{E}\sqrt{z}\mathcal{G}_{(j_2),kl}^{(j_1)}|$.

For convenience of notation, we introduce the control parameter:

$$\mathcal{E}_q := \frac{1}{N^q|z|^{q/2}} + \max\left\{\frac{[\mathrm{Im}(|z|S_\rho)]^q + \mathbb{E}|z\Lambda|^q}{(N\eta)^q}, \frac{|z|^q}{(N\eta)^{2q}}\right\}. \tag{4.104}$$

We now show how to estimate the last quantity in (4.103), using the following formulas (see for example (2.20) of [54]) (valid also for any j_1, j_2, with $G = G_{(j_2)}^{(j_1)}$ or $\mathcal{G} = \mathcal{G}_{(j_2)}^{(j_1)}$ with $k, l > \max\{j_1, j_2\}$):

$$G_{kl} = zG_{ll}G_{kk}^{(\{l\})}\left[(\mathbf{x}^l)^*\mathcal{G}^{(\{k,l\})}\mathbf{x}^k\right] =: \sqrt{z}G_{ll}\sqrt{z}G_{kk}^{(\{l\})}K_{kl}$$
$$\mathcal{G}_{kl} = z\mathcal{G}_{ll}\mathcal{G}_{(\{l\}),kk}\left[\mathbf{x}_l G_{(\{k,l\})}(\mathbf{x}_k)^*\right] =: \sqrt{z}\mathcal{G}_{ll}\sqrt{z}\mathcal{G}_{(\{l\}),kk}\mathcal{K}_{kl}. \tag{4.105}$$

We can define $K_{(j_2),kl}^{(j_1)}, \mathcal{K}_{(j_2),kl}^{(j_1)}$ analogously. The following lemma provides the necessary

bound on $\mathbb{E}|K_{kl}|^{2q}$.

Lemma 4.7.1. *Assume* (4.27) *and* (4.28) *for the entries of the matrix* X_N *as before and let* $z = E + i\eta$. *Then there exist constants* $c, C, M_1, M_2 > 0$ *such that*

$$|z|^q \max\{\mathbb{E}|K_{kl}|^{2q}, \mathbb{E}|\mathcal{K}_{kl}|^{2q}\} \leq (Cq)^{cq}\mathcal{E}_q \tag{4.106}$$

for $E, \eta \in Z_{E,\eta}$, $N > M_1$, $\frac{N\eta}{|\sqrt{z}|} > M_2$, $k \neq l \in \{1, ..., N\}$, $q \in \mathbb{N}$ *with* $q \leq c\left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/4}$. *Assuming* $|j_1 - j_2| \leq Cq$ *for some constant* C, *the same inequality holds for* $K^{(j_1)}_{(j_2),kl}, \mathcal{K}^{(j_1)}_{(j_2),kl}$. *with* $k, l > \max\{j_1, j_2\}$.

Proof. The following argument is identical for $K^{(j_1)}_{(j_2),kl}, \mathcal{K}^{(j_1)}_{(j_2),kl}$, so we work with K_{kl}. By the definition of K_{kl} and using the notation $\epsilon_{k1}, \epsilon_{k_2}$ for ϵ_1 and ϵ_2 as in (4.62) we get that:

$$|z|^q \mathbb{E}|K_{kl}|^{2q} \leq \frac{(Cq)^{cq}}{N^{2q}}\left(\mathbb{E}|\epsilon_{k2}|^{2q} + \mathbb{E}\sum_j |\mathcal{G}^{(kl)}_{jj} x_{kj} x_{lj}|^{2q}\right) \leq \frac{(Cq)^{cq}}{(N\eta)^q}|z|^{q/2} \tag{4.107}$$

where $\mathbb{E}|\epsilon_{k2}|$ is bounded using (4.69), (4.85) and $\mathbb{E}\sum_j |\mathcal{G}^{(kl)}_{jj} x_{kj} x_{lj}|^{2q}$ is bounded by Rosenthal's inequality like $\mathbb{E}|\epsilon_{k1}|^{2q}$ in (4.63). We also used Lemma 4.6.1 to bound $\mathbb{E}|\mathcal{G}_{kk}|^{2q}$. In summary, we used the bound (4.60) to bound this quantity.

We now improve this to (4.106). We use the bound (4.61), after we modify it using Lemma 4.6.1 and the fact that:

$$\frac{2}{(N\eta)^q N^{q/2}} \leq \frac{1}{(N\eta)^{2q}} + \frac{1}{N^q},$$

to obtain:

$$\mathbb{E}|\Upsilon|^{2q} \leq \frac{(Cq)^{cq}}{(N\eta)^{2q}}\mathbb{E}|\mathrm{Im}\,\mathrm{Tr}\,\mathcal{G}|^q + (Cq)^{cq}\left(\frac{1}{N^q} + \frac{1}{(N\eta)^{2q}}\right) \tag{4.108}$$

and using this, we can improve the bound on (4.107), which yields (4.106). $\qquad\square$

This gives the desired bound:

$$\mathbb{E}|\sqrt{z}G^{(\mathbb{J})}_{12}|^{2q} \leq (Cq)^{cq}\mathcal{E}_q, \tag{4.109}$$

where $|\mathbb{J}| \leq Cq$.

Lemma 4.7.2. *Assume (4.27) and (4.28) for the entries of X_N as before and let $z = E + i\eta \in Z_{E,\eta}$. Then, there exist constants $c, C, M > 0$ such that:*

$$\mathbb{E}\frac{1}{|\sqrt{z}G_{11}^{(\mathbb{J})}|^{2q}} \leq C^q,$$

for $\frac{N\eta}{\sqrt{|z|}} > M$, $q \leq \left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/4}$, with $|\mathbb{J}| \leq cq$.

Proof. We can take $|\mathbb{J}| = 0$ as the argument is similar in the general case. We use (4.23) to get that:

$$\mathbb{E}\frac{1}{|\sqrt{z}G_{11}|^{2q}} = \mathbb{E}\left|\sqrt{z}(1 + (\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1)\right|^{2q} \leq C^q + C^q|z|^q\mathbb{E}\left|(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1\right|^{2q}$$
$$\leq C^q\left(1 + |z|^q\mathbb{E}\left|(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1 - \mathbb{E}_{\mathbf{x}^1}(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1\right|^{2q} + \mathbb{E}\left|\mathbb{E}_{\mathbf{x}^1}\sqrt{z}(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1\right|^{2q}\right).$$

The second term on the RHS is small by Lemma 4.5.1. For the third term, we find that:

$$\mathbb{E}\left|\mathbb{E}_{\mathbf{x}^1}\sqrt{z}(\mathbf{x}^1)^*\mathcal{G}^{(1)}\mathbf{x}^1\right|^{2q} = \mathbb{E}\left|\frac{1}{N}\sqrt{z}\mathrm{Tr}(\mathcal{G}^{(1)})\right|^{2q} = \mathbb{E}\left|\frac{1}{N}\sqrt{z}\left(\frac{1}{z} + \mathrm{Tr}(G^{(1)})\right)\right|^{2q} \leq C^q,$$
$$(4.110)$$

where we used Lemma 4.6.1 and that $|S_N^{(1)} - S_N| \leq \frac{1}{N\eta}$ as in (4.44), as well as the bound $\frac{N\eta}{\sqrt{|z|}} > M$. $\qquad\square$

To estimate the third quantity in (4.103), we find by (4.108) that:

$$\mathbb{E}\left|(\mathbb{I} - \mathbb{E}_{\mathbf{x}^k})\frac{1}{\sqrt{z}G_{kk}}\right|^{2q} = \mathbb{E}\left|-\sqrt{z}\Upsilon^{(\{k\})}\right|^{2q} \leq (Cq)^{cq}\mathcal{E}_q, \qquad (4.111)$$

where we used (4.23) and the precise bound (4.61) along with Lemma 4.6.1.

Lastly, we also need a bound on $\mathbb{E}\left|\frac{1}{\mathbb{E}_{\mathbf{x}^1}\frac{1}{\sqrt{z}G_{11}}}\right|^q$ which we obtain in the following Lemma:

Lemma 4.7.3. *Let $E, \eta \in Z_{E,\eta}$, where $z = E + i\eta$. There exist constants $c, C, M > 0$ such that:*

$$\mathbb{E}\left|\frac{1}{\mathbb{E}_{\mathbf{x}^1}\frac{1}{\sqrt{z}G_{11}}}\right|^q \leq C^q,$$

for $N\eta \geq |\sqrt{z}|M$ and for $q \in \mathbb{N}$ with $q \leq c\left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/4}$.

Proof. The proof is similar to Lemma 5.1 in [15]. We define:

$$\widetilde{G_{11}} = \frac{1}{\mathbb{E}_{\mathbf{x}^1}\frac{1}{G_{11}}} = -\frac{1}{z\left(1 + \frac{\text{Tr}\mathcal{G}^{(\{1\})}}{N}\right)}.$$

We calculate that:

$$\left|\frac{d}{d\eta}\log\widetilde{G_{11}}(E + i\eta)\right| = \left|\frac{d}{d\eta}\log\left(\frac{1}{z}\right) + \frac{d}{d\eta}\log\left(\frac{1}{1 + \text{Tr}\frac{\mathcal{G}^{(\{1\})}}{N}}\right)\right| = \left|-\frac{i}{z} - \frac{\frac{d}{d\eta}\text{Tr}\mathcal{G}^{(\{1\})}}{N + \text{Tr}\mathcal{G}^{(\{1\})}}\right|.$$

We show that $\left|\frac{d}{d\eta}\text{Tr}\mathcal{G}^{(\{1\})}\right| \leq \frac{\text{Im}\text{Tr}\mathcal{G}^{(\{1\})}}{\eta}$ as follows:

$$\frac{d}{d\eta}\text{Tr}\mathcal{G}^{(1)} = \sum_{k=1}^{N}\frac{d}{d\eta}\mathcal{G}_{kk}^{(\{1\})}(\theta) = \sum_{k=1}^{N}i((\mathcal{G}^{(\{1\})})^2)_{kk} = \sum_{k=1}^{N}i\langle e_k, (\mathcal{G}^{(\{1\})})^2 e_k\rangle$$

$$\Rightarrow \left|\frac{d}{d\eta}\mathcal{G}^{(\{1\})}\right| \leq \sum_{k=1}^{N}\|(\mathcal{G}^{(\{1\})})^* e_k\|\,\|\mathcal{G}^{(\{1\})}e_k\| \leq \sum_{k=1}^{N}((\mathcal{G}^{(\{1\})})^*\mathcal{G}^{(\{1\})})_{kk}$$

$$= \sum_{k=1}^{N}\frac{\text{Im}(\mathcal{G}^{(\{1\})})_{kk}}{\eta} = \frac{\text{Im}\text{Tr}\mathcal{G}^{(\{1\})}}{\eta}. \tag{4.112}$$

We conclude that

$$\left|\frac{d}{d\eta}\log\widetilde{G_{11}}\right| \leq \frac{1}{|z|} + \frac{\text{Im}\text{Tr}\mathcal{G}^{(\{1\})}}{\eta|N + \text{Tr}\mathcal{G}^{(\{1\})}|} \leq \frac{2}{\eta}, \tag{4.113}$$

yielding that:

$$\left|\log\widetilde{G_{11}}(E + i\eta) - \log\widetilde{G_{11}}(E + i\eta/s)\right| = \left|\int_{\eta/s}^{\eta}\frac{d}{d\nu}\log\widetilde{G_{11}}(E + i\nu)d\nu\right| \leq \int_{\eta/s}^{\eta}\frac{2}{\nu}d\nu = \log s^2 \tag{4.114}$$

and thus

$$|\widetilde{G_{11}}(E + i\eta)| \le s^2 |\widetilde{G_{11}}(E + i\eta/s)|.$$

The proof now proceeds with induction on η just like in the proof of Lemma 4.6.1 using the identity

$$\sqrt{z}\widetilde{G_{11}} = \sqrt{z}G_{11} + \sqrt{z}G_{11}\sqrt{z}\widetilde{G_{11}}(\mathbb{I} - \mathbb{E}_{\mathbf{x}^1})(\sqrt{z}G_{11})^{-1}. \tag{4.115}$$

as well as (4.111) and the results of Lemma 4.6.1. Specifically, we notice that:

$$\mathbb{E}|\sqrt{z}\widetilde{G_{11}}|^q \le C^q + (Cq)^{cq}\mathbb{E}\left(|\sqrt{z}\widetilde{G_{11}}|^{3q}\right)^{1/3}\frac{|z|^{q/2}}{(N\eta)^q},$$

after which the induction is straight-forward just like Lemma 4.6.1. $\square$

Lastly, we use the matrix expansion algorithm to take advantage of the fluctuations. This technique was firstly introduced in the papers [24] and [25] and was later simplified and improved in the papers [19] and [20]. We will follow the analysis in [15].

Hence the following proposition, analogous to Lemma 4.1 of [15]:

Proposition 4.7.4. *Let $\mathcal{E}_q$ be the control parameter as in (4.104). There exist constants $C, M, c_0 > 0$ such that:*

$$\mathbb{E}\left|\frac{1}{N}\sum_k \sqrt{z}\Upsilon^{(\{k\})}\sqrt{z}G_{kk}\right|^{2q} \le (Cq)^{cq^2}\mathcal{E}_{4q}^{1/2}, \tag{4.116}$$

for $1 \le q \le c_0 \left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/8}$, $\frac{N\eta}{\sqrt{|z|}} \ge M$, $K > 0$, $z = E + i\eta \in Z_{E,\eta}$.

Proof. To match notation in [15], we introduce $W_k = \sqrt{z}\Upsilon_k\sqrt{z}G_{kk}$ and we split:

$$\frac{1}{N}\sum_k W_k = \frac{1}{N}\sum_k (\mathbb{I} - \mathbb{E}_k)W_k + \frac{1}{N}\sum_k \mathbb{E}_k W_k. \tag{4.117}$$

By Hölder's inequality,

$$\mathbb{E}\left|\frac{1}{N}\sum_k W_k\right|^{2q} \le C^q\mathbb{E}\left|\frac{1}{N}\sum_k (\mathbb{I} - \mathbb{E}_k)W_k\right|^{2q} + C^q\mathbb{E}|\mathbb{E}_1 W_1|^{2q}. \tag{4.118}$$

To bound the second term in (4.118) above, using that $\sqrt{\theta}\Upsilon^{(\{k\})} = -(\mathbb{I} - \mathbb{E}_k)\frac{1}{\sqrt{z}G_{kk}}$, we obtain

$$\mathbb{E}_k W_k = \frac{\mathbb{E}_k[\sqrt{z}G_{kk}(\sqrt{z}\Upsilon^{(\{k\})})^2]}{\left(\mathbb{E}_k \frac{1}{\sqrt{z}G_{kk}}\right)}. \tag{4.119}$$

and applying Lemma 4.7.3 to (4.119), we get that:

$$\mathbb{E}|\mathbb{E}_1 W_1|^{2q} \leq (\mathbb{E}|\sqrt{z}G_{11}|^{8q})^{\frac{1}{4}} \left(\mathbb{E} \left| \frac{1}{\mathbb{E}_1 \frac{1}{\sqrt{z}G_{11}}} \right|^{8q} \right)^{\frac{1}{4}} (\mathbb{E}|\sqrt{z}\Upsilon^{(\{1\})}|^{8q})^{\frac{1}{2}}$$

$$\leq (Cq)^{cq} (\mathcal{E}_{4q})^{1/2},$$

which is what we want.

In order to handle the first term of (4.118), we use the matrix expansion algorithm as in Section 5.2 of [15]. We notice that equations (5.7), (5.8), and (5.9) are the basis of the expansion algorithm, and they are equivalent to the following (see e.g. (2.18) in [54]):

$$\sqrt{z}G_{ij}^{(\mathbb{T})} = \sqrt{z}G_{ij}^{(\mathbb{T}k)} + \frac{\sqrt{z}G_{ik}^{(\mathbb{T})}\sqrt{z}G_{kj}^{(\mathbb{T})}}{\sqrt{z}G_{kk}^{(\mathbb{T})}} \text{ for } i,j,k \notin \mathbb{T} \text{ and } i,j \neq k,$$

$$\frac{1}{\sqrt{z}G_{ii}^{(\mathbb{T})}} = \frac{1}{\sqrt{z}G_{ii}^{(\mathbb{T}k)}} - \frac{\sqrt{z}G_{ik}^{(\mathbb{T})}\sqrt{z}G_{ki}^{(\mathbb{T})}}{\sqrt{z}G_{ii}^{(\mathbb{T})}\sqrt{z}G_{ii}^{(\mathbb{T}k)}\sqrt{z}G_{kk}^{(\mathbb{T})}} \text{ for } i,k \notin \mathbb{T} \text{ and } i \neq k \tag{4.120}$$

Using the above equation (4.120), we see that in our case the steps of the expansion algorithm (5.13), (5.14), (5.15) in [15] are the same except that each resolvent entry is multiplied by a factor of $\sqrt{z}$. Using our definition of W, equation (5.6) in [15] becomes analogous to

$$(\mathbb{I} - \mathbb{E}_{k_s})W_{k_s} = (\mathbb{I} - \mathbb{E}_{k_s})\left[(\mathbb{I} - \mathbb{E}_{k_s})\frac{1}{\sqrt{z}G_{k_s k_s}}\right]\sqrt{z}G_{k_s k_s}, s = 1, ..., 2q, \tag{4.121}$$

so the initial terms of the algorithm are $A^r := \sqrt{z}G_{k_r k_r}$ and $B^r := \frac{1}{\sqrt{z}G_{k_r k_r}}$ are the same as (5.16), (5.17) of [15] except that each resolvent entry is multiplied by a $\sqrt{z}$. Then (5.18), (5.19), and (5.20) of [15] carry over directly as well as properties (1) through (5) of

relevant strings. We then obtain the desired result

$$\mathbb{E}\left|\frac{1}{N}\sum_k(\mathbb{I}-\mathbb{E}_k)W_k\right|^{2q} \leq (Cq)^{cq^2}\mathcal{E}_{4q}^{1/2}$$

using the proof of (5.32) of [15]. It relies on counting the types of terms that result from the expansion algorithm. Since our algorithm yields the same type and number of terms in each step, the proof in our case will be identical. In [15], we notice the use of bounds (3.9) and Lemma 5.2 in (5.44) as well as in Case 2, bounds (5.26) and (3.4) in (5.43) and (5.49). We can replace (3.9), Lemma 5.2, (5.26), and (3.4) of [15] by our bounds on the relevant quantities in (4.103) as well as our (5.4). □

Proof of Theorem 4.2.1. By Proposition 4.4.1, in order to control Λ, we need to control high moments of $R = N^{-1}\sum\limits_{k=1}^{N} G_{kk}(T_k + \Upsilon^{(\{k\})})$. Taking expectation of $2q$ power we obtain:

$$\mathbb{E}|zR|^{2q} \leq C^q\left(\mathbb{E}\left|\frac{1}{N}\sum_k\sqrt{z}T_k\sqrt{z}G_{kk}\right|^{2q} + \mathbb{E}\left|\frac{1}{N}\sum_k\sqrt{z}\Upsilon^{(\{k\})}\sqrt{z}G_{kk}\right|^{2q}\right). \tag{4.122}$$

For the first term by (4.44), we obtain

$$\mathbb{E}\left|\frac{1}{N}\sum_k\sqrt{z}T_k\sqrt{z}G_{kk}\right|^{2q} \leq C^q\frac{1}{N^{2q}|z|^q}. \tag{4.123}$$

while the second term is handled in Proposition 4.7.4, yielding that

$$\mathbb{E}|zR|^{2q} \leq (Cq)^{cq^2}\mathcal{E}_{4q}^{1/2}.$$

Here we are able to simplify the analysis in [15] by only using the bounds proportional to R from Proposition 4.4.1 to control $E|\Lambda|^{2q}$ in $Z_{E,\eta}$ and $E|\text{Im}\Lambda|^{2q}$. Our simplifications carry over also to the Wigner case. We can assume that

$$[\text{Im}(S_\rho)]^{2q} + \mathbb{E}|\Lambda|^{2q} \geq \frac{1}{(N\eta)^{2q}},$$

(otherwise $\mathbb{E}|\Lambda|^{2q} \leq \frac{1}{(N\eta)^{2q}}$, as we want) and in this case:

$$\mathcal{E}_{2q} = \frac{1}{N^{2q}|z|^q} + \frac{\text{Im}(|z|S_\rho)]^{2q} + \mathbb{E}|z\Lambda|^{2q}}{(N\eta)^{2q}} \leq \frac{\eta^q + \text{Im}(|z|S_\rho)]^{2q} + \mathbb{E}|z\Lambda|^{2q}}{(N\eta)^{2q}},$$

Using the bound proportional to $|R|$ from Proposition 4.4.1, we obtain

$$\begin{aligned}
\mathbb{E}|z\Lambda|^q &\leq \frac{C^q\mathbb{E}|zR|^q}{|S_\rho + \frac{1}{2}|^q} \leq \frac{(Cq)^{cq^2}}{|S_\rho + \frac{1}{2}|^q} \left(\frac{\eta^q + [\text{Im}(|z|S_\rho)]^{2q}}{(N\eta)^{2q}}\right)^{1/2} \\
&= \frac{(Cq)^{cq^2}}{|S_\rho + \frac{1}{2}|^q} \frac{|z|^q}{(N\eta)^q} \left(\frac{\eta^q}{|z|^{2q}} + [\text{Im}(S_\rho)]^{2q}\right)^{1/2} \\
&\leq \frac{(Cq)^{cq^2}|z|^q}{(N\eta)^q} \left[\left(\frac{\sqrt{\eta}}{|z||S_\rho + \frac{1}{2}|}\right)^q + \left(\frac{\text{Im}S_\rho}{|S_\rho + \frac{1}{2}|}\right)^q\right].
\end{aligned}$$

To obtain the desired bound we now note that $\text{Im}S_\rho \leq |S_\rho + \frac{1}{2}|$ and $\frac{\sqrt{\eta}}{|z||S_\rho + \frac{1}{2}|} \leq C$ in our domain. The first one follows easily and for the second one we argue as follows:

$$\frac{\sqrt{\eta}}{|z||S_\rho + \frac{1}{2}|} = \frac{2\sqrt{\eta}}{\sqrt{|z|}\sqrt{|z - 4|}},$$

and by triangle inequality either $|z| \geq 2$ or $|z - 4| \geq 2$. Then in the first case, we use the bound $\sqrt{\eta} \leq \sqrt{|z - 4|}$ and in the second case the bound $\sqrt{\eta} \leq \sqrt{|z|}$.

Overall, this implies that:

$$\mathbb{P}\left(|S_N - S_\rho| \geq \frac{K}{N\eta}\right) \leq \frac{(N\eta)^q}{K^q}\mathbb{E}|\Lambda|^q \leq \frac{(Cq)^{cq^2}}{K^q}, \tag{4.124}$$

for $1 \leq q \leq c_0 \left(\frac{N\eta}{|\sqrt{z}|}\right)^{1/8}$, $\frac{N\eta}{\sqrt{|z|}} \geq M$, $K > 0$, $z = E + i\eta \in Z_{E,\eta}$. $\qquad\square$

4.8   Rate of convergence to the Marchenko-Pastur distribution

In this section we prove Theorem 4.2.2.

Let

$$\mathcal{N}(I) := \#\{i \leq N \mid \lambda_i \in I\}, \tag{4.125}$$

denote the number of eigenvalues in the interval I. We denote by ν the probability distribution of $\text{Re}(x_{ij})$ and $\text{Im}(x_{ij})$, for $i, j = 1, ..., N$.

Proof of Theorem 4.2.2. Let $0 < E \leq 4$. We will use a Pleijel argument from [55], recently used in obtaining estimates on a measure μ from estimates on its Stieltjes transform as in [23]. We start from the following equations (equations (13) and (14) in [23], following from equation (5) of [55]). Here μ is any probability measure supported on an interval $[-K, K]$, for a large $K > 0$:

$$\mu(-K, E) = \frac{1}{2\pi i} \int_{L(z_0)} S_\mu(z)dz + \frac{\eta_0}{\pi} \text{Re} S_\mu(z_0) + \mathcal{O}\left(\eta_0 \text{Im} S_\mu(z_0)\right) \qquad (4.126)$$

and

$$\mu(x, x') = \frac{1}{2\pi i} \int_{\gamma(x,x')} S_\mu(z)dz + \mathcal{O}\left(\eta_0 \left|S_\mu(x + i\eta_0)\right| + \eta_0 \left|S_\mu(x' + i\eta_0)\right|\right), \qquad (4.127)$$

where S_μ is the Stieltjes transform of μ and $L(z_0)$ is a contour as in Figure 4.1 (see also [23] Fig 1A), namely connects with line segments the points $E - i\eta_0, E - iQ, -1 - iQ, -1 + iQ, E + iQ, E + i\eta_0$ in that order with arbitrarily chosen constants -1 and Q, and $\gamma(x, x')$ is the chain connecting $x + i\eta_0$ with $x' + i\eta_0$ and $x' - i\eta_0$ with $x - i\eta_0$.

By Markov inequality, we obtain that

$$\mathbb{P}\left(\left|\rho_N(E) - P(E)\right| \geq \frac{C \log N}{N}\right) \leq \frac{N^q \mathbb{E}\left|\rho_N(E) - P(E)\right|^q}{(C \log N)^q}. \qquad (4.128)$$

Then using (4.126) and taking $z_0 := E + i\eta_0$ with $\eta_0 := \frac{M\sqrt{E}}{N}$ with M as in Theorem 4.2.1, we obtain that:

$$\mathbb{E}\left|\rho_N(E) - P(E)\right|^q = \mathbb{E}\left|\frac{1}{2\pi i} \int_{L(z_0)} \Lambda(z)dz + \frac{\eta_0}{\pi}\text{Re}\Lambda(z_0) + \mathcal{O}\left(\eta_0(\text{Im}S_N(z_0) + \text{Im}S_\rho(z_0))\right)\right|^q$$

$$\leq C^q \left(\mathbb{E}\left|\int_{L(z_0)} \Lambda(z)dz\right|^q + \mathcal{O}\left(\eta_0^q \mathbb{E}|\Lambda(z_0)|^q + \eta_0^q \text{Im}S_\rho(z_0)^q\right)\right), \qquad (4.129)$$

noting that the constant in the $\mathcal{O}$ comes from the Pleijel formula and is uniform in the matrix randomness. We study the above expression one term at a time. For $E \leq 4$ we

can bound the second term as follows

$$\eta_0^q \mathbb{E}|\Lambda(z_0)|^q \leq \eta_0^q \frac{(Cq)^{cq^2}}{(N\eta_0)^q} \leq \frac{(Cq)^{cq^2}}{N^q}. \tag{4.130}$$

The third term is bounded using the above inequality (4.130) on Λ as well as

$$\eta_0 \mathrm{Im} S_\rho \leq \frac{C\eta_0}{\sqrt{E}} = \frac{CM}{N}. \tag{4.131}$$

Now for the integral, we note that it suffices to study the part of the contour where $\mathrm{Im}(z) > 0$ since $\Lambda(\bar{z}) = \overline{\Lambda(z)}$. Thus we obtain:

$$\mathbb{E}\left|\int_{L(z_0)} \Lambda(z)dz\right|^q \leq C^q \left(\mathbb{E}\left|\int_0^{\eta_0} \Lambda(-1+iy)dy\right|^q \right.$$
$$\left. + \mathbb{E}\left|\int_{\eta_0}^Q \Lambda(-1+iy) - \Lambda(E+iy)dy\right|^q + \mathbb{E}\left|\int_{-1}^E \Lambda(x+iQ)dx\right|^q \right)$$

Since all eigenvalues are positive we bound $|\Lambda|$ for $-1 < 0$ by $|\Lambda(-1+i\eta)| \leq 2$ which yields:

$$\left|\int_0^{\eta_0} \Lambda(-1+iy)dy\right|^q \leq \left(\int_0^{\eta_0} |\Lambda(-1+iy)|dy\right)^q \leq C^q\eta_0^q. \tag{4.132}$$

Next we note that:

$$\mathbb{E}\left|\int_{-1}^E \Lambda(x+iQ)dx\right|^q \leq \frac{(Cq)^{cq^2}}{(NQ)^q} \tag{4.133}$$

Now we can bound the expected value of the integrals $\mathbb{E}\left(\int_{\eta_0}^Q |\Lambda(E+iy)|dy\right)^q$ and $\mathbb{E}\left(\int_{\eta_0}^Q |\Lambda(-1+iy)|dy\right)^q$ for $E \leq 4$, noting that the argument is identical at E and -1,

$$\mathbb{E}\left(\int_{\eta_0}^Q |\Lambda(E+iy)|dy\right)^q = \mathbb{E}\int_{\eta_0}^Q |\Lambda(E+iy_1)|dy_1 \int_{\eta_0}^Q |\Lambda(E+iy_2)|dy_2 \cdots \int_{\eta_0}^Q |\Lambda(E+iy_q)|dy_q$$
$$= \mathbb{E}\int_{\eta_0}^Q \cdots \int_{\eta_0}^Q \prod_{j=1}^q |\Lambda(E+iy_j)| \prod_{j=1}^q dy_j = \int_{\eta_0}^Q \cdots \int_{\eta_0}^Q \mathbb{E}\prod_{j=1}^q |\Lambda(E+iy_j)| \prod_{j=1}^q dy_j$$
$$\leq \int_{\eta_0}^Q \cdots \int_{\eta_0}^Q \prod_{j=1}^q (\mathbb{E}|\Lambda(E+iy_j)|^q)^{\frac{1}{q}} \prod_{j=1}^q dy_j \leq \frac{1}{N^q} \int_{\eta_0}^Q \cdots \int_{\eta_0}^Q \prod_{j=1}^q \frac{(Cq)^{cq}}{y_j} \prod_{j=1}^q dy_j$$
$$= \frac{(Cq)^{cq^2}}{N^q} \left(\int_{\eta_0}^Q \frac{1}{y}dy\right)^q \leq (Cq)^{cq^2} \frac{(\log N)^q}{N^q},$$

where we can apply (4.124) inside the integral because our estimates on Λ are uniform

on compact sets.

To prove the other part of (4.32) near the "hard" edge, we use the (4.127) and study the interval $[-E, E]$, noting that $\rho_N(E) = \mathcal{N}([-E, E])/N$ and $P(-E, E) = P(E)$. Similarly to the above, we have that:

$$\rho_N(-E, E) - \rho(-E, E) = \frac{1}{2\pi i} \int_{\gamma(-E,E)} \Lambda(z)dz + \mathcal{O}\left(\eta_0 \Lambda(E + i\eta_0)\right) + \mathcal{O}\left(\eta_0 S_\rho(E + i\eta_0)\right).$$

We now take expectation and power q and use Hölders inequality. The corresponding integral can be bounded similarly to above:

$$\mathbb{E}\left|\int_{\gamma(-E,E)} \Lambda(z)dz\right|^q = \mathbb{E}\left|\int_{-E}^{E} \Lambda(x + i\eta_0) - \Lambda(x - i\eta_0)\, dx\right|^q$$

$$= \mathbb{E}\left|\int_{-E}^{E} 2\mathrm{Im}\Lambda(x + i\eta_0)\right|^q \leq \frac{(Cq)^{cq^2} E^q}{(N\eta_0)^q} \leq \frac{(Cq)^{cq^2}(\sqrt{E})^q}{M^q}$$

and, similarly to (4.131):

$$\eta_0 S_\rho(E) \leq \frac{\eta_0}{\sqrt{E}} = \frac{M}{N}, \tag{4.134}$$

which together with (4.130) yields the $\sqrt{E}$ - bound of (4.32) for $E < 4$.

To establish the (4.32) for $E > 4$, we use (4.32) for $E = 4$ to establish bounds on the number of eigenvalues outside of the spectrum. Letting $\mathcal{N}_I$ be the number of eigenvalues in an interval I, we see that:

$$\mathcal{N}_{(4,\infty)} = N - N\rho_N(4) = N(P(4) - \rho_N(4)) \tag{4.135}$$

which by (4.32) for $E = 4$ yields that:

$$\mathbb{P}\left(\frac{\mathcal{N}_{(4,\infty)}}{N} > \frac{K \log N}{N}\right) \leq \frac{(Cq)^{q^2}}{K^q} \tag{4.136}$$

and for $E > 4$,

$$\mathbb{P}\left(|\rho_N(E) - P(E)| \geq \frac{K \log N}{N}\right) \leq \mathbb{P}\left(\frac{\mathcal{N}_{(4,\infty)}}{N} > \frac{K \log N}{N}\right), \tag{4.137}$$

thus (4.136) gives the desired bound. $\qquad\square$

4.9 Rigidity of the eigenvalues

The aim of this section is a proof of Theorem 4.2.3.

Proof of Theorem 4.2.3. Let $i \le \frac{N}{2}$. We will make use of the following inequalities near the "hard" edge and away from the "soft" edge:

$$c\sqrt{x} \le P(x) \le C\sqrt{x}, \tag{4.138}$$

and

$$cP(x)^{-1} \le \rho(x) \le CP(x)^{-1}. \tag{4.139}$$

valid for $x \in (0, 3]$. The second inequality implies that

$$c\frac{N}{i} \le \rho(\gamma_i) \le C\frac{N}{i}, \tag{4.140}$$

for any $i \le \frac{N}{2}$.

For $\epsilon > 0$, we have that:

$$
\begin{aligned}
&\mathbb{P}\left(|\lambda_i - \gamma_i| \ge K\epsilon\frac{i}{N}\right) \\
&\le \mathbb{P}\left(|\lambda_i - \gamma_i| \ge K\epsilon\frac{i}{N} \text{ and } \lambda_i \le \gamma_i\right) + \mathbb{P}\left(|\lambda_i - \gamma_i| \ge K\epsilon\frac{a}{N} \text{ and } \lambda_i > \gamma_i\right) \\
&= A + B.
\end{aligned}
\tag{4.141}
$$

We consider first the term A. We set

$$\ell = K\varepsilon\frac{i}{N}.$$

From $\lambda_i \le \gamma_i$ and $|\lambda_i - \gamma_i| \ge \ell$ we find that $\lambda_i \le \gamma_i - \ell$. This implies that $\rho_N(\gamma_i - \ell) \ge \frac{i}{N} = P(\gamma_i)$. By the mean value theorem for the function P, there exists a point $x^* \in [\gamma_i - \ell, \gamma_i]$

such that $P(\gamma_i) - P(\gamma_i - \ell) = \rho(x^*)\ell$, yielding that:

$$\rho_N(\gamma_i - \ell) - P(\gamma_i - \ell) = \rho_N(\gamma_i - \ell) - P(\gamma_i) + \rho(x^*)\ell \geq \rho(x^*)\ell$$

$$\geq \rho(\gamma_i)K\epsilon\frac{i}{N} \geq cK\epsilon, \tag{4.142}$$

because ρ is non-increasing, $i < N/2$, and we used (4.140). Setting $\varepsilon = \frac{\log N}{N}$ we deduce from Theorem 4.2.2 that:

$$A \leq \mathbb{P}\left(|\rho_N(\gamma_i - \ell) - P(\gamma_i - \ell)| \geq \frac{cK\log N}{N}\right) \leq \frac{(Cq)^{cq^2}}{K^q} \tag{4.143}$$

For $i \leq \log N$, set $\epsilon = \frac{i}{N} \geq c\sqrt{\gamma_i}$, from (4.138), to obtain the better bound:

$$A \leq \mathbb{P}\left(|\rho_N(\gamma_i - \ell) - P(\gamma_i - \ell)| \geq cK\sqrt{(\gamma_i - \ell)_+}\right) \leq \frac{(Cq)^{cq^2}}{K^q}, \tag{4.144}$$

because $c\sqrt{(\gamma_i - \ell)_+} \leq c\sqrt{\gamma_i} \leq \frac{i}{N} \leq \frac{\log N}{N}$ and we used Theorem 4.2.2.

We now estimate the term B. From the estimate (4.138) near the "hard" edge we have that $P(x) \sim \sqrt{x}$, so:

$$\gamma_i \leq C\left(\frac{i}{N}\right)^2,$$

for some constant $C > 0$ for all $i < N/2$. We consider the number

$$y = 2C\left(\frac{i}{N}\right)^2$$

and we further consider the cases that $\gamma_i + \ell \leq y$ or $\gamma_i + \ell > y$.

In the first case, since $\lambda_i > \gamma_i$ and $|\lambda_i - \gamma_i| \geq \ell$, we have that $\lambda_i > \gamma_i + \ell$ and so $\rho_N(\gamma_i + \ell) \leq \frac{i}{N} = P(\gamma_i)$.

Hence, from the mean value theorem, we find $x^* \in [\gamma_i, \gamma_i + \ell] \subset [\gamma_i, y]$ such that $P(\gamma_i + \ell) - P(\gamma_i) = \rho(x^*)\ell$, yielding that:

$$P(\gamma_i + \ell) - \rho_N(\gamma_i + \ell) = P(\gamma_i) - \rho_N(\gamma_i + \ell) + \rho(x^*)\ell$$

$$\geq \rho(x^*)\ell = \rho(x^*)K\epsilon\frac{i}{N} \geq \rho(y)K\epsilon\frac{i}{N} \geq cK\epsilon, \tag{4.145}$$

where we used that ρ is non-increasing and that $\rho(y) \geq \frac{c}{\sqrt{y}}$ near the "hard edge".

Setting $\epsilon = \frac{\log N}{N}$ and using Theorem 4.2.2, we conclude that

$$B \leq \mathbb{P}\left(|P(\gamma_i + \ell) - \rho_N(\gamma_i + \ell)| \geq cK\frac{\log N}{N}\right) \leq \frac{(Cq)^{cq^2}}{K^q}, \qquad (4.146)$$

as required. To obtain rigidity at the "hard" edge (equation (4.34), let $\epsilon = \frac{i}{N}$ to obtain:

$$B \leq \mathbb{P}\left(|P(\gamma_i + \ell) - \rho_N(\gamma_i + \ell)| \geq \frac{cKi}{N}\right)$$
$$\leq \mathbb{P}\left(|P(\gamma_i + \ell) - \rho_N(\gamma_i + \ell)| \geq c\sqrt{K}\sqrt{\gamma_i + \ell}\right) \leq \frac{(Cq)^{cq^2}}{K^{q/2}}, \qquad (4.147)$$

where the second line follows as before because $\sqrt{\gamma_i} \leq c\frac{i}{N}$ and $\ell = \frac{Ki^2}{N^2}$.

In the other case we have that $\gamma_i + \ell > y$ so the inequality $\lambda_i > \gamma_i + \ell$ implies that $\lambda_i > y$ and therefore $\rho_N(y) \leq \frac{i}{N} = P(\gamma_i)$. Hence from the mean value theorem, there exists $x^* \in [\gamma_i, y]$ such that $P(y) - P(\gamma_i) \geq \rho(x^*)\ell$, which yields:

$$P(y) - \rho_N(y) \geq P(\gamma_i) - \rho_N(y) + \rho(x^*)\ell \geq \rho(x^*)\ell = \rho(x^*)K\epsilon\frac{i}{N} \geq \rho(y)K\epsilon\frac{i}{N} \geq cK\epsilon,$$

and we can conclude (4.33) and (4.34) as above. This finishes the proof of Theorem 4.2.3. □

Chapter 5

The Dyson equation method

In this chapter we generalize the Stieltjes transform technique to the Dyson equation technique.

Let H be a $N \times N$ Hermitian random matrix, whose spectrum we want to compute for large N. We recall that with the Stieltjes transform method we had to define the *resolvent* matrix $G(z) = (H - zI)^{-1}$ and then the Stieltjes transform of the empirical spectral measure would be equal to the normalized trace of $G(z)$. We then had to show that the normalized trace of $G(z)$ convergences to a suitable Stieljes tranform $S_\mu(z)$ of a limiting measure μ. Usually, $S_\mu(z)$ satisfies an algebraic equation which is "close" to the probabilistic one satisfied by $\langle G(z) \rangle$, which can be found using resolvent identities.

In the Dyson equation technique, instead of studying the normalized trace $\langle G(z) \rangle$ of the resolvent matrix, we study the whole matrix $G(z)$ and deduce a basic probabilistic equation satisfied by it in the Gaussian case which we generalize for any distribution for the entries. If some conditions are satisfied about the initial Hermitian random matrix H, then we can simplify this equation to deduce a deterministic one which is "close" to the original and whose determistic solution $M(z)$ would be "close" to $G(z)$. Analogously, $M(z)$ turns out be the Stieltjes transform of a *matrix-valued* measure, whose normalized trace $\langle M(z) \rangle$ can give us the liming spectral measure for the eigenvalues of H.

This technique is especially useful when we have correlations inside the Hermitian matrix H. In section 5.1 we will apply this technique to the Gaussian Unitary Ensemble (GUE) and find out that it coincides with the Stieltjes transform technique. We then find out some differences for Wigner-type matrices and correlated Hermitian matrices in section 5.2.

The technique is then studied and established theoretically in the next sections. There are two steps for this method to work, after proving the existence and uniqueness of the solution $M(z)$:

1. Show that the probabilistic equation satisfied by $G(z)$ and the deterministic equation satisfied by $M(z)$ are "close" to each other. This is done by treating the first as a perturbation of the second one and showing that the error is small. This is achieved through the *multivariate cumulant expansion* technique and is summarized in 5.3.

2. Show that the deterministic equation is *stable*. This means that solutions to it are "close" to solutions of small perturbations of it. This is achieved by showing the invertibility of a suitable *linear stability operator*. This is summarized in section 5.4.

We finally give the resulting theorems in section 5.5.

This analysis is based on the lecture notes [26]. The publications [21] and [3] analyze these two steps and give the final theorems, while the existence and uniqueness of the solution for the Dyson equation can be found in [37] for the matrix case and in [1] for the vector case.

5.1 Gaussian Hermitian random matrix models

The definition of the resolvent $G = G(z)$ results in the basic identity:

$$HG = I + zG. \tag{5.1}$$

We will include in this identity the information provided by the expected value of HG, which reminds of the formula (for Gaussian random variables):

$$\mathbb{E}[hf(h)] = \mathbb{E}[h^2]\mathbb{E}[f'(h)], \tag{5.2}$$

where h is real centered Gaussian and f a smooth function. The version of this for complex Gaussian random variables is the following:

$$\mathbb{E}[zf(z)] = \mathbb{E}\|z\|^2 \mathbb{E}\left[\overline{\partial}f(z)\right], \tag{5.3}$$

where $\overline{\partial}$ denotes the Wirtinger derivative and f is sufficiently smooth. This gets generalized for Gaussian random vectors $V \in \mathbb{R}^n$ as follows:

$$\mathbb{E}[Vf(V)] = K\mathbb{E}\left[\nabla f(V)\right], \tag{5.4}$$

where $K = \mathbb{E}[VV^T]$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is sufficiently smooth.

The application of this formula to resolvent matrices in random matrices firstly appeared in [42]. In our case, where f represents the resolvent function of a complex centered Gaussian matrix, we have the following:

Lemma 5.1.1. *For a Hermitian Gaussian random matrix H and its resolvent G, it holds that:*

$$\mathbb{E}[HG] = -\mathbb{E}\left[\tilde{\mathbb{E}}[\tilde{H}G\tilde{H}]G\right], \tag{5.5}$$

where $\tilde{H}$ is an independent copy of H and the second expectation is with respect to $\tilde{H}$, i.e. $\tilde{\mathbb{E}}[\tilde{H}G\tilde{H}] = \mathbb{E}\left[HGH|G\right].$

Proof. For the proof, we apply (5.4), while identifying $\mathbb{C}$ with $\mathbb{R}^2$. We obtain that:

$$\begin{aligned}
&\mathbb{E}[H_{jk}G_{ab}] \\
&= \mathbb{E}\left[K_{11}\partial_{\mathrm{Re}(H_{jk})}G_{ab} + K_{12}\partial_{\mathrm{Im}(H_{jk})}G_{ab} + iK_{21}\partial_{\mathrm{Re}(H_{jk})}G_{ab} + iK_{22}\partial_{\mathrm{Im}(H_{jk})}G_{ab}\right] \\
&= -\mathbb{E}\left[(K_{11} - K_{22} + iK_{12} + iK_{21})G_{aj}G_{kb} + (K_{11} + K_{22} - iK_{12} + iK_{21})G_{ak}G_{jb}\right],
\end{aligned}$$

where K is the real $2N \times 2N$ covariance matrix of the k-th column of H which belongs in $\mathbb{C}^N$.

We observe that $K_{11} = K_{22} = \mathbb{E}[\mathrm{Re}H_{jk}]^2 = \mathbb{E}[\mathrm{Im}H_{jk}]^2$ and

$$K_{12} = K_{21} = \mathbb{E}[\mathrm{Re}H_{jk}\mathrm{Im}H_{jk}] = 0.$$

Therefore, we have that:

$$\mathbb{E}[H_{jk}G_{ab}] = -\mathbb{E}\left[\|H_{jk}\|^2 G_{ak}G_{jb}\right] = -\mathbb{E}\|H_{12}\|^2\mathbb{E}\left[G_{ak}G_{jb}\right], \tag{5.6}$$

for any $j, k, a, b \in \{1, ..., N\}$. Calculating the desired expected value, we see that:

$$\tilde{\mathbb{E}}[\tilde{H}G\tilde{H}]_{jk} = \sum_{a,b} \tilde{\mathbb{E}}[\tilde{H}_{ja}G_{ab}\tilde{H}_{bk}] = \sum_{a,b} G_{ab}\tilde{\mathbb{E}}[\tilde{H}_{ja}\tilde{H}_{bk}]$$
$$= \delta_{jk}\mathbb{E}\|H_{12}\|^2\sum_{a=1}^{N} G_{aa} = \delta_{jk}\mathbb{E}\|H_{12}\|^2\mathrm{Tr}[G], \tag{5.7}$$

and this means that:

$$\mathbb{E}\left[\tilde{\mathbb{E}}[\tilde{H}G\tilde{H}]G\right]_{jk} = \mathbb{E}\|H_{12}\|^2\mathbb{E}\left[G_{jk}\mathrm{Tr}(G)\right].$$

On the other hand, by (5.6), we have that:

$$\mathbb{E}[HG]_{jk} = \sum_{a=1}^{N} \mathbb{E}\left[H_{ja}G_{ak}\right] = -\sum_{a=1}^{N} \mathbb{E}\|H_{12}\|^2\mathbb{E}\left[G_{aa}G_{jk}\right]$$
$$= -\mathbb{E}\|H_{12}\|^2\mathbb{E}\left[G_{jk}\mathrm{Tr}(G)\right],$$

and the proof is complete. $\square$

We can now define a linear operator $S : \mathbb{C}^{N \times N} \mapsto \mathbb{C}^{N \times N}$, by:

$$S[R] = \tilde{\mathbb{E}}[\tilde{H}R\tilde{H}], \tag{5.8}$$

so that we may write:

$$\mathbb{E}[HG] = -\mathbb{E}[S[G]G]. \tag{5.9}$$

Plugging this information to (5.1) we obtain the *Dyson equation*:

$$I + (z + S[G])G = D, \quad D := HG + S[G]G, \tag{5.10}$$

where S defined in (5.8) is called the *self-energy* operator in the Dyson equation and D is called the *error matrix*. Observe that $\mathbb{E}[D] = 0$, so that we can always work with the probabilistically "close" equation:

$$I + (z + S[M])M = 0. \tag{5.11}$$

In the GUE ensemble we have that $K_{11} = K_{22} = \frac{1}{2N}$ and of course $K_{12} = K_{21} = 0$. This gives the following expression (see (5.7)) for the self-energy operator:

$$(S[R])_{jk} = \tilde{\mathbb{E}}[\tilde{H}R\tilde{H}]_{jk} = \delta_{jk}\frac{1}{N}\text{Tr}(R), \tag{5.12}$$

so that the Dyson equation becomes:

$$I + \left(z + \frac{1}{N}\text{Tr}(G)\right)G = 0. \tag{5.13}$$

Taking normalized trace this is identical to the Stieltjes transform equation:

$$1 + (z + S_N)S_N = 0. \tag{5.14}$$

5.2 Wigner-type and correlated Hermitian models

In the Wigner-type model we still assume that the entries of H are independent up to the Hermitian symmetry, but we drop the identically-distributed condition. We define

the variances of each random entry h_{ij} as:

$$s_{ij} = \mathbb{E}\|h_{ij}\|^2, \tag{5.15}$$

which correspond to the matrix of variances S with $(S)_{ij} := s_{ij}$. In the Wigner model, we have that $s_{ij} = \frac{1}{N}$ but here we just impose the condition that:

$$\frac{c}{N} \leq s_{ij} \leq \frac{C}{N}, \tag{5.16}$$

uniformly for each $i, j = 1, ..., N$ and some $c, C > 0$.

We calculate the self-energy operator for this model as follows:

$$\tilde{\mathbb{E}}[\tilde{H}R\tilde{H}]_{jk} = \sum_{a,b} R_{ab}\tilde{\mathbb{E}}[\tilde{H}_{ja}\tilde{H}_{bk}]$$

$$= \delta_{jk}\sum_{a=1}^{N} R_{aa}s_{aj} = \delta_{jk}\langle S_j, \mathrm{diag}(R)\rangle, \tag{5.17}$$

where S_j is the j-th column of S and $\mathrm{diag}(R)$ is the vector corresponding to the diagonal of R.

For the Dyson equation, by using (5.17), we now have to consider a system of N equations for the unknown vector $m = (m_1, ..., m_N)$ which represents the diagonal of G:

$$1 + (z + \langle S_j, m\rangle)m_j = 0, \tag{5.18}$$

for $j = 1, ..., N$.

For the stability analysis, the linear stability operator is given by:

$$I - m^2 S: \ \mathbb{C}^N \mapsto \mathbb{C}^{\mathbb{N}},$$

with:

$$\left[(I - m^2 S)(x)\right]_j := x_j - m_j^2 \sum_{a=1}^{N} s_{ja}x_a. \tag{5.19}$$

We will see a proof of this in section 5.4 where the linear stability operator is given in its general form.

In the correlated Hermitian model, we drop the independence condition in the matrix entries of H, so that the matrix elements may have non-trivial correlations in addition to the one required by the Hermitian symmetry.

This is the most general model and the one mostly used in applications where we use the *Hermitization* technique to "hermitize" the matrix and then apply the Dyson equation technique. This is what we will do with the model of the next chapter, see section 6.1.

The self-energy operator here is in its most general form:

$$(S[R])_{ij} = \mathbb{E}[HRH]_{ij} = \sum_{a,b} \mathbb{E}[H_{ia}H_{bj}]R_{ab},$$

and we have to use the analogue of (5.16) which is:

$$c\langle R \rangle \preceq S[R] \preceq C\langle R \rangle, \tag{5.20}$$

where $\langle \, \cdot \, \rangle$ represents the normalized trace of a matrix and $A \preceq B$ means that the matrix $B - A$ is positive semi-definite.

Conditions (5.16) and (5.20) are known as *mean-field* conditions which make the spectrum map into an interval or area of order 1.

The Dyson equation is in its general form:

$$I + (z + S[G])G = 0,$$

while the linear stability operator is also in its general form and given by:

$$I - C_M \circ S : \quad \mathbb{C}^{N \times N} \mapsto \mathbb{C}^{N \times N}, \tag{5.21}$$

where C_M is the "sandwich" operator $C_M[R] = MRM$, and M is the solution matrix, so that the linear stability operator becomes:

$$R \mapsto R - MS[R]M. \tag{5.22}$$

Notice that this form implies (5.19) for the vector Dyson equation. A summary for the proof of (5.22) will be given in the deterministic stability analysis in section 6.7.

5.3   The multivariate cumulant expansion

In this section we give a summary of the proof that the probabilistic Dyson equation is "close" to the original one, which is treated as a perturbation of the probabilistic one with error given by the matrix D:

$$I + (z + S[M])M = 0, \quad I + (z + S[M])M = D.$$

In order to prove this we have to bound expectations of high powers of the quantity $|D_{ij}|$, for any $i, j = 1, ...N$.

We will follow the multivariate cumulant expansion technique which was given in [21] and involves more general matrix cases with slow correlation decay.

Another approach for the proof was given in [3] for correlated matrices with fast exponential correlation decay. There they used the Schur complement method together with concentration estimates on quadratic functionals of independent or essentially independent random vectors.

Let:

$$D = HG + S[G]G. \tag{5.23}$$

We extend formula (5.2) to non-Gaussian random variables as follows:

Lemma 5.3.1. *Let h be a general non-Gaussian real random variable such that all its moments and cumulants exist. Then for $f \in C^\infty$, we have that:*

$$\mathbb{E}[hf(h)] = \sum_{k=0}^\infty \frac{\kappa_{k+1}}{k!} \mathbb{E}[f^{(k)}(h)], \tag{5.24}$$

where κ_k is the k-th cumulant of h.

We recall that the cumulants of a real random variable are defined by the relation:

$$\log \mathbb{E}[e^{th}] = \sum_{k=0}^\infty \frac{\kappa_k}{k!} t^k, \tag{5.25}$$

which is similar to the moment-generating formula:

$$\mathbb{E}[e^{th}] = \sum_{k=0}^\infty \frac{m_k}{k!} t^k, \tag{5.26}$$

and so the moments can define the cumulants and the cumulants can define the moments, by comparing the two power series.

Proof of (5.24). For the proof, we can use the Fourier transform of f,

$$\hat{f}(t) = \int_\mathbb{R} f(x) e^{itx} dx$$

and then differentiate to get that:

$$\hat{f}'(t) = i \int_\mathbb{R} x f(x) e^{itx} dx,$$

so by the Fourier-inversion formula for the function $xf(x)$ we have that:

$$xf(x) = i \int_\mathbb{R} \overline{\hat{f}'(t)} e^{itx} dt \Rightarrow \mathbb{E}[xf(x)] = i \int_\mathbb{R} \overline{\hat{f}'(t)} \hat{\mu}(t) dt, \tag{5.27}$$

where we defined the measure μ to be the distribution of the real random variable h and

$$\hat{\mu}(t) = \int_\mathbb{R} e^{itx} d\mu(x) = \mathbb{E}[e^{ith}].$$

We then perform integration by parts on the RHS of (5.27) to get that:

$$i \int_{\mathbb{R}} \overline{\hat{f}'(t)} \hat{\mu}(t) dt = -i \int_{\mathbb{R}} \overline{\hat{f}(t)} \hat{\mu}'(t) dt = -i \int_{\mathbb{R}} \overline{\hat{f}(t)} \hat{\mu}(t) (\log \hat{\mu}(t))' dt$$

$$= \sum_{k=0}^{\infty} \frac{\kappa_{k+1}}{k!} \int_{\mathbb{R}} (it)^k \overline{\hat{f}(t)} \hat{\mu}(t) dt = \sum_{k=0}^{\infty} \frac{\kappa_{k+1}}{k!} \mathbb{E}[f^{(k)}(h)],$$

where in the last step we used the Fourier-inversion formula once more for f and differentiated it k times. $\qquad \square$

In order to apply (5.24) to random matrices, we have to define the *joint cumulants* for a family of random variables, such as a random matrix.

If $\mathbf{h} = (h_1, h_2, \ldots h_m)$ is a collection of random variable, then

$$\kappa(\mathbf{h}) = \kappa(h_1, h_2, \ldots h_m)$$

are the coefficients of the logarithm of the moment-generating function:

$$\log \mathbb{E}[e^{\mathbf{t} \cdot \mathbf{h}}] = \sum_{\mathbf{k}} \frac{\mathbf{t}^{\mathbf{k}}}{\mathbf{k}!} \kappa_{\mathbf{k}}, \tag{5.28}$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n$ and $\mathbf{k} = (k_1, k_2, \ldots, k_n) \in \mathbb{N}^n$ is a multi-index with n components and:

$$\mathbf{t}^{\mathbf{k}} := \prod_{i=1}^{n} \mathrm{t}_{\mathrm{i}}^{k_i,}, \quad \mathbf{k}! := \prod_{i=1}^{n} k_i!, \quad \kappa_{\mathbf{k}} = \kappa(\mathrm{h}_1, \ \mathrm{h}_1, \ldots \mathrm{h}_2, \ \mathrm{h}_2, \ldots),$$

where h_j appears k_j-times. The analogue of (5.24) is:

$$\mathbb{E}[h_1 f(\mathbf{h})] = \sum_{\mathbf{k}} \frac{\kappa_{\mathbf{k}+\mathbf{e}_1}}{\mathbf{k}!} \mathbb{E}[f^{(\mathbf{k})}(\mathbf{h})], \quad \mathbf{h} = (\mathrm{h}_1, \ \mathrm{h}_2, \ldots, \mathrm{h}_n), \tag{5.29}$$

where $f^{(\mathbf{k})} = \frac{\partial^{k_1} \ldots \partial^{k_n}}{\partial h_1 \ldots \partial h_n} f$ and the summation is for all n multi-indices with:

$$\mathbf{k} + \mathbf{e}_1 = (k_1 + 1, k_2, \ldots, k_n).$$

The proof is similar with the proof of (5.24).

Now, back to the definition of the error matrix D in (5.23), we will use this multivariate cumulant expansion (5.29) to prove that the expectation of high powers of the quantity $|D_{ij}|$ is small. This is done by writing:

$$\mathbb{E}|D_{ij}|^{2p} = \mathbb{E}\left[(HG + \mathcal{S}[G]G)_{ij}D_{ij}^{p-1}\overline{D_{ij}^p}\right] \tag{5.30}$$

and then use (5.29) to do an integration by parts in the first H factor, while considering everything else as a function $f(H)$. This involves some heavy analysis of cumulant expansions and combinatorics, which can be found in [21].

The authors there arrive at the result of [Theorem 4.1] in [21], which takes into account a couple of different possible random matrix norms for the error matrix D. A simple corollary is for the following maximum-norm:

Theorem 5.3.2 (Bound on the error matrix D). *Under the mean-field condition* (5.20) *and a finite-moments condition* (5.33) *for the matrix H, we have that for any $\gamma, \epsilon, D > 0$ and some bounded $z = E + i\eta$ with $\eta \geq N^{-1+\gamma}$, there exists $C > 0$ such that:*

$$\mathbb{P}\left(\|D(z)\|_{\max} \geq \frac{N^\epsilon}{\sqrt{N\eta}}\right) \leq \frac{C}{N^D}. \tag{5.31}$$

Here the maximum-norm of a matrix T is defined by:

$$\|T\|_{\max} := \max_{i,j}|T_{ij}|, \tag{5.32}$$

while the finite-moments condition can be described by:

$$\max_{i,j}\mathbb{E}|\sqrt{N}H_{ij}|^p \leq \mu_p, \tag{5.33}$$

for a sequence of constants μ_p, where H is our initial Hermitian matrix.

5.4 The deterministic stability step

In this section we give a summary for the proof of the stability of the deterministic equation:

$$I + (z + \mathcal{S}[M])M = 0. \tag{5.34}$$

Since the solution M of the equation (5.34) as well as the solution M' of its perturbed version (5.10) can be expressed as a function of the variable z, to prove its stability it is enough to control the derivative:

$$\partial_z M = \frac{\partial M(z)}{\partial z},$$

therefore we differentiate (5.34) to get a formula for the derivative of the solution:

$$(I + \mathcal{S}[\partial_z M])M + (z + \mathcal{S}[M])\partial_z M = 0 \Rightarrow$$

$$M^2 + M\mathcal{S}[\partial_z M]M - \partial_z M = 0 \Leftrightarrow$$

$$\partial_z M - M\mathcal{S}[\partial_z M]M = M^2 \Leftrightarrow$$

$$(I - C_M \circ S)[\partial_z M] = M^2, \tag{5.35}$$

where C_M is the "sandwich" operator defined in (5.21).

If we now show that the "linear stability operator" $I - C_M \circ S$ is invertible, then the derivative $\partial_z M$ will have reasonable behaviour, which means that the equation (5.34) will be stable to perturbations.

To show invertibility, we have to bound the norm of the inverse of the linear stability operator. For this, we will use the "super-operator" norm $\| \cdot \|_{op}$ which is the operator norm over the Hilbert-Schmidt norm in the Hilbert space of complex matrices.

The idea of the proof in [3] is to write the linear stability operator as a product of invertible operators with an operator of the form $\mathcal{U} - \mathcal{T}$, where $\mathcal{U}$ is a unitary operator and $\mathcal{T}$ is a self-adjoint operator, called the "saturated self-energy" operator. In [21] the multivariate cumulant expansion technique is used instead, to give the same result.

They both arrive at the following theorem:

Theorem 5.4.1 (Stability of the Dyson equation). *If the mean-field condition* (5.20) *is satisfied, then for z in the "bulk" of the support of ρ, we have that:*

$$\left\| (I - C_{M(z)} \circ S)^{-1} \right\|_{op} \lesssim \frac{1}{\rho(z)^C},$$

for some constant $C > 0$, where ρ is called the "density of states" and is defined as:

$$\rho(z) := \frac{1}{\pi} \left\langle \operatorname{Im} M(z) \right\rangle. \tag{5.36}$$

We will see in section 5.5 that this is indeed a density for a measure and it is actually the harmonic extension of the real density of the limiting spectral measure of H.

The following Lemma is derived by the analysis in [3]:

Lemma 5.4.2 (Linear stability operator estimate). *Under the mean-field condition* (5.20), *we have the following estimate, for z in the "bulk" of ρ:*

$$\left\| (I - C_{M(z)} \circ S)^{-1} \right\|_{op} \lesssim \frac{1}{\rho(z)^C} \left\| (\mathcal{U} - \mathcal{T})^{-1} \right\|_{op},$$

for some constant $C > 0$, where $\mathcal{U}$ is a unitary operator and $\mathcal{T}$ a self-adjoint one.

Under some further conditions we can get the same result by the following supplementary Lemma in [3], which is of independent interest and inspired our stability analysis in section 6.7 for a different random matrix model:

Lemma 5.4.3 (Rotation-Inversion Lemma). *Let $\mathcal{T}$ be a self-adjoint and $\mathcal{U}$ a unitary operator on $\mathbb{C}^{N \times N}$. Suppose that $\mathcal{T}$ has a spectral gap, i.e., there is a constant $\theta > 0$ such that:*

$$Spec\left(\mathcal{T}\right) \subset \left[-\|\mathcal{T}\|_{op} + \theta, \|\mathcal{T}\|_{op} - \theta\right] \cup \left\{\|\mathcal{T}\|_{op}\right\},$$

with a non-degenerate largest eigenvalue $\|\mathcal{T}\|_{op} \leq 1$. Then there exists a constant $C > 0$ such

that:

$$\left\| (\mathcal{U} - \mathcal{T})^{-1} \right\|_{op} \le \frac{C}{\theta} \left| 1 - \|\mathcal{T}\|_{op} \langle K, \mathcal{U}[K] \rangle \right|^{-1},$$

where K is the normalized eigenmatrix of $\mathcal{T}$, corresponding to the largest eigenvalue $\|\mathcal{T}\|_{op}$.

The combination of Lemma 5.4.2 and Lemma 5.4.3 gives again Theorem 5.4.1.

5.5 Results about the Dyson equation

In this section, we give three results concerning the Dyson equation analysis of a Hermitian matrix. Notice that the first and second result is independent of the third one for which we need the small perturbation theorem as well as the stability theorem.

The first two results summarize the properties of the solution matrix M which can be regarded as a Stieltjes transform of a matrix-valued measure, which gives rise to a certain real density ρ.

The third result asserts that ρ should be the limiting spectral density of the initial Hermitian matrix H since its resolvent G is indeed "close" to the solution matrix M, as the dimensions go to infinity.

1. Solution of the Dyson equation.

We remind that the existence and uniqueness of the solution M for the Dyson equation, with a positive semi-definite imaginary part, is established in [37], through a certain fixed-point theorem and can be deduced by a well-defined iterative scheme.

Since $M(z) : \mathbb{C}_+ \mapsto \mathbb{C}_+^{N \times N}$ is actually a function from $\mathbb{C}_+ := \{z \in \mathbb{C} \mid \operatorname{Im}(z) > 0\}$ to the Hilbert space of complex matrices with $\operatorname{Im}(M(z)) \succeq 0$ (because $\operatorname{Im}(z) = \eta > 0$), we can view it as a matrix-valued Herglotz function and apply its Nevanlinna representation for a suitable matrix-valued measure on the real line. All the underlying theory can be found in section 5 of [30].

We get the following theorem for the solution matrix $M(z)$:

Theorem 5.5.1 (Stieltjes transform representation of M)**.** *Let M be the unique solution of* (5.34) *with positive semi-definite imaginary part. Then, M admits a Stieltjes transform representation:*

$$M_{xy} = \int_{\mathbb{R}} \frac{V_{xy}(dt)}{t - z},$$

for $x, y = 1, ..., N$ and $z \in \mathbb{C}_+$. The measure $V(dt) = (V_{xy}(dt))_{x,y=1}^N$ on the real line with values in positive semi-definite matrices is unique and satisfies the normalization $V(\mathbb{R}) = I$.

2. Density of states for the solution matrix.

Following the analysis from the Nevanlinna functions in [30], we have that the diagonal elements of V are actually finite Borel measures in $\mathbb{R}$. The normalization $V(\mathbb{R}) = I$ makes them probability measures. This makes the normalized trace of V a probability measure, as a convex combination of probability measures. Hence, from the previous theorem 5.5.1, we derive the following:

Theorem 5.5.2 (Density of states)**.** *Let $\langle M(z) \rangle : \mathbb{C}_+ \mapsto \mathbb{C}_+$ denote the normalized trace of the solution matrix. Then $\langle M(z) \rangle$ admits a Stieltjes tranform representation of a probability density ρ, such that:*

$$\langle V(dt) \rangle = \rho(t)dt,$$

i.e.

$$\langle M(z) \rangle = \int_{\mathbb{R}} \frac{\rho(t)}{t - z} dt.$$

We remark that under the flatness condition (5.20), the density ρ becomes Hölder continuous. [Proposition 2.2] in [3].

The density ρ is called the *density of states* as we have seen in section 5.4 because of the following fact. By the Stieltjes inversion formula, we have that:

$$\lim_{\eta \to 0^+} \frac{1}{\pi} \langle \mathrm{Im} M(z) \rangle = \rho(t),$$

hence

$$\rho(z) = \frac{1}{\pi} \langle \mathrm{Im} M(z) \rangle$$

is the harmonic extension of ρ.

We will see now that ρ is actually the limiting density for the eigenvalue counting function of the Hermitian matrix H.

3. Approximation of the resolvent matrix - Local law.

Here we use our two ingredients, the small error term D in section 5.3 and the stability of the Dyson equation in section 5.4, to present the last important theorem of this chapter, which is the approximation of the resolvent matrix $G = (H - zI)^{-1}$ by the solution matrix $M(z)$ for large dimensions.

Theorem 5.5.3 (Local law for the solution matrix). *Assume that the flatness condition* (5.20) *and the finite-moments condition* (5.33) *hold, as well as a suitable decay of correlations condition like* (5.38) *for the matrix H. Let M be the solution of the Dyson equation* (5.34) *with the self-energy operator defined as in* (5.8). *Then for the spectral parameter $z = E + i\eta$ inside the "bulk" of the support of ρ, where ρ is the density of states, i.e. with $\rho(E) \geq \delta$ and with $\eta \geq N^{-1+\gamma}$, for some $\gamma, \delta > 0$, we have the "entrywise local law":*

$$\mathbb{P}\left(|G_{ij}(z) - M_{ij}(z)| \geq \frac{N^\varepsilon}{\sqrt{N\eta}} \right) \leq \frac{C}{N^D}, \tag{5.37}$$

for some constant C depending on $\gamma, \delta, \varepsilon, D$ and the constants on (5.20) *and* (5.33).

We note the correlation decay condition for an exponential decay of the matrix correlations, is:

$$\mathrm{Cov}\left(\phi\left(W_A \right), \psi\left(W_B \right) \right) \leq C(\phi, \psi) e^{-d(A,B)}, \tag{5.38}$$

where $W = \sqrt{N}H$ is the re-scaled random matrix, A, B are two subsets of the index set $[1, N] \times [1, N]$ and the distance d is the usual Euclidean distance between the sets $A \cup A^{\mathrm{t}}$ and $B \cup B^{\mathrm{t}}$ and $W_A = (w_{ij})_{(i,j) \in A}$.

For a more general slow correlation decay, the condition is:

$$|\kappa\left(f_1(W), f_2(W)\right)| \leq \frac{C}{1 + d\left(\operatorname{supp} f_1, \operatorname{supp} f_2\right)^s}\, \|f_1\|_2\, \|f_2\|_2\,, \qquad (5.39)$$

for some $s > 12$ and f_1, f_2 square integrable functions on random $N \times N$ complex Hermitian matrices.

A simple corollary of Theorem 5.5.3 is the following "averaged local law":

Corollary 5.5.4. *With the same conditions as in Theorem 5.5.3, we have that:*

$$\mathbb{P}\left(\left|\frac{1}{N}\operatorname{Tr}[G(z) - M(z)]\right| \geq \frac{N^\varepsilon}{N\eta}\right) \leq \frac{C}{N^D}, \qquad (5.40)$$

again for a constant C depending on $\gamma, \delta, \varepsilon, D$ and the constants on (5.20) and (5.33).

The main key technical tool and observation is still the fact that the Stieltjes transform of the eigenvalue counting function ρ_N is given by the normalized trance of the resolvent, because:

$$S_N(z) = \int_{\mathbb{R}} \frac{\rho_N(t)}{t - z}\,dt = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} = \langle G(z)\rangle,$$

so that:

$$S_N(z) = \langle G(z)\rangle \rightarrow \langle M(z)\rangle = S_\rho(z)$$

and hence according to all of our analysis in section 3.4: $\rho_N \rightarrow \rho$, where ρ is the density of states.

Chapter 6

A non-Hermitian generalization of the Marchenko-Pastur distribution

6.1 The random matrix model

In order to define our model, we firstly define two auxiliary square random matrices P and Q of a given dimension $N \times N$. We assume that all the entries of P and Q are i.i.d. complex random variables with mean value 0 and complex variance $\frac{1}{2N}$.

The key difference in this random matrix model compared to the previous results stated in section 6.2 is that we allow the entries of P and Q to have any possible common distribution $\mathcal{F}$ with the following properties:

- $\mathcal{F}$ has mean zero, complex variance $\frac{1}{2N}$ and second moment zero, that is if $x \sim \mathcal{F}$ then $\mathbb{E}[\mathrm{Re}(x)] = \mathbb{E}[\mathrm{Im}(x)] = 0$, $\mathrm{Var}[\mathrm{Re}(x)] = \mathrm{Var}[\mathrm{Im}(x)] = \frac{1}{4N}$ and $\mathbb{E}[\mathrm{Re}(x)\mathrm{Im}(x)] = 0$. The last assumption is made to ease the computations of the model while the first two are vital for the results.

- $\mathcal{F}$ has all of its moments finite, this means that if $x \sim \mathcal{F}$ then $\mathbb{E}[x^n] < \infty$ for any $n \in \mathbb{N}$. This is necessary for the Dyson equation method to take place.

- $\mathcal{F}$ has a bounded density. Specifically, we make the assumption that there exist

$q \geq 1$, $\kappa > 0$ and a probability density $\psi \in L^q(\mathbb{C})$ with $\|\psi\|_q \leq N^\kappa$ such that:

$$\mathbb{P}\left(\sqrt{N}P_{ij} \in B\right) = \mathbb{P}\left(\sqrt{N}Q_{ij} \in B\right) = \int_B \psi(z)d^2z, \qquad (6.1)$$

for each $i, j = 1, ..., N$ and $B \subset \mathbb{C}$ a Borel set. This condition makes the least singular value problem (see section 6.8) much easier.

Our model is defined as follows. We pick

$$X_1 := \sqrt{1 + \tau}P + \sqrt{1 - \tau}Q \text{ and } X_2 := \sqrt{1 + \tau}P - \sqrt{1 - \tau}Q, \qquad (6.2)$$

for a real parameter $\tau \in [0, 1]$. Notice that X_1 and X_2 are correlated through this parameter. For $\tau = 0$, we have that X_1 and X_2 are completely uncorrelated, while for $\tau \in (0, 1)$ and $\tau \to 1$ they become more and more correlated, reaching complete correlation for $\tau = 1$ where they become identical matrices. Through X_1 and X_2 we now form the *correlated sample covariance ensemble*:

$$\mathbf{X} := X_1 X_2^* \qquad (6.3)$$

What can we say about the eigenvalues of this ensemble? For $\tau = 0, 1$ we get two well-known cases. Specifically, for $\tau = 0$ we have a product of two independent circular matrices, while for $\tau = 1$ we have a product of a circular matrix and its conjugate. Notice that in this case $\mathbf{X}$ becomes hermitian, that's why τ is called the *non-hermiticity* parameter.

In the first case, we have a finite product of circular random matrices, which means that the empirical spectral distribution converges to a power of a circular law. This distribution has the unit disk as support and exhibits a concentration of eigenvalues around the origin according to the number of product elements. In our case, we have the second power of the circular law. See [49] for this result.

In the second case, we have the classical sample covariance ensemble, which means that the empirical spectral distribution converges to the Marchenko-Pastur distribution. This

means that the support in this case collapses to the real line in $[0, 4]$. This is expected since the matrix becomes Hermitian and the eigenvalues real. See [44] for this result.

For $\tau \in (0, 1)$ we are expecting a transition between these two distributions. Since there is a concentration of eigenvalues near the origin both in the second power of the circular law as well as in the Marchenko-Pastur distribution, we are expecting this to hold for the transitional distribution while the initial disk moves to the right and shrinks to the real line. Of course, we are also expecting local laws to hold for this random matrix ensemble.

6.2 Previous results on the spectral distribution

This model was previously studied in [5] where it was assumed that $\mathcal{F} \sim \mathcal{CN}(0, \frac{1}{2N})$, that is $\mathcal{F}$ follows a complex normal distribution with mean 0 and variance $\frac{1}{2N}$. It was therefore given the name *non-Hermitian Wishart ensemble*.

In that study, since there was a Gaussian distribution assumption for the entries of P and Q, different and more concrete methods were used through orthogonal polynomials. This is because in this case we have specific potentials in our joint eigenvalue density for X_1 and X_2, which can then be used to find the joint eigenvalue density of $\mathbf{X}$ and proceed with asymptotic estimates.

These methods are rather concrete and powerful, so they gave rise to the calculation of the transitional distribution $(0 < \tau < 1)$ even for rectangular matrices $P, Q \in \mathbb{C}^{N \times (N+\nu)}$, where $\nu = \mathcal{O}(N)$ is a *non-square* parameter.

Using the parameter $\alpha := \lim_{N \to \infty} \frac{\nu}{N}$ they arrived at the following result concerning the transitional distribution of the non-Hermitian Wishart ensemble:

Theorem 6.2.1 (Non-Hermitian Wishart ensemble law from [5])**.** *As $N \to \infty$ the empirical spectral measure of the non-Hermitian Wishart ensemble converges to the following deterministic*

measure on the complex plane:

$$d\widehat{\mu}(\zeta) := \frac{1}{1-\tau^2} \frac{1}{\pi\sqrt{4|\zeta|^2 + (1-\tau^2)^2 \alpha^2}} \cdot \mathbf{1}_{\widehat{S}_{\tau,\alpha}}(\zeta)d^2\zeta, \tag{6.4}$$

where the support $\widehat{S}_{\tau,\alpha}$ of the spectrum is given by:

$$\widehat{S}_{\tau,\alpha} := \left\{ \zeta = x + iy : \left(\frac{x - \tau(2+\alpha)}{(1+\tau^2)\sqrt{1+\alpha}}\right)^2 + \left(\frac{y}{(1-\tau^2)\sqrt{1+\alpha}}\right)^2 \leq 1 \right\}. \tag{6.5}$$

| (A) $\tau = 0$ | (B) $\tau = 1/\sqrt{2}$ | (C) $\tau = 0.85$ | (D) $\tau = 1$ |
| (E) $\tau = 0$ | (F) $\tau = 1/\sqrt{2}$ | (G) $\tau = 0.85$ | (H) $\tau = 1$ |
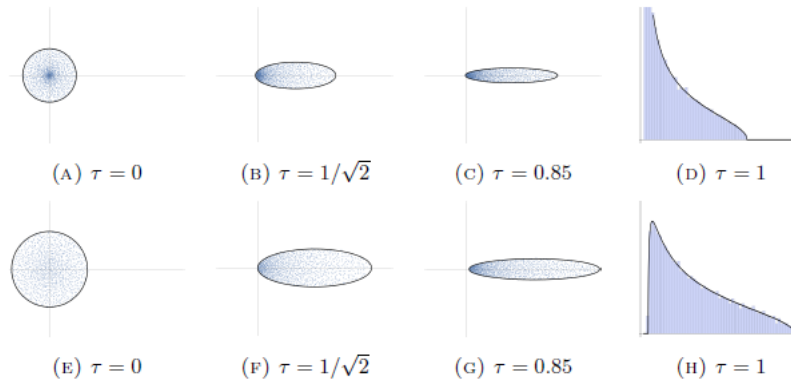
Figure 6.1: The eigenvalues of $\mathbf{X}$ for large N. We have that $\alpha = 0$ for the figures (A)-(D) and $\alpha = 1$ for (E)-(H). This figure is from [5].

6.3 Main theorems

Our main theorem is about the spectrum of the correlated covariance matrix $\mathbf{X_N} = X_1 X_2^*$ where X_1 and X_2 are square matrices and its corresponding limiting local law, as the dimension increases.

The empirical measure $\widehat{\mu}_N$ associated with $\mathbf{X}_N$ is given by:

$$\widehat{\mu}_N := \frac{1}{N}\sum_{j=1}^{N} \delta_{\widehat{\zeta}_j},$$

where $\{\widehat{\zeta}_j\}_{j=1}^N$ are the N complex eigenvalues of $\mathbf{X}_N$. We firstly introduce the limiting measure $\widehat{\mu}$ of the empirical spectral distribution of this matrix as follows:

Definition 6.3.1 (Limiting measure B)**.** *The measure $\widehat{\mu}$ is defined by*

$$d\widehat{\mu}(z) := \frac{1}{1-\tau^2}\frac{1}{2\pi|z|}\mathbf{1}_{\widehat{S}_\tau}(z)d^2z, \tag{6.6}$$

with the following support $\widehat{S}_\tau$:

$$\widehat{S}_\tau(z) := \left\{ z = x + iy : \left(\frac{x-2\tau}{1+\tau^2}\right)^2 + \left(\frac{y}{1-\tau^2}\right)^2 \leq 1 \right\}. \tag{6.7}$$

Notice that $\widehat{S}_\tau(z)$ defines a shifted ellipse centered at $(2\tau, 0)$ while the measure predicts a concentration of eigenvalues around the edge point of the ellipse at $(0,0)$.

The theorem about the correlated covariance matrix ensemble will follow from a similar theorem about the Dirac matrix, defined as follows:

$$\mathbf{Y}_N := \begin{pmatrix} 0 & X_1 \\ X_2^* & 0 \end{pmatrix} \in \mathbb{C}^{2N\times 2N}$$

The limiting measure μ of the empirical spectral distribution of the Dirac matrix is defined as follows:

Definition 6.3.2 (Limiting measure A)**.** *The measure μ is defined by*

$$d\mu(z) := \frac{1}{1-\tau^2}\frac{1}{\pi}\mathbf{1}_{S_\tau}(z)d^2z, \tag{6.8}$$

with the following support S_τ:

$$S_\tau := \left\{ z = x + iy : \left(\frac{x}{1+\tau}\right)^2 + \left(\frac{y}{1-\tau}\right)^2 \leq 1 \right\}. \tag{6.9}$$

Here, S_τ is a classical ellipse centered at $(0,0)$ with its eccentricity and edge points depending on τ.

For the proof of our theorems we will work primarily on the following domain for the Dirac matrix:

Definition 6.3.3 (Spectral parameter domain A). *For any $\delta, \kappa > 0$, define the ellipse that arises from the exclusion of the edges and a small ball around 0 as follows:*

$$S_{\tau,\delta,\kappa} := \left\{ \zeta = x + iy : \left(\frac{x}{1+\tau} \right)^2 + \left(\frac{y}{1-\tau} \right)^2 \leq 1 - \delta \right\} \backslash B_0(\kappa), \qquad (6.10)$$

where $B_0(\kappa)$ is a ball of radius κ around 0.

Analogously, we define the spectral domain for the correlated covariance matrix:

Definition 6.3.4 (Spectral parameter domain B). *For any $\delta, \kappa > 0$ define the shifted ellipse that arises from the exclusion of the edges and a small ball around 0 as follows:*

$$\widehat{S}_{\tau,\delta,\kappa} := \left\{ \zeta = x + iy : \left(\frac{x - 2\tau}{1+\tau^2} \right)^2 + \left(\frac{y}{1-\tau^2} \right)^2 \leq 1 - \delta \right\} \backslash B_0(\kappa). \qquad (6.11)$$

For the statement of the main theorem we need to define the *test functions* as complex functions $f : \mathbb{C} \to \mathbb{C}$ with compact support and at least two times complex-differentiable with continuous complex second derivative. Their corresponding *zoom functions* are defined as follows:

$$f_{\zeta_0,a}(\zeta) := N^{2a} f(N^a(\zeta - \zeta_0)), \qquad (6.12)$$

where N is the dimension of the matrix, $\zeta \in \mathbb{C}$ and $\zeta_0 \in \mathbb{C}$ is the point we want to "zoom in" at a zooming scale given by $a > 0$.

We are now ready to state our main theorem.

Theorem 6.3.1 (Local universal law for the spectrum of the correlated sample covariance ensemble). *Let $a \in (0, 1/2)$ and $\kappa, \delta \in (0, 1)$. Then, for any $\epsilon > 0$ and $\nu \in \mathbb{N}$, there is a constant $C > 0$, such that:*

$$\mathbb{P}\left(\left| \frac{1}{N} \sum_{\widehat{\zeta} \in Spec(X_1 X_2^*)} f_{\widehat{\zeta}_0,a}(z) - \int_{\mathbb{C}} f_{\widehat{\zeta}_0,a}(z) d\widehat{\mu}(z) \right| \leq N^{-1+2a+\epsilon} \|\Delta f\|_{L^1} \right) \geq 1 - CN^{-\nu},$$

uniformly for all $N \in \mathbb{N}$, $\widehat{\zeta}_0 \in \widehat{S}_{\tau,\delta,\kappa}$ and test functions f with corresponding zoom function $f_{\widehat{\zeta}_0,a}$, such that $\|\Delta f\|_{L^{2+a}} \leq N^D \|\Delta f\|_{L^1}$ for some $D \in \mathbb{N}$.

The Dirac matrix follows an elliptic law, as formulated in the following theorem.

Theorem 6.3.2 (Local universal law for the spectrum of the Dirac matrix). *Let $a \in (0, 1/2)$ and $\kappa, \delta \in (0, 1)$. Then, for any $\epsilon > 0$ and $\nu \in \mathbb{N}$, there is a constant $C > 0$, such that:*

$$\mathbb{P}\left(\left|\frac{1}{2N}\sum_{\zeta \in Spec(\mathbf{Y_N})} f_{\zeta_0, a}(\zeta) - \int_{\mathbb{C}} f_{\zeta_0, a}(z)d\mu(z)\right| \leq N^{-1+2a+\epsilon}\|\Delta f\|_{L^1}\right) \geq 1 - CN^{-\nu},$$

uniformly for all $N \in \mathbb{N}$, $\zeta_0 \in S_{\tau, \delta, \kappa}$ and test functions f with corresponding zoom function $f_{\zeta_0, a}$, such that $\|\Delta f\|_{L^{2+a}} \leq N^D\|\Delta f\|_{L^1}$ for some $D \in \mathbb{N}$.

Here we will recover the identical Dyson equation for the approximation of the trace of the resolvent of the Dirac matrix as in [7], and we can rely on the work done in [7] to deduce Theorem 6.3.2, as most proof details carry over directly to our case. We will focus on the differences between the case of the Dirac matrix and the elliptical law matrix from [7]. The two ingredients we must prove will be as follows:

1. We have to show that the trace of the resolvent for the Dirac matrix can be approximated by the v as in the 2×2 Dyson equation (6.22). The precise statement of this is formulated in Proposition 6.5.3. This furthermore requires a proof of stability for the 4×4 Dyson equation as formulated in Proposition 6.7.1 and proved in Section 6.7.

2. We need to control the relevant least singular values, formulated in Theorem 6.8.1 and proved in Section 6.8.

Proof of Theorem 6.3.2. The proof is identical to the proof of Theorem 2.1 in [7], using our Proposition 6.5.3 in place of their Proposition 3.1 and our Theorem 6.8.1 instead of their Theorem 3.4 to control the smallest singular value. □

6.4 Hermitization and Green function estimate

We start with Girko's formula

$$\frac{1}{2N} \sum_{\xi \in Spec(Y_N)} f(\xi) = \frac{1}{4\pi N} \int_{\mathbb{C}} \Delta f(\zeta) \log|\det \mathbf{H}_\zeta| d^2\zeta, \tag{6.13}$$

where we introduced the Hermitization:

$$\mathbf{H}_\zeta := \begin{pmatrix} 0 & \mathbf{Y}_N - \zeta \\ \mathbf{Y}_N^* - \bar{\zeta} & 0 \end{pmatrix} \in \mathbb{C}^{4N \times 4N} \tag{6.14}$$

The log-determinant of $\mathbf{H}_\zeta$ can be obtained from the resolvent matrix $\mathbf{G}(\zeta, \eta) := (\mathbf{H}_\zeta - i\eta)^{-1}$ through the identity:

$$\log|\det \mathbf{H}_\zeta| = -4N \int_0^T \langle \mathrm{Im}\mathbf{G}(\zeta, \eta)\rangle d\eta + \log|\det(\mathbf{H}_\zeta - iT)|, \tag{6.15}$$

valid for any $T > 0$, where we defined the normalized trace of a matrix $\mathbf{R} \in \mathbb{C}^{k \times k}$ as $\langle \mathbf{R}\rangle := \frac{1}{k}\mathrm{Tr}\mathbf{R}$.

Formula (6.13) is called the *logarithmic potential* method and we will use approximations of the normalized resolvent trace and of the quantity $\log|\det(H_\zeta - iT)|$ to estimate the left-hand side. The spectral distribution can be recovered from the Laplacian of the log-determinant of the Hermitization on the right-hand side of (6.13) or its approximation as the *spectral resolution* η goes to 0.

This method is based on the complex retrieval of measure identity in section 3.5 and we will not see the details here except from the approximation of the normalized trace of the resolvent as in proposition 6.5.3. The rest of the proof can be found in the proof of Theorem 2.1 in [7], including the proof of Lemma 3.5.

6.5 Solution of the Dyson equation

We consider now, as in [7], the Dyson equation, which will provide an approximation of the normalized resolvent trace. Since we are working with a 4×4 matrix here the equation will be modified:

$$-M^{-1} = \begin{pmatrix} iH & Z \\ \overline{Z} & iH \end{pmatrix} + \mathscr{P}(M), \ \ \mathscr{P}\left[(M_{ij})_{i,j \leq 4}\right] = \begin{pmatrix} M_{44} & 0 & \tau M_{42} & 0 \\ 0 & M_{33} & 0 & \tau M_{31} \\ \tau M_{24} & 0 & M_{22} & 0 \\ 0 & \tau M_{13} & 0 & M_{11} \end{pmatrix},$$

(6.16)

where $\zeta \in \mathbb{C}$, $\eta > 0$ and $M = M(\zeta, \eta) \in \mathbb{C}^{4 \times 4}$ is the unknown matrix. Since there are many possible solutions, we require

$$\mathrm{Im} M = \frac{1}{2i}(M - M^*)$$

(6.17)

to be positive definite. We have defined

$$Z := \begin{pmatrix} \zeta & 0 \\ 0 & \zeta \end{pmatrix} , \ \ iH := \begin{pmatrix} i\eta & 0 \\ 0 & i\eta \end{pmatrix}.$$

The operator $\mathscr{P}$ arises from the calculation in (6.38).

Proposition 6.5.1 (Solution of the Dyson equation). *There exists a unique solution M of* (6.16) *with the constraint* (6.17) *and it satisfies:*

$$M = \begin{pmatrix} iV & \overline{B} \\ B & iV \end{pmatrix},$$

(6.18)

where we defined

$$iV := \begin{pmatrix} iv(\zeta, \eta) & 0 \\ 0 & iv(\zeta, \eta) \end{pmatrix} , \ \ B := \begin{pmatrix} b(\zeta, \eta) & 0 \\ 0 & b(\zeta, \eta) \end{pmatrix}$$

(6.19)

for each $\eta > 0$, $\zeta \in \mathbb{C}$ and for some $v = v(\zeta, \eta) \in (0, \infty)$ and $b = b(\zeta, \eta) \in \mathbb{C}$.

Proof. Let $W = \begin{pmatrix} iH & Z \\ \overline{Z} & iH \end{pmatrix}$.

The claim follows from Theorem 2.1 of [37], which demonstrates the existence and uniqueness of the solution M and gives it as a limit of iterates $M_n = \mathcal{F}^n(M_0)$ with $\mathcal{F}(M_j) = -(W + \mathscr{P}(M_j))^{-1}$. We see that if we take M_0 of the form (6.18), it remains of the form (6.18) upon application of $\mathcal{F}$, thus M is of the form (6.18). Indeed, using a Schur complement formula for inversion and noting that all the 2×2 blocks commute, we see that for M_j of the form (6.18), explicitly, we compute

$$\mathcal{F}(M_j)_{11} = i\frac{v + \eta}{(v + \eta)^2 + |\zeta + \tau b|^2} = \mathcal{F}(M_j)_{33} \tag{6.20}$$

$$\mathcal{F}(M_j)_{13} = \frac{-\zeta - \tau b}{(v + \eta)^2 + |\zeta + \tau b|^2} = \overline{\mathcal{F}(M_j)_{31}}. \tag{6.21}$$

This completes the proof of 6.5.1. $\qquad \square$

Due to (6.19) and (6.18), we can take partial normalized traces on (6.16) (see definition after equation (6.42)) to recover a 2×2 Dyson Equation, which we will work with henceforth and which matches the equation in [7]:

$$-M^{-1} = \begin{pmatrix} i\eta & \zeta \\ \overline{\zeta} & i\eta \end{pmatrix} + \mathscr{P}_2(M), \quad \mathscr{P}_2\left[(M_{ij})_{1 \leq i,j \leq 2}\right] = \begin{pmatrix} M_{22} & \tau M_{21} \\ \tau M_{12} & M_{11} \end{pmatrix}, \tag{6.22}$$

where $\zeta \in \mathbb{C}$, $\eta > 0$ and $M = M(\zeta, \eta) \in \mathbb{C}^{2 \times 2}$ is the unknown matrix with

$$\mathrm{Im}M = \frac{1}{2i}(M - M^*) \succeq 0. \tag{6.23}$$

As the equation is identical to the equation in [7], the solution will satisfy all the same properties. Thus we can import without changes the following lemma and identities from [7]:

Lemma 6.5.2 (Lemma 4.1 in [7], Basic estimates of M). *We have the following estimates:*

(i) *Uniformly for all $\eta > 0$ and $\zeta \in \mathbb{C}$, we have that:*

$$\|M\| \leq \min\{1, \eta^{-1}\} \leq 2(1 + \eta^{-1}), \tag{6.24}$$

$$\|M^{-1}\| \lesssim 1 + \eta + |\zeta|. \tag{6.25}$$

(ii) *Let $r > 0$. Then uniformly for all $\eta > 0$ and $\zeta \in \mathbb{D}_r$, we have:*

$$M = i\eta^{-1} + \mathcal{O}_r(\eta^{-2}), \tag{6.26}$$

$$M^* M \gtrsim_r (1 + \eta)^{-2}, \tag{6.27}$$

$$\mathrm{Im} M \gtrsim_r \eta(1 + \eta)^{-2} \tag{6.28}$$

and in particular,

$$\|M\|^2 = v^2 + |b|^2 = \frac{v}{v + \eta}. \tag{6.29}$$

Here the $\lesssim$ symbol means that a quantity is less than another up to a constant. The symbol $\lesssim_a$ means that a quantity is less than another up to a constant that may depend on a.

Furthermore, we import equation (4.7) from [7], which gives the following expression for b:

$$b = -\frac{\mathrm{Re}\zeta}{1 + \tau + \eta/v} + i\frac{\mathrm{Im}\zeta}{1 - \tau + \eta/v}. \tag{6.30}$$

The following proposition will give an estimate for the normalized trace of the Green function of the Hermitization of the Dirac matrix, by the solution matrix of the Dyson equation, the proof of which will be the theme of section 6.6:

Proposition 6.5.3 (Local law for $\mathbf{H}_\zeta$, averaged version). *Let v be defined as in (6.18) and let $\gamma, \delta, \kappa > 0$. Then for any $\epsilon > 0$ and $\nu > 0$, there exists $C_{\epsilon,\nu} > 0$ such that:*

$$\mathbb{P}\left(|\langle \mathbf{G}(\zeta, \eta) \rangle - iv(\zeta, \eta)| \leq \frac{N^\epsilon}{N\eta} \right) \geq 1 - C_{\epsilon,\nu} N^{-\nu},$$

uniformly for all bounded $\eta \geq N^{-1+\gamma}$, $\zeta \in S_{\tau,\delta,\kappa}$ and $N \in \mathbb{N}$.

6.6 Local law for the Hermitization matrix

In this section we prove Proposition 6.5.3.

To express the approximation $\mathbf{G} \overset{(N\to\infty)}{\approx} \mathbf{M}$ we introduce as in [7] some appropriate norms. For any random matrix $R \in \mathbb{C}^{l \times l}$ in dimension $l \in \mathbb{N}$ we define the p-norms

$$\|R\|_p := \|R\|_p^{iso} := \sup_{\|x\|,\|y\|\leq 1} (\mathbb{E}|\langle x, Ry\rangle|^p)^{1/p},$$

$$\|R\|_p^{avg} := \sup_{\|W\|\leq 1} (\mathbb{E}|\langle WR\rangle|^p)^{1/p}, \tag{6.31}$$

where the supremum is taken over $x, y \in \mathbb{C}^l$ and $W \in \mathbb{C}^{l \times l}$, respectively. The symbol $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product on $\mathbb{C}^l$ and the symbol $\langle \cdot \rangle$ the normalized trace on $\mathbb{C}^{l \times l}$.

Notice that for any finite $l \in \mathbb{N}$ the norms are equivalent to each other and also equivalent to the standard norm of the random variable $\|R\|$ where $\|\cdot\|$ denotes the operator norm. ($\|R\|$ is a positive real random variable).

The main Theorem 6.6.1 states the approximation of the $4N \times 4N$ resolvent matrix $\mathbf{G} = \mathbf{G}(\zeta, \eta) = (\mathbf{H}_\zeta - i\eta)^{-1} \in \mathbb{C}^{4N \times 4N}$ to the deterministic matrix $\mathbf{M} = \mathbf{M}(\zeta, \eta) \in \mathbb{C}^{4N \times 4N}$, as $N \to \infty$. We recall that $\mathbf{M}$ is defined as $\mathbf{M} := M \otimes \mathbf{1}_{N \times N}$ which means that

$$\mathbf{M}(\zeta, \eta) = \begin{pmatrix} iv(\zeta, \eta) & \overline{b(\zeta, \eta)} \\ b(\zeta, \eta) & iv(\zeta, \eta) \end{pmatrix} \tag{6.32}$$

where every entry in this 2×2 block structure is a multiple of the identity matrix $\mathbf{1}_{2N \times 2N}$, where $v(\zeta, \eta)$ and $b(\zeta, \eta)$ are defined as in 6.18.

Theorem 6.6.1 (Local law for the Hermitization matrix $\mathbf{H}_\zeta$, main version)**.** *Let $\gamma \in (0,1)$ and $p \in \mathbb{N}$. Uniformly, for all bounded $\eta \geq N^{-1+\gamma}$ and $\zeta \in E_{\rho,\gamma}$, the following local law holds:*

$$\|\mathbf{G} - \mathbf{M}\|_p^{iso} \lesssim_{p,\gamma} \frac{N^\gamma}{\sqrt{N\eta}} \quad , \quad \|\mathbf{G} - \mathbf{M}\|_p^{avg} \lesssim_{p,\gamma} \frac{N^\gamma}{N\eta}. \tag{6.33}$$

Proof of Proposition 6.5.3. By the definition of $\mathbf{M}$ in (6.32) we have that $\langle \mathbf{M} \rangle = iv$. Thus, for suitably chosen γ and p, we get as in [7] that

$$\mathbb{P}\left(|\langle \mathbf{G}(\zeta,\eta) \rangle - iv(\zeta,\eta)| \geq \frac{N^\epsilon}{N\eta} \right) \leq \frac{(N\eta)^p}{N^{\epsilon p}} \mathbb{E} \, |\langle \mathbf{G}(\zeta,\eta) \rangle - iv(\zeta,\eta)|^p \leq$$
$$\leq \frac{(N\eta)^p}{N^{\epsilon p}} \left(\|\mathbf{G} - \mathbf{M}\|_p^{avg} \right)^p,$$

by Markov's inequality and the result follows.

Here we prove Theorem 6.6.1 in the local regime $\eta \ll 1$ so as to get our universal local spectral distribution for the Dirac matrix.

Proof of Theorem 6.6.1. As in [7] the proof follows from the fact that $\mathbf{G}$ approximately satisfies the matrix Dyson equation:

$$1 + (i\eta + \mathbf{Z} + \boldsymbol{S}\mathbf{M})\mathbf{M} = 0, \tag{6.34}$$

where each constant is a multiple of the identity matrix $\mathbf{1}_{4N \times 4N}$, $\mathbf{Z} := -\mathbb{E}\mathbf{H}_\zeta$, and the unknown matrix $\mathbf{M}$ will approximate the resolvent matrix $\mathbf{G}$ in $\mathbb{C}^{4N \times 4N}$. Here $\boldsymbol{S}$ is the natural extension of $\mathscr{P}$ in (6.16) to $\mathbb{C}^{4N \times 4N}$, i.e. for

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} & \mathbf{A}_{14} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} & \mathbf{A}_{24} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} & \mathbf{A}_{34} \\ \mathbf{A}_{41} & \mathbf{A}_{42} & \mathbf{A}_{43} & \mathbf{A}_{44} \end{pmatrix} \in \mathbb{C}^{4N \times 4N}, \tag{6.35}$$

$$
\mathbf{Z} = \begin{pmatrix} 0 & \zeta \mathbf{1}_{2N \times 2N} \\ \bar{\zeta} \mathbf{1}_{2N \times 2N} & 0 \end{pmatrix}, \quad \boldsymbol{S}\mathbf{A} = \begin{pmatrix} \langle \mathbf{A}_{44} \rangle & 0 & \tau \langle \mathbf{A}_{42} \rangle & 0 \\ 0 & \langle \mathbf{A}_{33} \rangle & 0 & \tau \langle \mathbf{A}_{31} \rangle \\ \tau \langle \mathbf{A}_{24} \rangle & 0 & \langle \mathbf{A}_{22} \rangle & 0 \\ 0 & \tau \langle \mathbf{A}_{13} \rangle & 0 & \langle \mathbf{A}_{11} \rangle \end{pmatrix}, \quad (6.36)
$$

where each entry is a multiple of the $N \times N$ identity matrix. The matrix $\mathbf{M}$ defined in (6.32) solves (6.34).

As in [7] and [21], we introduce the *self-energy* operator:

$$
\hat{\boldsymbol{S}} : \mathbb{C}^{4N \times 4N} \mapsto \mathbb{C}^{4N \times 4N} \text{ by}
$$

$$
\hat{\boldsymbol{S}}\mathbf{A} = \mathbb{E}\left\{ (\mathbf{H}_\zeta + \mathbf{Z})\mathbf{A}(\mathbf{H}_\zeta + \mathbf{Z}) \right\}. \tag{6.37}
$$

We show that $\hat{\boldsymbol{S}} = \boldsymbol{S}$ which justifies the form that we have for the Dyson equation (6.34). Careful calculation of the self-energy operator for a general matrix $\mathbf{A}$ shows that:

$$
\hat{\boldsymbol{S}}\mathbf{A} = \mathbb{E}\{(\mathbf{H}_\zeta + \mathbf{Z})\mathbf{A}(\mathbf{H}_\zeta + \mathbf{Z})\} = \mathbb{E}\left\{ \begin{pmatrix} 0 & \mathbf{Y_N} \\ \mathbf{Y_N}^* & 0 \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} 0 & \mathbf{Y_N} \\ \mathbf{Y_N}^* & 0 \end{pmatrix} \right\}
$$

$$
= \mathbb{E} \begin{pmatrix} \mathbf{Y_N} A_{22} \mathbf{Y_N}^* & \mathbf{Y_N} A_{21} \mathbf{Y_N} \\ \mathbf{Y_N}^* A_{12} \mathbf{Y_N}^* & \mathbf{Y_N}^* A_{11} \mathbf{Y_N} \end{pmatrix},
$$

where $\mathbf{A}_{ij}$ are $2N \times 2N$ matrices for $i, j = 1, 2$. We now calculate each term in the resulting block matrix. We only show what happens in the first and second block entry and the rest calculations will be similar:

$$
\mathbb{E}[\mathbf{Y_N} A_{22} \mathbf{Y_N}^*] = \mathbb{E}\left\{ \begin{pmatrix} 0 & X_1 \\ X_2^* & 0 \end{pmatrix} \begin{pmatrix} [\mathbf{A}_{22}]_{11} & [\mathbf{A}_{22}]_{12} \\ [\mathbf{A}_{22}]_{21} & [\mathbf{A}_{22}]_{22} \end{pmatrix} \begin{pmatrix} 0 & X_2 \\ X_1^* & 0 \end{pmatrix} \right\}
$$

$$
= \begin{pmatrix} \langle [\mathbf{A}_{22}]_{22} \rangle & [\mathbf{A}_{22}]_{21}^T \left[(1+\tau)\mathbb{E}p_{11}^2 - (1-\tau)\mathbb{E}q_{11}^2 \right] \\ [\mathbf{A}_{22}]_{12}^T \left[(1+\tau)\mathbb{E}p_{11}^2 - (1-\tau)\mathbb{E}q_{11}^2 \right] & \langle [\mathbf{A}_{22}]_{11} \rangle \end{pmatrix}
$$

$$
= \begin{pmatrix} \langle [\mathbf{A}_{22}]_{22} \rangle & 0 \\ 0 & \langle [\mathbf{A}_{22}]_{11} \rangle \end{pmatrix},
$$

where $[\mathbf{A}_{22}]_{kl}$ are $N \times N$ matrices for $k, l = 1, 2$. We used the first assumption 6.1 for the distribution of the matrix entries in the last equality. We now compute the second block:

$$
\mathbb{E}[\mathbf{Y_N A}_{21} \mathbf{Y_N}] = \mathbb{E}\left\{ \begin{pmatrix} 0 & X_1 \\ X_2^* & 0 \end{pmatrix} \begin{pmatrix} [\mathbf{A}_{21}]_{11} & [\mathbf{A}_{21}]_{12} \\ [\mathbf{A}_{21}]_{21} & [\mathbf{A}_{21}]_{22} \end{pmatrix} \begin{pmatrix} 0 & X_1 \\ X_2^* & 0 \end{pmatrix} \right\}
$$

$$
= \begin{pmatrix} \tau\langle[\mathbf{A}_{21}]_{22}\rangle & [\mathbf{A}_{21}]_{21}^T(1+\tau)\mathbb{E}p_{11}^2 + [\mathbf{A}_{21}]_{21}^T(1-\tau)\mathbb{E}q_{11}^2 \\ [\mathbf{A}_{12}]_{12}^T(1+\tau)\mathbb{E}p_{11}^2 + [\mathbf{A}_{12}]_{12}^T(1-\tau)\mathbb{E}q_{11}^2 & \tau\langle[\mathbf{A}_{21}]_{11}\rangle \end{pmatrix}
$$

$$
= \begin{pmatrix} \tau\langle[\mathbf{A}_{21}]_{22}\rangle & 0 \\ 0 & \tau\langle[\mathbf{A}_{21}]_{11}\rangle \end{pmatrix},
$$

where $[\mathbf{A}_{21}]_{kl}$ are $N \times N$ matrices for $k, l = 1, 2$ and A^T denotes the transpose matrix of the matrix A.

The result will be

$$
\hat{\mathbf{S}}\mathbf{A} = \begin{pmatrix} \langle\mathbf{A}_{44}\rangle & 0 & \tau\langle\mathbf{A}_{42}\rangle & 0 \\ 0 & \langle\mathbf{A}_{33}\rangle & 0 & \tau\langle\mathbf{A}_{31}\rangle \\ \langle\mathbf{A}_{24}\rangle & 0 & \langle\mathbf{A}_{22}\rangle & 0 \\ 0 & \tau\langle\mathbf{A}_{13}\rangle & 0 & \langle\mathbf{A}_{11}\rangle \end{pmatrix}, \tag{6.38}
$$

with each entry being a multiple of the $N \times N$ identity matrix, which coincides with $\mathbf{S}\mathbf{A}$.

We continue with the proof and mark that (6.34) is an approximation of the resolvent identity:

$$
1 + (i\eta + \mathbf{Z} + \mathbf{S}\mathbf{G})\mathbf{G} = \mathbf{D}, \quad \mathbf{D} := (\mathbf{H}_\zeta + \mathbf{Z} + \mathbf{S}\mathbf{G})\mathbf{G}, \tag{6.39}
$$

which can be regarded as a perturbation of (6.34) with error matrix $\mathbf{D}$.

The following theorem from [21] gives bounds for the error matrix $\mathbf{D}$ when $\mathbf{S}$ is given by the self-energy operator $\hat{\mathbf{S}}$ of our random matrix model:

Theorem 6.6.2 (Bound on the error matrix, from [21], Theorem 4.1). *Under our assumptions 6.1 for our random matrix model, we have the following for the error matrix of the Dyson equation. Let $\epsilon > 0$ and $p \in \mathbb{N}$. There exists a constant $C_* > 0$ such that, uniformly for $\eta \in [\frac{1}{N}, 1]$, we have the following bounds on the error matrix:*

$$\|\mathbf{D}\|_p^{iso} \lesssim_{p,\epsilon} N^\epsilon \sqrt{\frac{\|\mathrm{Im}\mathbf{G}\|_q}{N\eta}}(1 + \|G\|_q)^{C_*}\left(1 + \frac{\|\mathbf{G}\|_q}{N^{1/4}}\right)^{C_* p} \tag{6.40}$$

$$\|\mathbf{D}\|_p^{avg} \lesssim_{p,\epsilon} N^\epsilon \frac{\|\mathrm{Im}\mathbf{G}\|_q}{N\eta}(1 + \|G\|_q)^{C_*}\left(1 + \frac{\|\mathbf{G}\|_q}{N^{1/4}}\right)^{C_* p}, \tag{6.41}$$

with $q = C_* p^4/\epsilon$, where $\mathbf{D}$ is defined by equation (6.39).

With the help of this proposition we are able to prove the local law (6.33) for the Hermitization matrix $\mathbf{H}_\zeta$. By subtracting (6.34) from (6.39) and defining our target matrix $\mathbf{\Delta} := \mathbf{G} - \mathbf{M}$ we arrive at the equation:

$$\mathbf{\Delta} - \mathbf{M}(S\mathbf{\Delta})\mathbf{M} = \mathbf{M}(S\mathbf{\Delta})\mathbf{\Delta} - \mathbf{MD}. \tag{6.42}$$

We now use the *partial trace operator*: $\mathbb{C}^{2N \times 2N} \mapsto \mathbb{C}^{2 \times 2}, \ \mathbf{A} \to \underline{\mathbf{A}}$, where:

$$\underline{\mathbf{A}} := \begin{pmatrix} \langle \mathbf{A_{11}} \rangle & \langle \mathbf{A_{12}} \rangle \\ \langle \mathbf{A_{21}} \rangle & \langle \mathbf{A_{22}} \rangle \end{pmatrix},$$

whenever

$$\mathbf{A} = \begin{pmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{pmatrix},$$

with each entry A_{ij} being an $N \times N$ block matrix. This allows us to reduce our equation from $\mathbb{C}^{4N \times 4N}$ to $\mathbb{C}^{4 \times 4}$ and deduce a first bound for the matrix $\Delta := \underline{\mathbf{\Delta}}$. We arrive at the equation:

$$\mathscr{L}\Delta = M(\mathscr{P}\Delta)\Delta - MD, \tag{6.43}$$

where $D := \underline{\mathbf{D}}$, and $\mathscr{L} : \mathbb{C}^{4 \times 4} \mapsto \mathbb{C}^{4 \times 4}$ is the linear stability operator from (4.10) of [7] - see section 6.7. From Corollary 6.7.2 we have that $\mathscr{L}$ leaves a certain subspace of $\mathbb{C}^{4 \times 4}$

which we call A_+ invariant - see section 6.7, and

$$K_{\delta,\kappa} := \sup_{\zeta \in S_{\tau,\delta,\kappa}} \|\mathscr{L}^{-1}|_{A_+}\| \lesssim_{\delta,\kappa} 1. \tag{6.44}$$

Note, importantly, that $\Delta \in A_+$ because $\langle \mathbf{G}_{11} \rangle = \langle \mathbf{G}_{22} \rangle$ (which can be seen using the Schur formula for inverse) and $\langle \mathbf{M}_{11} \rangle = \langle \mathbf{M}_{22} \rangle = iv$ (see (6.32)).

Thus, we can invert $\mathscr{L}$ on A_+ in (6.43) and take the p-norm defined on (6.31) and derive the inequality:

$$\|\Delta\|_p (1 - K_{\delta,\kappa} \|\Delta\|_p) \lesssim K_{\delta,\kappa} \|D\|_p,$$

where we used the comparability of the norms for $l = 2$. This means that if we choose c as a small enough constant, then:

$$\|\Delta \mathbf{1}(\|\Delta\|_p \leq c/K_{\delta,\kappa})\|_p \lesssim K_{\delta,\kappa} \|D\|_p. \tag{6.45}$$

We also note two important inequalities that connect D with $\mathbf{D}$ and Δ with $\boldsymbol{\Delta}$. We have that:

$$\|D\|_p \lesssim \|\mathbf{D}\|_p^{avg} \tag{6.46}$$

and

$$\|\boldsymbol{\Delta}\|_p^{\#} \lesssim \|\Delta\|_p + \|\Delta\|_{2p} \|\boldsymbol{\Delta}\|_{2p}^{\#} + \|\mathbf{D}\|_p^{\#}, \tag{6.47}$$

which comes from (6.42) after taking norms and using Cauchy-Schwartz as well as the definition of S. This relationship between norms of $\boldsymbol{\Delta}$ and Δ is analogous to (5.22) in [7] and is the final ingredient that is needed to complete the proof of Lemma 6.6.3, which yields the next Proposition 6.5.3. The proof of Lemma 6.6.3 is identical to the proof of Lemma 5.4 in [7]. □

Lemma 6.6.3 (Bootstrapping). *There is a constant $c_* > 0$ depending only on the distribution of the entries of the random matrices P and Q such that $\|\boldsymbol{\Delta}\|_p \lesssim_{p,\delta,\gamma} N^{-\gamma/6}$ for all $p \in \mathbb{N}$ in $\mathbb{A}_{\delta,\gamma}$ implies $\|\boldsymbol{\Delta}\|_p \lesssim_{p,\delta,\gamma} N^{-\gamma/6}$ for all $p \in \mathbb{N}$ in $\mathbb{A}_{\delta,(1-c_*)\gamma}$.*

Here we recall the main steps of the bootstrapping argument from [7]. Inequality (6.45), together with the inequalities (6.46) , (6.47) and (6.40) allows us to show proposition 6.6.1, i.e. $\|\boldsymbol{\Delta}\|_p \ll 1$, by bootstapping from $\eta \sim 1$ to $\eta \sim N^{-1+\gamma}$, for $1 \gg \gamma > 0$. We remark a sketch of the proof and the rest of the details can be found in [7]:

- Implication inside induction. Through proposition (6.40) a bound of the form $\|\boldsymbol{\Delta}\|_p \lesssim 1$ implies that $\|D\|_p \ll 1$ because then $\|\mathbf{G}\|_p \leq \|\mathbf{M}\|_p + \|\boldsymbol{\Delta}\|_p \lesssim 1$, for all $p \in \mathbb{N}$ and $N\eta \gg 1$. This in turn, through (6.45), implies that $\|\Delta\|_p \ll 1$. Finally we estimate that $\|\boldsymbol{\Delta}\|_p \ll 1$ because of (6.47). Altogether, this argument shows that $\|\boldsymbol{\Delta}\|_p \lesssim 1$ implies $\|\boldsymbol{\Delta}\|_p \ll 1$ on all of $\mathbb{A}_{\delta,\gamma}$, for any $\gamma > 0$, where we introduced the parameter set:

$$(\zeta, \eta) \in \mathbb{A}_{\delta,\gamma} := S_{\tau,\delta,\kappa} \times [N^{-1+\gamma}, 1].$$

  This implication is bootstrapped from $\eta \sim 1$ all the way to $\eta \sim N^{-1+\gamma}$, and is formulated in the following bootstrapping lemma 6.6.3 which we import from [7]. Its proof in our model is similar.

- Induction basis. The regime $\eta \sim 1$ is established by the global law in [21], [Theorem 2.1] in combination with [Lemma 5.4.1] and the bound $\|G\| \leq \eta^{-1} \leq N$ on $\mathbb{A}_{\delta,\gamma}$. Notice that the M matrix there can be replaced by our solution matrix $\mathbf{M}$. Specifically, we get the following proposition:

  Proposition 6.6.4 (Global law). *There exists a universal constant $c > 0$ such that for any $\epsilon > 0$:*

$$\|\boldsymbol{\Delta}\|_p \lesssim \frac{N^\epsilon}{\sqrt{N\eta}}, \quad \|\boldsymbol{\Delta}\|_p^{avg} \lesssim \frac{N^\epsilon}{N\eta} \tag{6.48}$$

  for any bounded ζ and $\eta \in [N^{-c\epsilon}, 1]$.

6.7 Stability of the Dyson equation

To relate Δ and $\boldsymbol{\Delta}$ in the boostrapping argument above, we will need to show the invertibility of the *stability operator* defined by:

$$\mathscr{L} : \mathbb{C}^{4\times4} \mapsto \mathbb{C}^{4\times4}$$

$$\mathscr{L}(R) := R - M\mathscr{P}(R)M \tag{6.49}$$

Specifically, we will need bounds on the norm of the inverse of this operator, in a subspace of $\mathbb{C}^{4\times4}$ where it can be inverted.

We introduce the 4×4 matrix:

$$E_- := \begin{pmatrix} I_{2\times2} & 0 \\ 0 & -I_{2\times2} \end{pmatrix}$$

We will work on the subspace $A_+ := E_-^{\perp}$ orthogonal to this matrix where orthogonality is understood with respect to the Hilbert-Schmidt inner product on $\mathbb{C}^{4\times4}$. We note that:

$$A_+ = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathbb{C}^{4\times4} \,\middle|\, \mathrm{Tr}A = \mathrm{Tr}D \right\}$$

We write $\mathscr{L}^{-1}|_{A_+}$ to denote the restriction of the inverse of $\mathscr{L}$ in the domain A_+, i.e. the operator is first restricted on A_+ and then inverted. In the proof of this proposition we will use $\|.\|$ for the operator norm.

Proposition 6.7.1 (Linear stability estimate). *For any $\zeta \in \mathbb{C}$ and $\eta > 0$, the operator $\mathscr{L}$ leaves A_+ invariant and is invertible on that subspace. Let*

$$\alpha := 1 - |v^2 - |b|^2| \tag{6.50}$$

The norm of the inverse of the restriction operator on A_+ satisfies:

$$\|\mathscr{L}^{-1}|_{A_+}\| \lesssim \alpha^{-2}, \tag{6.51}$$

uniformly for all bounded ζ and $\eta \in (0, 1]$.

Proof. As in [7], the linear stability operator $\mathscr{L}$ can be written in this form $\mathcal{U} - \mathcal{T}$, where $\mathcal{U}$ is a unitary matrix and $\mathcal{T}$ a self-adjoint matrix. From (6.18), we deduce that the solution matrix M is normal and therefore there exists a polar decomposition, $M = |M|U = U|M|$, for some unitary matrix U. By direct computation, we find that:

$$|M|^2 = M^*M = \|M\|^2 \cdot I_4, \quad \|M\|^2 = v^2 + |b|^2 = \frac{v}{\eta + v}$$

and so we have that U should be equal to:

$$U = (v^2 + |b|^2)^{-1/2} M.$$

We write the linear stability operator as $\mathscr{L} = \mathcal{U}^*(\mathcal{U} - \|M\|^2 \mathscr{P})$, where we defined $\mathscr{P}$ as in (6.16) and $\mathcal{U}$ is defined as $\mathcal{U}R = U^*RU^*$, for any $R \in \mathbb{C}^{4 \times 4}$. Note that $\mathcal{U}$ and $\mathscr{P}$ leave A_+ invariant and $\mathscr{P}$ is self-adjoint whereas $\mathcal{U}$ is unitary. We observe that:

$$\mathrm{Spec}(\mathscr{P}|_{A_+}) = \{1, -1, 0, \tau, -\tau\},$$

with the following eigenspaces:

- $P_1 = \left\langle \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right\rangle$

- $P_{-1} = \left\langle \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \right\rangle$

Hence, for $\mathcal{T} := \|M\|^2 \mathscr{P}$, we have that $\|\mathcal{T}\| = \|M\|^2 \cdot \|\mathscr{P}\| = \|M\|^2 = \frac{v}{\eta+v} \leq 1$. To complete the proof, we will take inspiration from [2], (Lemma 5.8) and [7], (Lemma

4.5). In our case, the Hermitian operator has no spectral gap. Instead it has a double eigenvalue at 1 and a non-degenerate eigenvalue at -1. However, using a projection onto the span of the eigenvectors corresponding to ± 1 instead of a projection onto the top eigenvector in the case of a non-trivial spectral gap and performing explicit computations for our particular operator $\mathcal{U}$, we are able to recover necessary estimates on the inverse of the stability operator.

We will show that

$$\|(\mathcal{U} - \mathcal{T})^{-1}\| \leq C \left(\min \left\{ (1 - v^2 + |b|^2)^2, (1 + v^2 - |b|^2)^2 \right\} \right)^{-1}, \tag{6.52}$$

which is equivalent to proving that

$$\|(\mathcal{U} - \mathcal{T})w\| \geq c \min \left\{ (1 - v^2 + |b|^2)^2, (1 + v^2 - |b|^2)^2 \right\} \|w\| \tag{6.53}$$

for some constant $c > 0$, for any unit vector w. Let $w \in \mathbb{C}^{4 \times 4}$ with $\|w\| = 1$. We will decompose w according to the spectral projections of $\mathcal{T}$,

$$w = P_1 w + P_2 w \tag{6.54}$$

where P_1 is the projection onto $span(h_1, ..., h_k)$ and P_2 is the projection onto its orthogonal complement. The matrices $h_1, ..., h_k$ are defined as the eigenvectors corresponding to the eigenvalues $\pm \|M\|^2$ of $\mathcal{T}$ which correspond to the eigenvalues ± 1 of $\mathscr{P}$. Notice that $k = 3$.

We will consider 2 cases:

Case 1: $\|P_2 w\|^2 \geq c\alpha^2$

Case 2: $\|P_2 w\|^2 < c\alpha^2$,

for some suitably small $c > 0$.

In Case 1, using triangle inequality we obtain

$$\|(\mathcal{U} - \mathcal{T})w\| \geq \|w\| - \|\mathcal{T}w\| = 1 - (\|\mathcal{T}P_1w\|^2 + \|\mathcal{T}P_2w\|^2)^{1/2}$$

$$\geq 1 - (\|\mathcal{T}\|^2\|P_1w\|^2 + \|\mathcal{T}P_2w\|^2)^{1/2}$$

and using that $1 - \sqrt{1 - y} \geq y/2$ for any $y \in [0, 1]$ we obtain:

$$2\|(\mathcal{U} - \mathcal{T})w\| \geq 1 - \|\mathcal{T}\|^2\|P_1w\|^2 - \|\mathcal{T}P_2w\|^2$$

$$= 1 - \|\mathcal{T}\|^2\|P_1w\|^2 - \tau^2\|P_2w\|^2$$

$$\geq 1 - (\|P_1w\|^2 + \|P_2w\|^2) + (1 - \tau^2)\|P_2w\|^2$$

$$= (1 - \tau^2)\|P_2w\|^2. \tag{6.55}$$

Using the definition of Case 1 in the above inequality, we obtain:

$$2\|(\mathcal{U} - \mathcal{T})w\| \geq (1 + \tau)(1 - \tau)c\alpha^2 \geq c\alpha^2. \tag{6.56}$$

. For Case 2, we note:

$$\|(\mathcal{U} - \mathcal{T})w\| = \|(I - \mathcal{U}^*\mathcal{T})w\| \geq \|P_1(1 - \mathcal{U}^*\mathcal{T})w\|$$

$$\geq \|P_1(1 - \mathcal{U}^*\mathcal{T})P_1w\| - \|P_1(1 - \mathcal{U}^*\mathcal{T})P_2w\|.$$

Looking at the first term, we obtain by direct computation:

$$\|P_1(1 - \mathcal{U}^*\mathcal{T})P_1w\|^2 = \left\|\sum_{i=1}^{3}\langle h_i, (1 - \mathcal{U}^*\mathcal{T})\sum_{j=1}^{3}\langle h_j, w\rangle h_j\rangle h_i\right\|^2$$

$$= \left\|\sum_{i,j=1}^{3}\langle h_j, w\rangle(\langle h_i, h_j\rangle - \langle h_i, \mathcal{U}^*\mathcal{T}h_j\rangle)h_i\right\|^2$$

$$= \left\|\sum_{i,j=1}^{3}\langle h_j, w\rangle(\langle h_i, h_j\rangle - \lambda_j\langle h_i, \mathcal{U}^*h_j\rangle)h_i\right\|^2$$

$$= \sum_{i=1}^{3}\left(\sum_{j=1}^{3}\langle h_j, w\rangle(\langle h_i, h_j\rangle - \lambda_j\langle h_i, \mathcal{U}^*h_j\rangle)\right)^2$$

We expand now everything to get that:

$$\|P_1(1 - \mathcal{U}^*\mathcal{T})P_1 w\|^2$$

$$= \left(\langle h_1, w\rangle(1 + v^2) - |b|^2\langle h_2, w\rangle\right)^2 + \left(-|b|^2\langle h_1, w\rangle(1 + v^2)\langle h_2, w\rangle\right)^2$$

$$+ \langle h_3, w\rangle^2(1 - v^2 + |b|^2)^2$$

$$= \left((1 + v^2)^2 + |b|^4\right)\langle h_1, w\rangle^2 + \left((1 + v^2)^2 + |b|^4\right)\langle h_2, w\rangle^2 - 4|b|^2(1 + v^2)\langle h_1, w\rangle\langle h_2, w\rangle$$

$$+ \langle h_3, w\rangle^2(1 - v^2 + |b|^2)^2$$

$$= \left((1 + v^2 - |b|^2)^2\right)\left(\langle h_1, w\rangle^2 + \langle h_2, w\rangle^2\right) + 2|b|^2(1 + v^2 + |b|^2)(\langle h_1, w\rangle - \langle h_2, w\rangle)^2$$

$$+ \langle h_3, w\rangle^2(1 - v^2 + |b|^2)^2$$

$$\geq \|P_1 w\|^2\min\left\{(1 - v^2 + |b|^2)^2, (1 + v^2 - |b|^2)^2\right\}. \tag{6.57}$$

Now to upper bound the second term:

$$\left\|\sum_{i=1}^3\langle h_i, \mathcal{U}^*\mathcal{T}P_2 w\rangle h_i\right\|^2 = \sum_{i=1}^3\langle h_i, \mathcal{U}^*\mathcal{T}P_2 w\rangle^2 = \sum_{i=1}^3\langle P_2\mathcal{U}h_i, \mathcal{T}P_2 w\rangle^2$$

$$\leq \|\mathcal{T}\|^2\|P_2 w\|^2\sum_{i=1}^3\|P_2\mathcal{U}h_i\|^2\leq 3\|\mathcal{T}\|^2 c\alpha^2 \leq c\alpha^2 \tag{6.58}$$

From (6.30), we notice that $0 \leq \alpha \leq 1$, and we have that:

$$\|P_1 w\|^2 = 1 - \|P_2 w\|^2\geq 1 - c\alpha^2. \tag{6.59}$$

Putting this together, we obtain that:

$$\|(\mathcal{U} - \mathcal{T})w\|\geq |\alpha|\sqrt{1 - c\alpha^2} - c|\alpha|\geq ca^2 \tag{6.60}$$

which yields the result. $\qquad\square$

Corollary 6.7.2 (Linear stability estimate in the bulk). *Let $\delta, \kappa \in (0, 1)$. Then uniformly, for all $\zeta \in S_{\tau,\delta,\kappa}$ and $\eta \in (0, 1]$, we have:*

$$\|\mathcal{L}^{-1}|_{A_+}\|\lesssim_{\delta,\kappa} 1 \tag{6.61}$$

Proof. It follows from (6.51) and (6.30), and noting that

$$1 - v^2 + |b|^2 = \frac{\eta}{\eta + v} + 2|b|^2, \text{ and}$$

$$1 + v^2 - |b|^2 = \frac{\eta}{\eta + v} + 2v^2.$$

Notice that we have $\alpha = 0$ when $\eta \to 0$ and $u = b = 0$, i.e. when we are near the origin, that's why we exclude it from our domain when we get our local laws. $\square$

6.8 The least singular value problem

We prove here an important theorem about the least singular value of the matrix $\mathbf{Y_N} - \zeta$.

Not so surprisingly, the Dirac matrix

$$\mathbf{Y_N} = \begin{pmatrix} 0 & X_1 \\ X_2^* & 0 \end{pmatrix}$$

has almost elliptic correlations. This can be seen as follows. We pick elements $\mathbf{X_1}_{ij}$ and $\mathbf{X_2}_{ab}^*$ for $i, j = 1, ..., N$ and $a, b = 1, ..., N$ and compute that:

$$X_{1ij}X_{2ab}^* = (1 + \tau)P_{ij}P_{ab}^* - (1 - \tau)Q_{ij}Q_{ab}^* - \sqrt{1 - \tau^2}P_{ij}Q_{ab}^* + \sqrt{1 - \tau^2}Q_{ij}P_{ab}^*$$

$$= (1 + \tau)P_{ij}\overline{P_{ba}} - (1 - \tau)Q_{ij}\overline{Q_{ba}} - \sqrt{1 - \tau^2}P_{ij}\overline{Q_{ba}} + \sqrt{1 - \tau^2}Q_{ij}\overline{P_{ba}}$$

If we take expectation in this expression the only non-zero terms in this correlation will appear for $b = i$ and $a = j$, which means that:

$$N\mathbb{E}X_{1ij}X_{2ab}^* = \begin{cases} \tau, & \text{if } a = j, b = i \\ 0, & \text{otherwise} \end{cases} \tag{6.62}$$

The matrix $\sqrt{N}\mathbf{Y_N}$ however doesn't fully satisfy the assumptions of [[48],Theorem 1.9] with the elliptic correlation parameter $\tau \in [0, 1]$, (see also [34]), because of the deterministic zero entries.

Due to this we are assuming a bounded density on the entries (see 6.1) and follow the proof in [6] to deduce a lower bound for the least singular value of $\mathbf{Y_N} - \zeta$ with high probability.

Theorem 6.8.1 (Smallest singular value of $\mathbf{Y_N} - \zeta$.). *Let $\sigma_{2N}(\zeta)$ denote the least singular value of the random matrix $\mathbf{Y_N} - \zeta$. Then for any $B > 0$ there exists $A > 0$ and $C > 0$ such that*

$$\mathbb{P}\left(\sigma_{2N}(\zeta) \leq N^{-A}\right) \leq CN^{-B},$$

uniformly for all suitably large $N \in \mathbb{N}$ and bounded $\zeta \in \mathbb{C}$.

Proof. We will use assumption 6.1. We begin with the usual estimate about the least singular value connecting it with the rows of the matrix $\mathbf{Y}_N - \zeta$. This estimate was firstly established in [56] and the general technique can be found for example in (Lemma 4.12, [11]) and (Proposition 7.1, [6]):

$$\sigma_{2N}(\zeta) \geq (2N)^{-1/2} \min_{i \in [2N]} \text{dist}(R_i, R_{-i}),$$

where $R_1, ..., R_{2N}$ are the rows of $\mathbf{Y}_N - \zeta$. Continuing with a union bound, we get that, for $0 < u < 1$:

$$\mathbb{P}(\sigma_{2N}(\zeta) \leq u) \leq 2N \max_{i \in [2N]} \mathbb{P}[(2N)^{-1/2}\text{dist}(R_i, R_{-i}) \leq u] \tag{6.63}$$

We fix $i \in [2N]$ and pick a unit vector y orthogonal to R_{-i} and measurable with respect to $A_i := \{R_j \mid j \neq i\}$. We remind that $R_{-i} := \text{span}(A_i)$. We notice that:

$$\text{dist}(R_i, R_{-i}) = \|\pi_i(R_i)\|\cdot\|y\| \geq |\langle R_i, y \rangle|,$$

by the Cauchy-Schwartz inequality, where π_i is the orthogonal projection onto the orthogonal complement of R_{-i}. Since y is normalized there exists $j \in [2N]$ such that $|y_j| = \max_k |y_k| \geq (2N)^{-1/2}$. We then estimate that:

$$\mathbb{P}\left(|\langle R_i, y \rangle| \leq u\sqrt{2N}\right) = \mathbb{E}\left[\mathbb{P}\left(|\langle R_i, y \rangle| \leq u\sqrt{2N}\Big|y\right)\right]. \tag{6.64}$$

We now estimate the inequality $|\langle R_i, y \rangle| \leq u\sqrt{2N}$ as:

$$|\langle R_i, y \rangle| = \left| \sum_k \overline{x_{ik}} y_k - \overline{\zeta} y_i \right|,$$

where the sum is over N indices and x_{ik} are elements of either X_1 or X_2^*.

To deduce the bound $u\sqrt{2N}$ it is enough that:

$$\left| \overline{x_{ik}} y_k - \frac{\overline{\zeta}}{N} y_i \right| \leq \frac{1}{N} u\sqrt{2N}, \quad \text{for each } k,$$

or:

$$|y_k| \left| x_{ik} - \frac{\zeta}{N} \frac{\overline{y_i}}{\overline{y_k}} \right| \leq \frac{1}{N} u\sqrt{2N}, \quad \text{for each } k \text{ with } y_k \neq 0.$$

But,

$$|y_k| \left| x_{ik} - \frac{\zeta}{N} \frac{\overline{y_i}}{\overline{y_k}} \right| \leq |y_j| \left| x_{ik} - \frac{\zeta}{N} \frac{\overline{y_i}}{\overline{y_k}} \right|,$$

so it is enough that:

$$\left| \sqrt{N} x_{ik} - \frac{\zeta}{\sqrt{N}} \frac{\overline{y_i}}{\overline{y_k}} \right| \leq \frac{\sqrt{2} u}{|y_j|}, \quad \text{for each } k \text{ with } y_k \neq 0, \tag{6.65}$$

Thus,

$$\mathbb{P}\left(|\langle R_i, y \rangle| \leq u\sqrt{2N} \Big| y \right) \leq \sum_k \mathbb{P}\left(\left| \sqrt{N} x_{ik} - \frac{\zeta}{\sqrt{N}} \frac{\overline{y_i}}{\overline{y_k}} \right| \leq \frac{\sqrt{2} u}{|y_j|} \right)$$

$$= \sum_k \int_{B\left(\frac{\zeta}{\sqrt{N}} \frac{\overline{y_i}}{\overline{y_k}}, \frac{\sqrt{2}u}{|y_j|} \right)} \psi(z) d^2 z \leq N \left(\pi \frac{2u^2}{|y_j|^2} \right)^{\frac{q-1}{q}} \|\psi\|_q$$

$$\leq N^{\kappa+1} \left(\frac{\pi u^2}{N} \right)^{\frac{q-1}{q}} = \pi N^{\kappa - \frac{1}{q}} u^{2 - \frac{2}{q}}.$$

From (6.63), we deduce that, for every $u \in (0, 1)$:

$$\mathbb{P}\left(\sigma_{2N}(\zeta) \leq u \right) \leq 2\pi N^{\kappa + 1 - \frac{1}{q}} u^{2 - \frac{2}{q}}. \tag{6.66}$$

Let $B > 0$. We pick $A > 0$ and $q > 1$ such that:

$$\kappa + 1 - \frac{1}{q} - 2A + \frac{2A}{q} \leq -B,$$

or:

$$A \geq \frac{B + \kappa + 1 - \frac{1}{q}}{2 - \frac{2}{q}},$$

and then choose $u = N^{-A} \in (0, 1)$ to get the result we want. $\qquad\square$

6.9 Universality results

We now perform the change of variables to deduce the local law for the correlated covariance matrix. We follow the linearization of a random matrix technique from [51].

Proof of Theorem 6.3.1. For the Dirac matrix $\mathbf{Y_N}$, it holds that:

$$\mathbf{Y_N^2} = \begin{pmatrix} X_1 X_2^* & 0 \\ 0 & X_2^* X_1. \end{pmatrix}$$

Notice that $X_1 X_2^*$ has the same eigenvalues with $X_2^* X_1$. Define now a test function $f : \mathbb{C} \mapsto \mathbb{C}$ with corresponding zoom function $f_{\widehat{\zeta}_0, a}$ for $\widehat{\zeta}_0 \in \widehat{S}_{\tau, \delta, \kappa}$ and $a \in (0, 1/2)$ such that $\|\Delta f\|_{L^{2+a}} \leq N^D \|\Delta f\|_{L^1}$ for some $D \in \mathbb{N}$. Observe that:

$$\begin{aligned}
\int_{\mathbb{C}} f_{\widehat{\zeta}_0, a}(z) d\widehat{\mu}_N &= \frac{1}{N} \sum_{\zeta \in Spec(X_1 X_2^*)} f_{\widehat{\zeta}_0, a}(\zeta) \\
&= \frac{1}{2N} \sum_{\zeta \in Spec(\mathbf{Y_N^2})} f_{\widehat{\zeta}_0, a}(\zeta) \\
&= \int_{\mathbb{C}} f_{\widehat{\zeta}_0, a}(z^2) d\mu_N(z),
\end{aligned}$$

where $\widehat{\mu}_N$ is the counting measure for the correlated covariance matrix, whereas μ_N is the counting measure for the Dirac matrix. By Theorem 6.3.2 we get the local weak convergence $\mu_N \to \mu$, where μ is the measure for the spectrum of the Dirac matrix,

which means that:

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{\zeta\in Spec(X_1 X_2^*)} f_{\widehat{\zeta_0},a}(\zeta) - \int_{\mathbb{C}} f_{\widehat{\zeta_0},a}(z^2)d\mu(z)\right| \leq N^{-1+2a+\epsilon}\right) \geq 1 - CN^{-\nu}, \quad (6.67)$$

for any $\epsilon > 0$, $\nu \in \mathbb{N}$ and some suitable $C > 0$.

We now perform the change of variables $z^2 \mapsto z$. The Jacobian of the transformation is equal to:

$$|J(z)| = \left|\frac{d}{dz}(\sqrt{z})\right|^2 = \frac{1}{4|z|}. \quad (6.68)$$

We are left to prove the transformation from the centered ellipse to the shifted one. Indeed, by firstly expanding the elliptic spectral domain of the Dirac matrix to an equivalent form and then introducing complex variables $z = x + iy$ we arrive at the desired domain:

$$\frac{x^2}{(1+\tau)^2} + \frac{y^2}{(1-\tau)^2} - 1 \leq 0 \Leftrightarrow$$

$$\left[\frac{x^2}{(1+\tau)^2} + \frac{y^2}{(1-\tau)^2} - 1\right]\left[\frac{x^2}{(1-\tau)^2} + \frac{y^2}{(1+\tau)^2} + 1\right] \leq 0 \Leftrightarrow$$

$$\frac{x^4}{(1-\tau^2)^2} + \frac{y^4}{(1-\tau^2)^2} + \frac{x^2 y^2}{(1+\tau)^4} + \frac{x^2 y^2}{(1-\tau)^4} + \frac{x^2 - y^2}{(1+\tau)^2} - \frac{x^2 - y^2}{(1-\tau)^2} \leq 1 \Leftrightarrow$$

$$\frac{x^4 + y^4}{(1-\tau^2)^2} + \frac{(1+\tau)^4 + (1-\tau)^4}{(1-\tau^2)^4}x^2 y^2 + \frac{(1-\tau)^2 - (1+\tau^2)}{(1-\tau^2)^2}(x^2 - y^2) \leq 1 \Leftrightarrow$$

$$x^4 + y^4 - 4\tau(x^2 - y^2) + 2\frac{1 + 6\tau^2 + \tau^4}{(1-\tau^2)^2}x^2 y^2 \leq (1-\tau^2)^2 \Leftrightarrow$$

$$x^4 + y^4 - 4\tau(x^2 - y^2) + 2\frac{(1-\tau^2)^2 + 8\tau^2}{(1-\tau^2)^2}x^2 y^2 \leq (1-\tau^2)^2 \Leftrightarrow$$

$$(x^2 + y^2)^2 + \frac{16\tau^2}{(1-\tau^2)^2}x^2 y^2 - 4\tau(x^2 - y^2) \leq (1-\tau^2)^2.$$

In this more convenient form we introduce complex variables before making the substitution:

$$(x^2 + y^2)^2 + \frac{16\tau^2}{(1-\tau^2)^2}x^2y^2 - 4\tau(x^2 - y^2) \leq (1-\tau^2)^2 \Leftrightarrow$$

$$|z|^4 + \frac{16\tau^2}{(1-\tau^2)^2}x^2y^2 - 4\tau(z^2 - 2ixy) \leq (1-\tau^2)^2 \Leftrightarrow$$

$$|z|^4 - 4\tau z^2 + 8\tau i\frac{z+\bar{z}}{2}\frac{z-\bar{z}}{2i} + \frac{16\tau^2}{(1-\tau^2)^2}\left(\frac{z+\bar{z}}{2}\frac{z-\bar{z}}{2i}\right)^2 \leq (1-\tau^2)^2 \Leftrightarrow$$

$$|z|^4 - 4\tau z^2 + 2\tau(z^2 - \bar{z}^2) - \frac{\tau^2}{(1-\tau^2)^2}(z^2 - \bar{z}^2)^2 \leq (1-\tau^2)^2.$$

This is the right place to make the substitution $z^2 \mapsto z$ and then we will get back to real variables, expand everything and complete the squares. We arrive at the new domain $\widehat{S}_\tau$ as follows:

$$|z|^2 - 4\tau z + 2\tau(z - \bar{z}) - \frac{\tau^2}{(1-\tau^2)^2}(z - \bar{z})^2 \leq (1-\tau^2)^2 \Leftrightarrow$$

$$x^2 + y^2 - 4\tau x + \frac{4\tau^2}{(1-\tau^2)^2}y^2 \leq 1 + \tau^4 - 2\tau^2 \Leftrightarrow$$

$$x^2 - 4\tau x + 2\tau^2 + \frac{(1-\tau^2)^2y^2 + 4\tau^2y^2}{(1-\tau^2)^2} \leq 1 + \tau^4 \Leftrightarrow$$

$$x^2 - 4\tau x + 2\tau^2 + \frac{y^2}{(1-\tau^2)^2}(1+\tau^2)^2 \leq 1 + \tau^4 \Leftrightarrow$$

$$(x - 2\tau)^2 + \frac{y^2}{(1-\tau^2)^2}(1+\tau^2)^2 \leq (1+\tau^2)^2 \Leftrightarrow$$

$$\frac{(x - 2\tau)^2}{(1+\tau^2)^2} + \frac{y^2}{(1-\tau^2)^2} \leq 1.$$

That means that

$$\int_{\mathbb{C}} f_{\widehat{\zeta}_0,a}(z^2)d\mu(z) = \int_{\mathbb{C}} f_{\widehat{\zeta}_0,a}(z)d\widehat{\mu}(z), \tag{6.69}$$

as we wanted, where we used the fact that the transformation $z \mapsto z^2$ maps the complex plane two times onto itself and the identity for the Jacobian (6.68) to retrieve the measure $\widehat{\mu}$ for the correlated covariance matrix.

Inequality 6.67 together with equality 6.69 implies theorem 6.3.1. $\qquad\square$

Chapter 7

Conclusion

7.1 Overview

In this thesis we started by giving some preliminary definitions and results about the usual deterministic matrices of linear algebra and functional analysis. We then moved on to define them as random objects in some proper probability spaces in which they are called random matrices.

We then gave definitions regarding the convergence behaviour of a sequence of usual deterministic measures to another measure. We managed to transform the space of all probability measures in a field into a measurable space, which allowed us to study measure-valued random variables, which are called random measures. Just like the convergence behaviour of usual deterministic measures, we discussed about the probabilistic convergence of a sequence of random measures to another deterministic measure. A classical example of a random measure in random matrix theory which encodes a lot of information about the eigenvalues of a random matrix as we saw was the empirical spectral measure, which as a sequence depends on the dimension of the matrix. Its convergence properties as well as its deterministic limiting measure can give valuable insights about the initial random matrix.

After that, we defined an appropriate transformation of a measure to a complex analytic function which encodes a lot of the properties of the measure, such as its convergence. This was about the Stieltjes tranformation which gives a 1-1 correspondence between each finite Borel measure and its Stieltjes function. We saw that there were easy formulas to recover the measure by its Stieltjes transformation, while the limiting measure corresponds to the limiting Stieltjes function in terms of uniform convergence of functions. This means that in order to establish the convergence of the measures it is enough to establish the convergence of the Stieltjes functions. What's more, we saw that this technique is specifically useful for establishing local laws for the limiting measure, as the Stieltjes function also encodes the scale of the convergence of the measures.

We used the Stieltjes transform technique to analyse the convergence of the empirical spectral measure of a random sample covariance matrix to the deterministic Marchenko-Pastur measure. Initially, we proved an optimal rate of convergence of the Stieltjes functions of the sequence of the spectral measures to the Stieltjes function of the Marchenko-Pastur distribution. We proved this in the optimal scaling for the Stieltjes functions. Special care was taken for the concentration of the eigenvalues around zero and the corresponding singularity of the limiting measure. Using this functional convergence we could go back to the measure-theoretic convergence of the spectral measures to the Marchenko-Pastur distribution and establish it in an almost optimal rate. We then used this result to prove the rigidity of the eigenvalues depending on their position with special treatment given near zero. Their locations turn out to be really close to the ones predicted by the Marchenko-Pastur measure while the size of the fluctuations proved was almost optimal.

In the second part of this thesis, we started analysing the Dyson equation method. It turns out that the Stieltjes transform method is a special case of this technique, as the random empirical spectral measure gets Stieltjes-transformed to the complex function which matches with the trace of the resolvent of the initial random matrix. The Dyson equation method is more general as it involves the treatment of the resolvent as a whole matrix and not just its trace. Of course, this means that the resolvent matrix

includes almost all the information about the empirical spectral measure in which we are interested. The Dyson equation method gives a self-consistent matrix equation for the resolvent matrix which takes into account only the first and second order correlations of the initial random matrix. A stability analysis is always needed then for such a technique so as to prove that the solution matrix of the Dyson equation remains close to the resolvent matrix subject to perturbations. If some assumptions are satisfied for the initial random matrix, then the method works well and we get as a theorem that the solution matrix is indeed close to the resolvent matrix and admits now a Stieltjes transform representation of a matrix whose trace can give us the limit measure of the empirical spectral measure sequence.

We used the Dyson equation technique in a random matrix model whose limitng spectral measure lies on the complex plane and interpolates between the second power of the circular law, the second power of a shifted elliptical law and the Marchenko-Pastur distribution. There were previous results about this limit measure when the entries of the random matrix model were Gaussian random variables. The Dyson equation technique could then be used for proving the same result for a universal case of random variables. We began by hermitizing the random matrix and setting up a Dyson equation for the new hermitized matrix. We solved this equation and proved that the trace of the solution was close to the trace of the resolvent matrix of the initial random matrix. For the solution, we needed to apply a stability analysis. After that, we were able to retrieve the same limit measure from the trace of the solution matrix while also using a least singular value control of the hermitized matrix.

7.2   Further directions

There were some missing targets while producing all the previous results. Specifically, in the local Marchenko-Pastur analysis, we wanted to prove a rate of convergence of the rate of $\mathcal{O}\left(\frac{\sqrt{\log N}}{N}\right)$ and a fluctuation estimate for the bulk eigenvalues of order $\mathcal{O}\left(\frac{\sqrt{\log N}}{N}\right)$ as these are the optimal bounds and they have been proved for the Wishart ensemble. Secondly, in the Correlated Covariance matrices, the initial goal was to prove everything for rectangular random matrices X_1 and X_2 and not just square ones, as was the case in the previous result for the non-Hermitian Wishart ensemble. What's more, our analysis avoided some local laws around the origin of the spectrum of this ensemble as well as in its edges.

We invite all random matrix theory researchers to deal with these issues by providing even more refined results in this analysis while advancing and pushing forward this rather new and exciting scientific field.

Bibliography

[1] Oskari Ajanki, László Erdős, and Torben Krüger. Quadratic vector equations on complex upper half-plane. *Memoirs of the American Mathematical Society*, 261(1261), 2019. 99

[2] Oskari Ajanki, Torben Krüger, and László Erdős. Singularities of solutions to quadratic vector equations on the complex upper half-plane. *Communications on Pure and Applied Mathematics*, 70(9):1672–1705, 2017. 134

[3] Oskari H Ajanki, László Erdős, and Torben Krüger. Stability of the matrix Dyson equation and random matrices with correlations. *Probability Theory and Related Fields*, 173:293–373, 2019. 99, 105, 109, 110, 112

[4] Gernot Akemann, Jinho Baik, and Philippe Di Francesco. *The Oxford Handbook of Random Matrix Theory*. Oxford University Press, 2015. 7

[5] Gernot Akemann, Sung-Soo Byun, and Nam-Gyu Kang. A non-Hermitian generalisation of the Marchenko–Pastur distribution: from the circular law to multi-criticality. *Annales Henri Poincaré*, 22:1035–1068, 2021. 20, 21, 117, 118

[6] Johannes Alt and Torben Krüger. Inhomogeneous circular law for correlated matrices. *Journal of Functional Analysis*, 281(7):109120, 2021. 22, 139

[7] Johannes Alt and Torben Krüger. Local elliptic law. *Bernoulli*, 28(2):886–909, 2022. 20, 21, 22, 121, 122, 123, 124, 125, 126, 127, 128, 130, 131, 132, 134

[8] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2009. 37

[9] M. Artin. *Algebra*. Pearson Education, 2011. 25

[10] Florent Benaych-Georges and Antti Knowles. Lectures on the local semi–circle law for wigner matrices. *arXiv: Probability*, 2016. 42

[11] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probability Surveys*, 9:1 – 89, 2012. 139

[12] Thomas Bothner and Alex Little. The complex elliptic ginibre ensemble at weak non-hermiticity: edge spacing distributions. *arXiv:2208.04684*, 2022. 12, 14

[13] Jennifer Bryson, Roman Vershynin, and Hongkai Zhao. Marchenko–Pastur law with relaxed independence conditions. *Random Matrices: Theory and Applications*, 10(04):2150040, 2021. 11

[14] Claudio Cacciapuoti, Anna Maltsev, and Benjamin Schlein. Local Marchenko-Pastur law at the hard edge of Sample Covariance Matrices. *Journal of Mathematical Physics*, 54(4):043302, 2013. 17, 18, 51, 55, 69

[15] Claudio Cacciapuoti, Anna Maltsev, and Benjamin Schlein. Bounds for the Stieltjes transform and the density of states of Wigner matrices. *Probability Theory and Related Fields*, 163(1):1–59, 2015. 17, 18, 54, 67, 78, 80, 84, 87, 88, 89, 90

[16] Kai Lai Chung. *A course in Probability Theory*. Academic press, 2001. 36

[17] R.M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002. 28, 29

[18] Rick Durrett. *Probability: Theory and Examples*. Cambridge university press, 2019. 36

[19] László Erdős, Antti Knowles, and Horng-Tzer Yau. Averaging fluctuations in resolvents of random band matrices. *Annales Henri Poincaré*, 14(8):1837–1926, 2013. 88

[20] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. The local semi–circle law for a general class of Random Matrices. *Electronic Journal of Probability*, 18:1–58, 2012. 88

[21] László Erdős, Torben Krüger, and Dominik Schröder. Random matrices with slow correlation decay. *Forum of Mathematics, Sigma*, 7:e8, 2019. 99, 105, 108, 109, 128, 129, 130, 132

[22] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin. The local relaxation flow approach to universality of the local statistics for Random Matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(1):1 – 46, 2012. 17

[23] László Erdős and Dominik Schröder. Fluctuations of functions of Wigner matrices. *Electronic Communications in Probability*, 21:1–15, 2016. 92

[24] László Erdos, Horng-Tzer Yau, and Jun Yin. Universality for generalized Wigner matrices with Bernoulli distribution. *arXiv:1003.3813*, 2010. 88

[25] László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229(3):1435–1515, 2012. 88

[26] László Erdős. The matrix Dyson equation and its applications for Random Matrices. *Random Matrices*, 2019. 99

[27] Adina Roxana Feier. *Methods of proof in Random Matrix Theory*. Harvard University, 2012. 9

[28] Michael Fleermann. *Global and local semi–circle laws for Random Matrices with correlated entries*. PhD thesis, FernUniversität in Hagen, 2019.

[29] Y. Fyodorov. Random Matrix theory. *Scholarpedia*, 6(3):9886, 2011.

[30] Fritz Gesztesy and Eduard Tsekanovskii. On matrix–valued Herglotz functions. *Mathematische Nachrichten*, 218(1):61–138, 2000. 111, 112

[31] Jean Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *Journal of Mathematical Physics*, 6(3):440–449, 1965. 12

[32] Vyacheslav L Girko. Circular law. *Theory of Probability & Its Applications*, 29(4):694–706, 1985. 12

[33] Sheng Gong and Youhong Gong. *Concise complex analysis (revised edition)*. World Scientific Publishing Company, 2007.

[34] Friedrich Götze, Alexey Naumov, and Alexander Tikhomirov. On minimal singular values of random matrices with correlated entries. *Random Matrices: Theory and Applications*, 4(02):1550006, 2015. 138

[35] Friedrich Götze and Alexander Tikhomirov. Optimal bounds for convergence of expected spectral distributions to the semi-circular law. *Probability Theory and Related Fields*, 165(1-2):163–233, 2016. 18, 58, 71, 72, 76, 77, 80

[36] Jonas Gustavsson. Gaussian fluctuations of eigenvalues in the GUE. In *Annales de l'IHP Probabilités et statistiques*, volume 41, pages 151–178, 2005. 20

[37] J William Helton, Reza Rashidi Far, and Roland Speicher. Operator-valued semi-circular elements: solving a quadratic matrix equation with positivity constraints. *International Mathematics Research Notices*, 2007(9), 2007. 99, 111, 124

[38] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002. 53

[39] Olav Kallenberg. *Foundations of modern Probability*. Springer, 1997. 27

[40] Olav Kallenberg. *Random measures, theory and applications*. Springer, 2017. 37

[41] Vladislav Kargin and Elena Yudovina. Lecture notes on Random Matrix Theory. *arXiv:1305.2153*, 2013.

[42] Alexei M Khorunzhy, Boris A Khoruzhenko, and Leonid A Pastur. Asymptotic properties of large random matrices with independent entries. *Journal of Mathematical Physics*, 37(10):5033–5060, 1996. 100

[43] Jiri Lebl. *Tasty bits of several complex variables*. Lebl, Jiri, 2019. 43, 44

[44] Vladimir Alexandrovich Marchenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967. 10, 48, 117

[45] D Martin and LV Ahlfors. *Complex Analysis*. McGraw-Hill, New York, 1966.

[46] Madan Lal Mehta. *Random Matrices*. Elsevier, 2004. 8

[47] Alexey Naumov. Elliptic law for real random matrices. *Moscow University Computational Mathematics and Cybernetics*, pages 31–38, 2013.

[48] Hoi H Nguyen and Sean O'Rourke. The elliptic law. *International Mathematics Research Notices*, 2015(17):7620–7689, 2015. 138

[49] Sean O'Rourke and Alexander Soshnikov. Products of independent non-Hermitian random matrices. *Electronic Journal of Probability*, 16, 2011. 116

[50] Adam Osekowski. A note on Burkholder-Rosenthal inequality. *Bull. Pol. Acad. Sci. Math*, 60(2):177–185, 2012. 77

[51] Sean O'Rourke, David Renfrew, Alexander Soshnikov, and Van Vu. Products of independent elliptic random matrices. *Journal of Statistical Physics*, 160(1):89–119, 2015. 22, 141

[52] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962. 41

[53] Leonid A Pastur. Spectra of random self–adjoint operators. *Russian mathematical surveys*, 28, 1973.

[54] Natesh S Pillai and Jun Yin. Universality of covariance matrices. *The Annals of Applied Probability*, 24(3):935–1001, 2014. 17, 84, 89

[55] Åke Pleijel. On a theorem by P. Malliavin. *Israel Journal of Mathematics*, 1(3):166–168, 1963. 92

[56] Mark Rudelson and Roman Vershynin. The Littlewood–Offord problem and invertibility of Random Matrices. *Advances in Mathematics*, 218(2):600–633, 2008. 139

[57] Walter Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., USA, 1987. 43

[58] Peter J Smith. A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *The American Statistician*, 49(2):217–218, 1995.

[59] Hans Juergen Sommers, Andrea Crisanti, Haim Sompolinsky, and Yaakov Stein. Spectrum of large random asymmetric matrices. *Physical review letters*, 60(19):1895, 1988. 13

[60] Zhonggen Su. Gaussian fluctuations in complex sample covariance matrices. *Electronic Journal of Probability*, 11:1284–1320, 2006. 20

[61] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Society, 2023. 32, 34

[62] AN Tikhomirov. Rate of convergence of the expected spectral distribution function to the Marchenko–Pastur law. *Siberian Advances in Mathematics*, 19:277–286, 2009. 18, 58, 71

[63] Claire Voisin. *Hodge Theory and Complex Algebraic Geometry*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002. 45

[64] Eugene P Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958. 9