UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

PhD thesis

Multivariate Polynomials and Rational Functions of Random Matrices - A Case Study

Jacob Fronk

Advisor: Torben Krüger

Submitted: 30 September 2023

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen

Jacob Fronk
QMATH
Department of Mathematical Sciences
University of Copenhagen
jf@math.ku.dk
jacob.fronk@gmail.com

Date of submission	September 30, 2023
Edited for print	December 1, 2023
Date of defense	December 15, 2023
Advisor	**Torben Krüger** University of Copenhagen, Copenhagen, Denmark University of Erlangen–Nuremberg, Erlangen, Germany
Assessment committee	**Albert H. Werner** (Chair) University of Copenhagen, Copenhagen, Denmark **Johannes Alt** University of Bonn, Bonn, Germany **Margherita Disertori** University of Bonn, Bonn, Germany

ISBN 978-87-7125-224-8

Abstract

Local laws describe the phenomenon that eigenvalues of random matrices converge to deterministic densities on scales just slightly above the typical eigenvalue spacing as the dimension of the matrices tends to infinity. This thesis explores both the nature of such limiting densities as well as the local laws that describe the convergence to these limiting densities for specific ensembles of both Hermitian and non-Hermitian random matrices.

The class of random matrices we are studying in the Hermitian realm are non-commutative quadratic polynomials of multiple independent Wigner matrices. We prove that, with the exception of some specific reducible cases, the limiting spectral density of the polynomials always has a square root growth at its edges and prove an optimal local law around these edges. Combining these two results, we establish that, as the dimension N of the matrices grows to infinity, the operator norm of such polynomials converges to a deterministic limit with a rate of convergence of $N^{-2/3+o(1)}$. Here, the exponent in the rate of convergence is optimal. For the specific reducible cases, we also provide a classification of all possible edge behaviours.

For non-Hermitian matrices $\mathbf{X}$ and $\mathbf{Y}$ with centered, independent and identically distributed (i.i.d.) entries, it has been shown that the eigenvalues of the rational function $\mathbf{Y}^{-1}\mathbf{X}$ converge to a uniform density after being projected onto the Riemann sphere. We provide a simple proof of a local law for this ensemble by reducing the problem to the local circular law, a well-established result that the eigenvalues of a single centered i.i.d. matrix converge to the uniform distribution on the complex unit disc on all scales larger than the typical eigenvalue spacing.

Resumé

Lokale love beskriver det fænomen, at egenværdier af tilfældige matricer konvergerer til deterministiske tætheder på skalaer, der en smule større end den typiske afstand mellem egenværdier, når matricernes dimension går mod uendelig. Denne afhandling undersøger både arten af sådanne begrænsende tætheder og de lokale love, der beskriver konvergensen til disse begrænsende tætheder for specifikke ensembler af både hermitiske og ikke-hermitiske tilfældige matricer.

Den klasse af tilfældige matricer, vi studerer i det hermitiske område, er ikke-kommutative kvadratiske polynomier af flere uafhængige Wigner-matricer. Med undtagelse af nogle specifikke reducerbare tilfælde beviser vi, at den begrænsende spektrale tæthed af polynomierne altid har en kvadratrodsvækst ved sin rand, og vi beviser en optimal lokal lov omkring randen. Ved at kombinere disse to resultater fastslår vi, at når dimensionen N af matricerne går mod uendelig, konvergerer operatornormen for sådanne polynomier til en deterministisk grænse med en konvergensrate på $N^{-2/3+o(1)}$. Her er eksponenten i konvergenshastigheden optimal. For de specifikke reducerbare tilfælde giver vi også en klassifikation af al mulig randadfærd.

For ikke-Hermitiske matricer $\mathbf{X}$ og $\mathbf{Y}$ med centrerede, uafhængige og identisk fordelte (i.i.d.) indgange er det blevet vist, at egenværdierne af den rationelle funktion $\mathbf{Y}^{-1}\mathbf{X}$ konvergerer til en ensartet tæthed efter projektion på Riemannsfæren. Vi giver et simpelt bevis på en lokal lov for dette ensemble ved at reducere problemet til den lokale cirkulære lov, et veletableret resultat, at egenværdierne for en enkelt centreret i.i.d. matrix konvergerer til den ensartede fordeling på den komplekse enhedsskive på alle skalaer, der er større end den typiske egenværdiafstand.

Contributions and structure

The thesis is split into two parts. The first part, consisting of Chapters 1-3, is dedicated to the study of Hermitian random matrices and mostly focuses on multivariate quadratic polynomials of independent random matrices. Our main results for this part are given in Section 2.2. The second part discusses non-Hermitian random matrix theory and its main result is given in Theorem 5.2.1.

- Chapter 1 provides an introduction to local laws and Hermitian random matrix theory. The resolvent method and the Dyson equation are introduced and we explain in detail how eigenvalue rigidity and eigenvector delocalization can be derived from local laws for the resolvent.
- Chapter 2 consists of a reproduction of the preprint arXiv:2308.16778: [35]. In it, we analyse the edge behaviour of multivariate quadratic polynomials of independent Wigner matrices. We establish the asymptotic behaviour of the limiting spectral density for all such polynomials and prove that, with the exception of some specific reducible cases, their density always shows a square root growth. Combined with an optimal local law, which we also prove around all edges with a square root growth, we then derive a norm convergence rate with an optimal exponent. Note that there is some overlap between the explanations given in Chapter 1 and those given in Chapter 2 due to the self-containedness of Chapter 2. The preprint is a joint work with Torben Krüger and Yuriy Nemish.
- In Chapter 3 we give some examples of polynomials to highlight some notable features of the limiting spectral measure analysed in Chapter 2. The author gratefully acknowledges the help of Paula Belzig in the creation of the plots displayed in this chapter.
- Capter 4 contains a short introduction to non-Hermitian random matrices. In particular, we introduce Girko's trick and explain how it is used to prove the circular law.
- In Chapter 5 we introduce the spherical ensemble, an example of a rational function of non-Hermitian random matrices. We prove a local law for the spherical ensemble and include a short discussion on how this trick can be extended to more general rational functions.

Contents

Chapter 1

Introduction

Random matrices were first introduced by Wishart in 1928, who studied the distribution of random covariance matrices [69]. The natural question about their spectral statistics, that is, the nature of their eigenvalue and eigenvector distributions, was only brought up about thirty years later in a seminal work by Wigner [68].

The crucial observation by Wigner was that as the dimension N of many random matrices grows to infinity, their eigenvalue distribution tends to a deterministic limiting distribution. This statement can be formalized in the following way. Let $\mathbf{H}$ be a random matrix. Throughout this introduction, we always assume that $\mathbf{H}$ is Hermitian, i.e. its entries h_{ij} satisfy $h_{ij} = \bar{h}_{ji}$. Let λ_i, $i \in \{1, \dots, N\}$, denote the eigenvalues of $\mathbf{H} \in \mathbb{C}^{N\times N}$ with algebraic multiplicity. Then all eigenvalues of $\mathbf{H}$ are real and we define the **empirical eigenvalue distribution** ρ_N of $\mathbf{H}$ by

$$\rho_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i}, \tag{1.0.1}$$

where δ_x denotes the normalized point mass at $x \in \mathbb{R}$. Clearly, ρ_N is a probability measure for all $N \in \mathbb{N}$ and since $\mathbf{H}$ is a random matrix, ρ_N is itself a random variable. If there is a deterministic probability measure ρ such that $\lim_{N\to\infty} \rho_N = \rho$ weakly in probability, then we call ρ the **limiting spectral measure** and we say that $\mathbf{H}$ obeys a **global law**.

A famous example of a random matrix ensemble that obeys such a global law is the **Wigner ensemble**. We define Wigner matrices $\mathbf{W} \in \mathbb{C}^{N\times N}$ as follows.

Definition 1.0.1 (Wigner matrix)**.** *Let ξ_0 be a real random variable and let ξ_1 be a random variable independent of ξ_0 that is either real or complex. Let them both be centered, have unit variance and finite moments, i.e.*

$$\mathbb{E}[\xi_i] = 0, \quad \mathbb{E}[|\xi_i|^2] = 1 \quad \textit{and} \quad \mathbb{E}[|\xi_i|^p] \leq C_p \tag{1.0.2}$$

for $i = 0, 1$ and some constant C_p for all $p \in \mathbb{N}$. Here, $\mathbb{E}[\cdot]$ denotes the expectation value of a random variable. A Hermitian matrix $\mathbf{W} \in \mathbb{C}^{N\times N}$ with entries $w_{i,j}$ is called a Wigner matrix if (w_{ij}, w_{ji}) are independent copies of $(\xi_1, \bar{\xi}_1)$ for $i > j$ and

w_{ii} are independent copies of ξ_0. In other words, Wigner matrices are Hermitian random matrices with centered and up to the Hermitian symmetry independent entries with uniform variance.

Remark. *The condition that all moments of ξ_0 and ξ_1 are finite is not necessary for the statements discussed here. We still demand it for simplicity, since we are not trying to optimize the moment conditions.*

Remark. *The $\sqrt{N}$ normalization ensures that most eigenvalues λ_i will be of order* 1*, since we have*

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\lambda_i^2] = \frac{1}{N}\mathbb{E}[\operatorname{Tr}\mathbf{W}^2] = \frac{1}{N}\sum_{i,j=1}^{N}\mathbb{E}[|w_{ij}|^2] = 1. \tag{1.0.3}$$

It has been shown that the limiting spectral measure of $\mathbf{W}$ is given by the semi-circle distribution ρ_{sc}, an absolutely continuous probability measure with density given by

$$\rho_{\mathrm{sc}}(E) = \begin{cases} \frac{1}{\pi}\sqrt{4-E^2} & \text{if } |E| \le 2 \\ 0 & \text{if } |E| > 2. \end{cases} \tag{1.0.4}$$

This is known as the **semi-circle law** and it was first proved by Wigner himself [68]. However, since the eigenvalues of Wigner matrices are all functions of the same random variables, that is, the entries of the matrix, we expect them to be strongly correlated and one can ask the question if the convergence of the empirical density to the limiting spectral measure also occurs on a local scale. More precisely, while a global law implies that

$$\lim_{N\to\infty}\frac{1}{\varepsilon}\left|\{\lambda\in\operatorname{Spec}(\mathbf{W}) : \lambda\in[a,a{+}\varepsilon]\}\right| = \frac{1}{\varepsilon}\int_a^{a+\varepsilon}\rho_{\mathrm{sc}}(E)\mathrm{d}E \tag{1.0.5}$$

in probability for all $\varepsilon > 0$ and $a \in \mathbb{R}$, we ask if

$$\lim_{N\to\infty}\frac{1}{\varepsilon_N}\left|\{\lambda\in\operatorname{Spec}(\mathbf{W}) : \lambda\in[a,a{+}\varepsilon_N]\}\right| = \lim_{N\to\infty}\frac{1}{\varepsilon_N}\int_a^{a+\varepsilon_N}\rho_{\mathrm{sc}}(E)\mathrm{d}E = \rho_{\mathrm{sc}}(a). \tag{1.0.6}$$

also holds in probability for sequences $(\varepsilon_N)_{N\in\mathbb{N}}$ with $\varepsilon_N \to 0$. We call such results **local laws**. It is clear that for a local law to hold for all $a \in \mathbb{R}$, we must require $\varepsilon_N \gg N^{-1}$, as for $\varepsilon_N \sim N^{-1}$, we would start to see the individual eigenvalues and we can no longer expect them to average out their density. However, it turns out to be sufficient to choose $\varepsilon_N = N^{-1+\gamma}$ for fixed $\gamma > 0$ arbitrarily small. This is formalized in the following theorem.

Theorem 1.0.2 (Local law for the empirical eigenvalue density)**.** *Let $\mathbf{W}$ be a Wigner matrix as introduced in Definition 1.0.1. For all $\gamma \in (0,1)$ and all $D \in \mathbb{N}$ there is a $C_{\gamma,D} > 0$ such that*

$$\mathbb{P}\left(N^{1-\gamma}\left|\left|\operatorname{Spec}(\mathbf{W})\cap[a,a{+}N^{-1+\gamma}]\right| - \int_a^{a+N^{-1+\gamma}}\rho_{\mathrm{sc}}(E)\mathrm{d}E\right| \ge N^{-\frac{\gamma}{2}}\right) \le C_{\gamma,D}N^{-D} \tag{1.0.7}$$

holds uniformly for all $a \in \mathbb{R}$.

Notations

We denote the average trace of a matrix $\mathbf{R} \in \mathbb{C}^{n\times n}$ by $\langle \mathbf{R} \rangle = \frac{1}{n} \operatorname{Tr} \mathbf{R}$ and its operator norm by $\|\mathbf{R}\|$. The Hermitian adjoint of $\mathbf{R}$ is denoted by $\mathbf{R}^*$. For vectors $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ we use $\langle \mathbf{v}, \mathbf{w} \rangle$ to denote their standard scalar product. The matrix $\mathbf{R}$ is called positive (semi-)definite if it satisfies $\langle \mathbf{v}, \mathbf{R}\mathbf{v} \rangle > 0$, ($\langle \mathbf{v}, \mathbf{R}\mathbf{v} \rangle \geq 0$) for all $\mathbf{v} \in \mathbb{C}^n \setminus \{0\}$. It is called negative (semi-)definite if $-\mathbf{R}$ is positive (semi-)definite. The identity matrix is denoted by $\mathbf{I}_n \in \mathbb{C}^{n\times n}$. If this does not create any ambiguity, we often compactly write $\zeta \mathbf{I}_n = \zeta$ for multiples of the identity matrix.

The letter z always denotes a complex number. Throughout Chapters 1-3 we always use $E = \operatorname{Re} z$ and $\eta = \operatorname{Im} z$. In Chapters 4 and 5 we deviate from this notation and instead η denotes a positive number independent of z. The imaginary unit is denoted by i to distinguish it from summation indices often denoted by i. We still usually avoid using them together in the same expression, instead opting for other summation indices if the imaginary unit is explicitly displayed.

We say that a sequence of event $A = A_N$, $N \in \mathbb{N}$, holds with high probability if for all $D \in \mathbb{N}$ there is a constant $C = C(D)$ such that we have

$$\mathbb{P}(A) \geq 1 - CN^{-D}. \tag{1.0.8}$$

The notation $[\![n]\!] := \{1, 2, \ldots, n\}$ is used for the natural numbers up to n. If X and Y are positive quantities, we use the notation $X \lesssim Y$ if there is some constant $c > 0$ such that $X \leq cY$. If the constant c depends on any parameters α, we write $X \lesssim_\alpha Y$. In particular, the constant will never depend on the dimension of our random matrices, N. If both $X \lesssim Y$ and $Y \lesssim X$ hold true, we write $X \sim Y$. If $X \lesssim Y$, we sometimes also write $X = \mathcal{O}(Y)$ and if $\lim_{N\to\infty} Y^{-1}X = 0$, the notation $X = o(Y)$ is used.

Note that some more notations that are commonly used only in the preprint that makes up Chapter 2 are given in the dedicated notations section of the preprint in Section 2.1.

1.1 The resolvent method

A method that has proven extraordinarily successful at proving local laws is the resolvent method and we introduce it in this section. It allows for much stronger local law statements than Theorem 1.0.2 and we will show how Theorem 1.0.2 follows as a corollary of Theorem 1.1.2 below. See e.g. [9, 31, 63] and references therein for an overview of the resolvent method and its historical development.

1.1.1 The Stieltjes transform

Let μ be a probability measure with real support. For $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, we define its **Stieltjes transform** m_μ by

$$m_\mu(z) := \int_{\mathbb{R}} \frac{\mu(\mathrm{d}x)}{x - z}. \tag{1.1.1}$$

Furthermore, let $\mathbb{H}$ denote the upper half-plane, i.e.

$$\mathbb{H} := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}. \tag{1.1.2}$$

Since μ has real support, m_μ is well defined on all of $\mathbb{H}$. The following lemma summarizes some basic properties of the Stieltjes transform.

Lemma 1.1.1. *Let μ be a real probability measure and let m_μ be its Stieltjes transform, which was introduced in (1.1.1). Then the following holds true.*

1. *The function m_μ is analytic on all of $\mathbb{H}$.*
2. *The probability measure μ is uniquely defined by its Stieltjes transform on $\mathbb{H}$, i.e. if μ' is another probability measure such that $m_\mu = m_{\mu'}$ on all of $\mathbb{H}$, then we have $\mu = \mu'$.*
3. *Let $(\mu_N)_{N\in\mathbb{N}}$ be a sequence of real probability measures such that*

$$\lim_{N\to\infty} m_{\mu_N}(z) = m_\mu(z) \tag{1.1.3}$$

   for all $z \in \mathbb{H}$. Then, μ_N converges weakly to μ.
4. *For any interval I with neither endpoint on an atom of μ we have*

$$\mu(I) = \lim_{\eta\to 0} \frac{1}{\pi} \int_I \operatorname{Im} m_\mu(E + \mathrm{i}\eta)\mathrm{d}E. \tag{1.1.4}$$

Proof. Statements 1. and 2. are clear. For Statement 3. and Statement 4. see e.g. [6, Chapter 2.4]. □

The resolvent $\mathbf{G}$ of a matrix $\mathbf{H}$ at spectral parameter $z \in \mathbb{C} \setminus \operatorname{Spec}(\mathbf{H})$ is defined by

$$\mathbf{G} := (\mathbf{H} - z)^{-1} \in \mathbb{C}^{N\times N}, \tag{1.1.5}$$

where we have used the notation $z = \mathbf{I}_N z$. Since $\mathbf{H} - z$ is non-invertible if and only if $z \in \operatorname{Spec}(\mathbf{H})$, the resolvent is well defined and for Hermitian $\mathbf{H}$, we have

$\mathrm{Spec}(\mathbf{H}) \subset \mathbb{R}$ and its resolvent is defined on all of $\mathbb{H}$. The Stieltjes transform of the empirical measure ρ_N, introduced in (1.0.1), of $\mathbf{H}$ is given by

$$m_{\rho_N}(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} = \frac{1}{N}\operatorname{Tr}(\mathbf{G}) =: \langle \mathbf{G}\rangle, \tag{1.1.6}$$

where we have introduced the shorthand notation $\langle\cdot\rangle$ for the averaged trace of a matrix. Therefore, we can establish a global law for a random matrix $\mathbf{H}$, if we are able to prove that the averaged trace of its resolvent $\langle\mathbf{G}\rangle$ converges pointwise to the Stieltjes transform $m = m_\rho$ of some deterministic measure ρ and this also proves that ρ is the limiting spectral measure.

1.1.2 The Dyson equation and the local law for the resolvent

To prove convergence of $\langle\mathbf{G}\rangle$, we first look for a promising candidate for the limiting object. To do so, we first consider a Wigner matrix $\mathbf{W}$ with complex Gaussian entries and note that $\mathbf{W}$ and its resolvent $\mathbf{G}$ satisfy

$$\mathbf{WG} = \mathbf{I}_N + z\mathbf{G}. \tag{1.1.7}$$

Using the cumulant expansion formula, [29, Proposition 3.2], combined with the fact that all cumulants of order greater than two vanish for Gaussian random variables, we find that the expectation value of $\mathbf{WG}$ satisfies

$$\mathbb{E}[\mathbf{WG}] = \mathbb{E}[\widetilde{\mathbb{E}}[\widetilde{\mathbf{W}}\nabla_{\widetilde{\mathbf{W}}}\mathbf{G}]] = -\mathbb{E}[\widetilde{\mathbb{E}}[\widetilde{\mathbf{W}}\mathbf{G}\widetilde{\mathbf{W}}]], \tag{1.1.8}$$

where $\widetilde{\mathbf{W}}$ denotes an independent copy of $\mathbf{W}$ and $\widetilde{\mathbb{E}}$ its expectation. $\nabla_R\mathbf{F}(\mathbf{H}) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbf{F}(\mathbf{H}+\varepsilon\mathbf{R})|_{\varepsilon=0}$ denotes the directional derivative for any differentiable matrix valued function $\mathbf{F} : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N}$. In the second step we have used $\nabla_{\mathbf{R}}\mathbf{G} = -\mathbf{GRG}$, which holds since $\mathbf{G}$ is a resolvent. This motivates the definition of $\mathbf{D}$ as an error term by

$$\mathbf{D} := \mathcal{S}[\mathbf{G}]\mathbf{G} + \mathbf{WG} \quad \text{with} \quad \mathcal{S}[\mathbf{R}] := \mathbb{E}[\mathbf{WRW}], \tag{1.1.9}$$

where $\mathcal{S}$ is called the self-energy operator. Indeed, using [29, Proposition 3.2] twice, we find

$$\mathbb{E}|\langle\mathbf{D}\rangle|^2 \lesssim \frac{|\langle\mathbf{G}^*\mathbf{G}\rangle|^2}{N} = \frac{|\langle\operatorname{Im}\mathbf{G}\rangle|}{N\eta}, \tag{1.1.10}$$

strongly indicating that $\mathbf{D}$ is a small object in some sense for $\eta \gg N^{-1}$ (See [29, Equation (4.4) and below] for the detailed calculation). In the last step in (1.1.10) we have used the **Ward identity**

$$\mathbf{G}^*\mathbf{G} = \frac{\operatorname{Im}\mathbf{G}}{\operatorname{Im} z}, \tag{1.1.11}$$

that holds true for the resolvent of all Hermitian matrices. With the definition of the error term (1.1.9), the resolvent $\mathbf{G}$ satisfies the equation

$$\mathbf{I}_N + (z + \mathcal{S}[\mathbf{G}])\mathbf{G} = \mathbf{D}. \tag{1.1.12}$$

If the error term in (1.1.12) vanishes, the equation takes the form

$$\mathbf{I}_N + (z + \mathcal{S}[\mathbf{M}])\mathbf{M} = 0 \tag{1.1.13}$$

and it is known as the **Dyson equation**. The resolvent $\mathbf{G}$ has positive imaginary part for all $z = E + \mathrm{i}\eta \in \mathbb{H}$ since it is a normal matrix and its eigenvalues

$$\frac{1}{\lambda_j - z} = \frac{1}{(\lambda_j - E)^2 + \eta^2}(\lambda_j - E + \mathrm{i}\eta) \tag{1.1.14}$$

have positive imaginary part for all $j \in [\![N]\!]$. As we are interested in solutions to (1.1.13) that are in some sense close to $\mathbf{G}$, we also condition on $\operatorname{Im}\mathbf{M} > 0$. Under this condition it has been shown in [46] that there is a unique solution to (1.1.13).

Remark. *The set-up presented here holds in a much more general setting. For a random matrix* $\mathbf{H}$ *with non-vanishing expectation value the Dyson equation* (1.1.13) *becomes*

$$\mathbf{I}_N + (z - \mathbb{E}[\mathbf{H}] + \mathcal{S}[\mathbf{M}])\mathbf{M} = 0 \quad \textit{with} \quad \mathcal{S}[\mathbf{R}] = \mathbb{E}[(\mathbf{H} - \mathbb{E}[\mathbf{H}])\mathbf{R}(\mathbf{H} - \mathbb{E}[\mathbf{H}])]. \tag{1.1.15}$$

The existence of a unique solution the constraint $\operatorname{Im}\mathbf{M} > 0$ *is guaranteed for all* $z \in \mathbb{H}$ *by [46]. For a rigorous result on the smallness of the error term* (1.1.9) *in appropriate norms for a wide range of random matrix ensembles we refer to [29, Theorem 4.1].*

Recall that the Stieltjes transform of the empirical measure ρ_N is given by $\langle \mathbf{G} \rangle$. Therefore, assuming that $\mathbf{G}$ is close to $\mathbf{M}$ our candidate for the Stieltjes transform of the limiting spectral measure is given by

$$m_{\mathrm{sc}} = \lim_{N\to\infty} \langle \mathbf{M} \rangle. \tag{1.1.16}$$

In case the non-diagonal entries of $\mathbf{W}$ have vanishing second moment, i.e. $\mathbb{E}[\xi_1^2] = 0$, the self-energy operator takes the simple form

$$\mathcal{S}[\mathbf{R}] = \langle \mathbf{R} \rangle \mathbf{I}_N. \tag{1.1.17}$$

Then it is easy to see that $\mathbf{M}$ is of the form $\mathbf{M} = m_{\mathrm{sc}}\mathbf{I}_N$ and m_{sc} satisfies the quadratic equation

$$m_{\mathrm{sc}}^2 + z m_{\mathrm{sc}} + 1 = 0. \tag{1.1.18}$$

For $z \in \mathbb{H}$, the unique solution to (1.1.18) with positive imaginary part is given by

$$m_{\mathrm{sc}} = m_{\mathrm{sc}}(z) = \frac{1}{2}\left(\sqrt{z^2 - 4} - z\right), \tag{1.1.19}$$

where $\sqrt{\cdot}$ denotes the square root function that maps the positive real axis to itself and with $\operatorname{Im}\sqrt{w} \geq 0$ for all $w \in \mathbb{C}$. Using Lemma 1.1.1, we find that m_{sc} is indeed the Stieltjes transform of ρ_{sc} introduced in (1.0.4).

Remark. *For $\mathbb{E}[\xi_1^2] \neq 0$, the self-energy (1.1.17) has an additional $\mathcal{O}(N^{-1})$ term but this does not affect the limiting Stieltjes transform $m_{\mathrm{sc}} = \lim_{N\to\infty}\langle \mathbf{M}\rangle$.*

Indeed, it turns out that $\mathbf{M}$ is a good approximation for $\mathbf{G}$ for large N as can be seen from the following Theorem 1.1.2, below. This kind of result is also referred to as a local law and in fact, we will see that it is a generalization to Theorem 1.0.2. To distinguish the result from Theorem 1.0.2, we denote it by the name **local law for the resolvent** here. For the result in this precise form see [29] (bulk regime and regime away from the spectrum) as well as [3] (edge regime).

Theorem 1.1.2 (Local law for the resolvent)**.** *For a large, fixed $C > 0$ define the set*

$$\mathbb{D}_\gamma := \{z \in \mathbb{H} : \operatorname{Im} z \geq N^{-1+\gamma}, |z| \leq N^C\}. \tag{1.1.20}$$

Let $\mathbf{W}$ be a Wigner matrix as in Definition 1.0.1. For all $\varepsilon, \gamma, D > 0$ and $z = E + \mathrm{i}\eta \in \mathbb{D}_\gamma$ the isotropic local law

$$\mathbb{P}\left(|\langle \mathbf{x}, (\mathbf{G} - m_{\mathrm{sc}}\mathbf{I}_N)\mathbf{y}\rangle| > \|\mathbf{x}\|\|\mathbf{y}\|N^\varepsilon\left(\frac{\operatorname{Im} m_{\mathrm{sc}}}{\sqrt{N\eta}} + \frac{1}{N\eta}\right)\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{1.1.21}$$

holds for all deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$. Moreover, the averaged local law

$$\mathbb{P}\left(|\langle \mathbf{B}(\mathbf{G} - m_{\mathrm{sc}}\mathbf{I}_N)\rangle| > \|\mathbf{B}\|\frac{N^\varepsilon}{N\eta}\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{1.1.22}$$

holds for all deterministic $\mathbf{B} \in \mathbb{C}^{N\times N}$. If additionally $|E| > 2$, an improved averaged local law of the form

$$\mathbb{P}\left(|\langle \mathbf{B}(\mathbf{G} - m_{\mathrm{sc}}\mathbf{I}_N)\rangle| > \|\mathbf{B}\|\left(\frac{N^\varepsilon}{N(\eta+\kappa_E)} + \frac{N^\varepsilon}{(N\eta)^2\sqrt{\eta+\kappa_E}}\right)\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{1.1.23}$$

holds. Here, $\kappa_E := \min\{|E+2|, |E-2|\}$ denotes the distance of E to the edge of the limiting density.

1.2 Eigenvalue rigidity and eigenvector delocalization

Theorem 1.1.2 implies immediately that $\mathbf{W}$ obeys a global law in the sense of (1.0.5). It does however not only do that. For one, the quantitative bounds allow us to establish eigenvalue rigidity, meaning that all eigenvalues are very close to their position predicted by the limiting density. This result is presented in Corollary 1.2.1. Additionally, Theorem 1.1.2 does not only track the Stieltjes transform of the empirical density, i.e. the averaged trace of the resolvent, but also of the entries of the resolvent itself in an arbitrary deterministic basis and thereby also contains information about the eigenvectors of $\mathbf{W}$. This can be used to prove eigenvector delocalization in the form stated in Corollary 1.2.2.

Corollary 1.2.1 (Eigenvalue rigidity)**.** *Let* $\mathbf{W}$ *be a Wigner matrix as introduced in Definition 1.0.1 and denote the classical index of the eigenvalue close to energy* $E \in \mathrm{supp}(\rho_{\mathrm{sc}}) \setminus \partial\mathrm{supp}(\rho_{\mathrm{sc}}) = (-2,2)$ *by*

$$k(E) := \left\lceil N \int_{-\infty}^{E} \rho_{\mathrm{sc}}(E')\mathrm{d}E' \right\rceil, \tag{1.2.1}$$

with $\lceil \cdot \rceil$ *being the ceiling function. All eigenvalues of* $\mathbf{W}$ *are close to their classical position in the following sense. For all* $\varepsilon, D > 0$ *and* $E \in (-2,2)$ *we have*

$$\mathbb{P}\left(\mathrm{sup}_E |\lambda_{k(E)} - E| \geq \min\left\{ \frac{N^{\varepsilon}}{N\kappa_E}, \frac{N^{\varepsilon}}{N^{\frac{2}{3}}} \right\} \right) \lesssim_{D,\varepsilon} N^{-D}, \tag{1.2.2}$$

where $\kappa_E := \min\{|E+2|, |E-2|\}$ *denotes the distance of* E *to the edge of the limiting density.*

Corollary 1.2.2 (Eigenvector delocalization)**.** *Let* $\mathbf{W}$ *be a Wigner matrix as introduced in Definition 1.0.1 and let* $\mathbf{v}$ *be a normalized eigenvector of* $\mathbf{W}$ *with eigenvalue* λ*. For all* $\varepsilon, D > 0$ *we have*

$$\sup_{\mathbf{x} \in \mathbb{C}^{N\times N}, \|\mathbf{x}\|=1} \mathbb{P}\left(|\langle \mathbf{x}, \mathbf{v} \rangle| \geq N^{-\frac{1}{2}+\varepsilon} \right) \lesssim_{D,\varepsilon} N^{-D}. \tag{1.2.3}$$

We prove both Corollary 1.2.1 and Corollary 1.2.2 assuming Theorem 1.1.2. Additionally, we conclude the eigenvalue density form of the local law, Theorem 1.0.2, to prove that it is indeed a direct consequence of Theorem 1.1.2. Our proof of Corollary 1.2.1 follows the strategy of [31, Chapter 11], see also [27, Theorem 7.6]. Before we can start the proof, we need some preparatory notions and lemmata. We start by giving the definition of stochastic domination, a notion of a high probability bound tailored to our needs. It was first introduced in [26] and we use it in the form of [27].

Definition 1.2.3 (Stochastic domination)**.** *Let* $X = (X^{(N)})_{N\in\mathbb{N}}$ *and* $Y = (Y^{(N)})_{N\in\mathbb{N}}$ *be families of non-negative random variables. We say* X *is stochastically dominated by* Y *if for all (small)* $\varepsilon > 0$ *and (large)* $D > 0$

$$\mathbb{P}\left(X^{(N)} \geq N^{\varepsilon} Y^{(N)} \right) \lesssim_{D,\varepsilon} N^{-D}. \tag{1.2.4}$$

We denote this relation by $X \prec Y$*. If the constant in the definition depends on any other parameters* α*, we write* $X \prec_{\alpha} Y$*.*

Next, we give the following lemma that asserts that with very high probability $\mathbf{W}$ has no eigenvalues away from the support of ρ_{sc}. More precisely, we have the following statement.

Lemma 1.2.4 (Exclusion of eigenvalues away from $\mathrm{supp}(\rho_{\mathrm{sc}})$). *Let* $\mathbf{W}$ *be a Wigner matrix as introduced in Definition 1.0.1. For all* $\varepsilon > 0$ *there is no eigenvalue further away from the support of the limiting spectral measure* $\mathrm{supp}(\rho_{\mathrm{sc}}) = [-2, 2]$ *than* $N^{-2/3+\varepsilon}$ *with high probability, i.e. we have*

$$\mathbb{P}\left(\exists \lambda \in \mathrm{Spec}(\mathbf{W}) : |\lambda| \geq 2 + N^{-\frac{2}{3}+\varepsilon}\right) \lesssim_{\varepsilon, D} N^{-D}. \tag{1.2.5}$$

Remark. *The scale* $N^{-2/3}$ *is indeed the optimal scale, since close to an edge, where the density grows like a square root, the typical eigenvalue spacing is* $N^{-2/3}$ *and we expect to see the fluctuation of the individual eigenvalues on that scale.*

Proof. We only prove the absence of eigenvalues λ to the right of the spectrum with $\lambda \geq 2 + N^{-2/3+\varepsilon}$, the absence of eigenvalues with $\lambda \leq -2 - N^{-2/3+\varepsilon}$ follows by symmetry. Now, fix $\varepsilon > 0$ and assume there is a $\lambda \in \mathrm{Spec}(\mathbf{W})$ with $\lambda \geq 2+N^{-2/3+\varepsilon}$. Then, $\mathrm{Im}\,\mathbf{G}$ has eigenvalue η^{-1} at spectral parameter $z = \lambda + \mathrm{i}\eta$. Since $\mathrm{Im}\,\mathbf{G}$ is positive definite this implies

$$\langle \mathrm{Im}\,\mathbf{G} \rangle \geq \frac{1}{N\eta}. \tag{1.2.6}$$

On the other hand we have with high probability from (1.1.23) that

$$\langle \mathrm{Im}\,\mathbf{G} \rangle \leq \mathrm{Im}\, m_{\mathrm{sc}} + |\langle \mathbf{G} \rangle - m_{\mathrm{sc}}| \lesssim \frac{\eta}{\sqrt{\kappa}} + \frac{N^{\frac{\varepsilon}{3}}}{N\kappa} + \frac{N^{\frac{\varepsilon}{3}}}{(N\eta)^2\sqrt{\kappa}} \tag{1.2.7}$$

with $\kappa = \lambda - 2 \geq N^{-2/3+\varepsilon}$. Note that we replaced ε from (1.1.23) by $\frac{\varepsilon}{3}$, which is possible since the statement holds with high probability for $\varepsilon > 0$ arbitrary. We now choose $\eta = N^{-2/3}$, then (1.2.6) is bounded from below by $N^{-1/3}$ and (1.2.7) is bounded from above by $CN^{-1/3-\varepsilon/6}$ for some $C > 0$. These two bounds are mutually exclusive and therefore there are with high probability no eigenvalues λ with $\lambda \geq N^{-2/3+\varepsilon}$. □

We denote the cumulative distribution functions of the empirical density by n_N and the cumulative distribution function of the semi-circle density by n_{sc}, i.e.

$$n_N(E) := \frac{1}{N}\left|\{\lambda \in \mathrm{Spec}(\mathbf{W}) : \lambda \leq E\}\right|, \quad n_{\mathrm{sc}}(E) := \int_{-\infty}^{E} \rho_{\mathrm{sc}}(x)\mathrm{d}x. \tag{1.2.8}$$

and we prove the following convergence result for n_N.

Lemma 1.2.5 (Convergence of the cumulative distribution function). *For all* $D, \varepsilon > 0$ *we have*

$$\mathbb{P}\left(\sup_{E\in\mathbb{R}} |n_N(E) - n_{\mathrm{sc}}(E)| \geq N^{-1+\varepsilon}\right) \lesssim_{\varepsilon, D} N^{-D}. \tag{1.2.9}$$

Proof. By Lemma 1.2.4 we have $n_N(E) = n_{\mathrm{sc}}(E)$ for all E with $|E| > 3$ with high probability. Therefore it is sufficient to prove Lemma (1.2.5) for $|E| \leq 3$. Let $m^\Delta := \langle \mathbf{G} \rangle - m_{\mathrm{sc}}$ and $\rho^\Delta := \rho_N - \rho_{\mathrm{sc}}$, i.e. ρ^Δ is a signed measure and m^Δ its Stieltjes transform if we extend the definition to also include signed measures. By the averaged local law (1.1.22), we have

$$|m^\Delta| \prec_\gamma \frac{1}{N\eta} \quad \text{for} \quad N^{-1+\gamma} \leq |\eta| \leq 1. \tag{1.2.10}$$

Fix a $\gamma > 0$ and let $0 < \eta \leq \eta_0 =: N^{-1+\gamma}$. From the spectral decomposition of $\mathbf{G}$, it follows that $\eta \operatorname{Im}\langle \mathbf{G}(E + \mathrm{i}\eta)\rangle$ increases monotonously in $\eta > 0$ for all $E \in \mathbb{R}$. Therefore we have

$$\begin{aligned} \eta |\operatorname{Im} m^\Delta(E + \mathrm{i}\eta)| \leq \eta Im\langle \mathbf{G}(E + \mathrm{i}\eta)\rangle + \eta \operatorname{Im} m(E + \mathrm{i}\eta) \\ \leq \eta_0 |m^\Delta(E + \mathrm{i}\eta_0)| + 2\eta_0 |m(E + \mathrm{i}\eta_0)| \prec_\gamma \eta_0, \end{aligned} \tag{1.2.11}$$

where we have also used the monotonicity of $\eta \mapsto \eta \operatorname{Im} m(E + \mathrm{i}\eta)$ in the second step and (1.2.10) as well as $|m| \lesssim 1$ in the last estimate. Dividing (1.2.11) by η and combining it with (1.2.10), we obtain

$$|\operatorname{Im} m^\Delta| \prec_\gamma \frac{N^\gamma}{N\eta} \tag{1.2.12}$$

for all $\gamma > 0$ and $0 < \eta < 1$. It is easy to check that $A \prec_\gamma N^\gamma B$ for all $\gamma > 0$ implies $A \prec B$ and therefore (1.2.12) implies

$$|\operatorname{Im} m^\Delta| \prec \frac{1}{N\eta} \tag{1.2.13}$$

for all $\eta \in (0, 1]$. To prove Lemma 1.2.5, we translate the bounds on m^Δ, Equation (1.2.10) and (1.2.13), to a bound on ρ^Δ. This is done in the following lemma.

Lemma 1.2.6. *Let $E \in \mathbb{R}$, $\eta > N^{-1+\gamma}$ and let $f_{E,\eta} \in C^2(\mathbb{R})$ be a monotonously decreasing function such that $f_{E,\eta}(x) = 1$ for $x \leq E$, $f_{E,\eta}(x) = 0$ for $x \geq E + \eta$ as well as $|f'_{E,\eta}(x)| \lesssim \eta^{-1}$ and $|f''_{E,\eta}(x)| \lesssim \eta^{-2}$. Then*

$$\left| \int_{\mathbb{R}} f_{E,\eta}(\lambda) \rho^\Delta(\mathrm{d}\lambda) \right| \prec_\gamma \eta. \tag{1.2.14}$$

Proof. Let χ be a real and smooth function such that $\chi(x) = 1$ for $|x| \leq \frac{1}{2}$ and $\chi(x) = 0$ for $|x| \geq 1$. For any twice differentiable, compactly supported, real function $f \in C_0^2(\mathbb{R})$ we have the Helffer-Sjöstrand formula

$$f(\lambda) = \frac{1}{\pi} \int_{\mathbb{R}^2} \frac{\mathrm{i}yf''(x)\chi(y) + (\mathrm{i}f(x) - yf'(x))\chi'(y)}{\lambda - x - \mathrm{i}y} \mathrm{d}x\mathrm{d}y. \tag{1.2.15}$$

Let $E_0 < E$ and $\widetilde{f}_{E,\eta} \in C_0^2(\mathbb{R})$ such that $\widetilde{f}_{E,\eta}(x) = f_{E,\eta}(x)$ for $x \geq E_0$, $\widetilde{f}_{E,\eta}(x) = 0$ for $x \leq E_0 - \eta$ as well as $|\widetilde{f}'_{E,\eta}(x)| \lesssim \eta^{-1}$ and $|\widetilde{f}''_{E,\eta}(x)| \lesssim \eta^{-2}$. Integrating $\widetilde{f}_{E,\eta}$ against ρ^Δ and using (1.2.15) we obtain

$$\begin{aligned}\left|\int_{\mathbb{R}} \widetilde{f}_{E,\eta}(\lambda)\rho^\Delta(\mathrm{d}\lambda)\right| &= \left|\operatorname{Re}\int_{\mathbb{R}} \widetilde{f}_{E,\eta}(\lambda)\rho^\Delta(\mathrm{d}\lambda)\right| \\ &\lesssim \left|\int_{\mathbb{R}^2} y\widetilde{f}''_{E,\eta}(x)\chi(y)\operatorname{Im} m^\Delta(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y\right| \\ &\quad + \int_{\mathbb{R}^2} |\widetilde{f}_{E,\eta}(x)||\chi'(y)||\operatorname{Im} m^\Delta(x+\mathrm{i}y)|\mathrm{d}x\mathrm{d}y \\ &\quad + \int_{\mathbb{R}^2} |y||\widetilde{f}'_{E,\eta}(x)||\chi'(y)||\operatorname{Re} m^\Delta(x+\mathrm{i}y)|\mathrm{d}x\mathrm{d}y.\end{aligned} \tag{1.2.16}$$

We separately bound all terms in (1.2.16) and we split the y integral in the first term in the regimes $|y| \leq \eta$ and $|y| \geq \eta$. We only prove the bounds for the $y > 0$ regime. The bounds for $y < 0$ follow by symmetry since $m^\Delta(E+\mathrm{i}\eta) = \bar{m}^\Delta(E-\mathrm{i}\eta)$. For $y \leq \eta$ we find

$$\left|\int_{\mathbb{R}}\int_{0<y\leq\eta} y\widetilde{f}''_{E,\eta}(x)\chi(y)\operatorname{Im} m^\Delta(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y\right| \prec \frac{\eta}{N}\int_{\mathbb{R}} |\widetilde{f}''_{E,\eta}(x)|\mathrm{d}x \lesssim \frac{1}{N} \leq \eta, \tag{1.2.17}$$

where we have used $|y\operatorname{Im} m^\Delta(x+\mathrm{i}y)| \prec \frac{1}{N}$ from (1.2.13) in the first estimate. The second step follows from $|\widetilde{f}''_{E,\eta}(x)| \lesssim \eta^{-2}$ as well as $\mu_\ell(\operatorname{supp}(f''_{E,\eta})) \lesssim \eta$, where μ_ℓ denotes the Lebesgue measure. For $y \geq \eta$ we have by partial integration in x that

$$\begin{aligned}&\int_{\mathbb{R}}\int_{y\geq\eta} y\widetilde{f}''_{E,\eta}(x)\chi(y)\operatorname{Im} m^\Delta(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y \\ &= -\int_{\mathbb{R}}\int_{y\geq\eta} y\widetilde{f}'_{E,\eta}(x)\chi(y)\partial_x \operatorname{Im} m^\Delta(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y.\end{aligned} \tag{1.2.18}$$

By analyticity of m^Δ away from the real axis we have $\partial_x \operatorname{Im} m^\Delta(x+\mathrm{i}y) = -\partial_y \operatorname{Re} m^\Delta(x+\mathrm{i}y)$. After a partial integration in y we arrive at

$$\begin{aligned}&\int_{\mathbb{R}}\int_{y\geq\eta} y\widetilde{f}''_{E,\eta}(x)\chi(y)\operatorname{Im} m^\Delta(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y \\ &= -\int_{\mathbb{R}}\int_{y\geq\eta} \widetilde{f}'_{E,\eta}(x)(y\chi(y))'\operatorname{Re} m^\Delta(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y \\ &\quad - \int_{\mathbb{R}} \widetilde{f}'_{E,\eta}(x)\eta\chi(\eta)\operatorname{Re} m^\Delta(x+\mathrm{i}\eta)\mathrm{d}x.\end{aligned} \tag{1.2.19}$$

Using $|(y\chi(y))'| \lesssim 1$, $|\widetilde{f}'_{E,\eta}(x)| \lesssim \eta^{-1}$ and $\mu_\ell(\operatorname{supp}(f'_{E,\eta})) \lesssim \eta$, as well as (1.2.10) we obtain

$$\int_{\mathbb{R}}\int_{y\geq\eta} |\widetilde{f}'_{E,\eta}(x)||(y\chi(y))'||\operatorname{Re} m^\Delta(x+\mathrm{i}y)|\mathrm{d}x\mathrm{d}y \prec_\gamma \frac{1}{N}|\log(\eta)| \lesssim_\gamma \eta \tag{1.2.20}$$

for the first term on the right-hand side of (1.2.19). The second term (1.2.19) is bounded by

$$\int_{\mathbb{R}} |\widetilde{f}'_{E,\eta}(x)|\eta\chi(\eta)||\operatorname{Re} m^{\Delta}(x+\mathrm{i}\eta)|\mathrm{d}x \prec_{\gamma} \frac{1}{N} \leq \eta \tag{1.2.21}$$

by an analogous argument. Similarly, we find

$$\int_{\mathbb{R}^2} |\widetilde{f}_{E,\eta}(x)||\chi'(y)||\operatorname{Im} m^{\Delta}(x+\mathrm{i}y)|\mathrm{d}x\mathrm{d}y \prec \frac{1}{N} \leq \eta \tag{1.2.22}$$

and

$$\int_{\mathbb{R}^2} |y||\widetilde{f}'_{E,\eta}(x)||\chi'(y)||\operatorname{Re} m^{\Delta}(x+\mathrm{i}y)|\mathrm{d}x\mathrm{d}y \prec \eta \int_{\mathbb{R}} |\widetilde{f}'_{E,\eta}(x)|\mathrm{d}x \lesssim \eta. \tag{1.2.23}$$

for the second and third term on the right-hand side of (1.2.16). Combining (1.2.20), (1.2.21), (1.2.22) and (1.2.23) we find

$$\left|\int_{\mathbb{R}} \widetilde{f}_{E,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda)\right| \prec_{\gamma} \eta \tag{1.2.24}$$

and Lemma 1.2.6 follows since

$$\int_{\mathbb{R}} \widetilde{f}_{E,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda) = \int_{\mathbb{R}} f_{E,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda) \tag{1.2.25}$$

if we choose $E_0 \leq \min \operatorname{supp} \rho^{\Delta}$. □

We have

$$n_N(E) - n_{\mathrm{sc}}(E) = \int_{\mathbb{R}} f_{E,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda) - \int_{E}^{E+\eta} f_{E,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda). \tag{1.2.26}$$

The last term on the right hand side of (1.2.26) satisfies

$$-\int_{E}^{E+\eta} f_{E,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda) \leq \int_{E}^{E+\eta} f_{E,\eta}(\lambda)\mathrm{d}\rho_{\mathrm{sc}}(\lambda) \leq C\eta \tag{1.2.27}$$

for some $C > 0$. Here, we have used the positivity of ρ_N in the first estimate and the boundedness of the density of ρ_{sc} in the second. We combine (1.2.14) with (1.2.27) to achieve a high probability upper bound to (1.2.26) in the form of

$$n_N(E) - n_{\mathrm{sc}}(E) \leq N^{-1+\varepsilon}. \tag{1.2.28}$$

The corresponding lower bound

$$n_N(E) - n_{\mathrm{sc}}(E) \geq -N^{-1+\varepsilon} \tag{1.2.29}$$

is obtained by an analogous argument using

$$n_N(E) - n_{\mathrm{sc}}(E) = \int_{\mathbb{R}} f_{E-\eta,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda) + \int_{E-\eta}^{E} (1 - f_{E-\eta,\eta}(\lambda)\rho^{\Delta}(\mathrm{d}\lambda). \tag{1.2.30}$$

□

Proof of Corollary 1.2.1. We only treat the case $E \in (-2, 0]$. The case $E > 0$ then follows by symmetry. Recall the definition of $k(E)$ in (1.2.1). We have

$$\int_{\lambda_{k(E)}}^{E} \rho_{\mathrm{sc}}(x)\mathrm{d}x = (n_{\mathrm{sc}}(E) - n_N(\lambda_{k(E)})) + (n_N(\lambda_{k(E)}) - n_{\mathrm{sc}}(\lambda_{k(E)})). \quad (1.2.31)$$

We first show that the above integral is small and conclude that $\lambda_{k(E)}$ is close to E with high probability from there. We begin by considering the first summand, we have

$$\begin{aligned} n_N(\lambda_{k(E)}) &= \frac{1}{N}\left|\{\lambda \in \mathrm{Spec}(\mathbf{W}) : \lambda \leq \lambda_{k(E)}\}\right| = \frac{k(E)}{N} \\ &= \frac{1}{N}\left\lceil N \int_{-\infty}^{E} \rho_{\mathrm{sc}}(x)\mathrm{d}x \right\rceil = \frac{1}{N}\left\lceil N n_{\mathrm{sc}}(E)\right\rceil . \end{aligned} \quad (1.2.32)$$

and therefore

$$|n_{\mathrm{sc}}(E) - n_N(\lambda_{k(E)})| = \frac{1}{N}|N n_{\mathrm{sc}}(E) - \lceil N n_{\mathrm{sc}}(E)\rceil\,| \leq \frac{1}{N}. \quad (1.2.33)$$

The second summand on the right hand side of (1.2.31) is bounded by $N^{-1+\varepsilon}$ for all $\varepsilon > 0$ with high probability by Lemma 1.2.5. Combined with (1.2.33) we have

$$\left|\int_{\lambda_{k(E)}}^{E} \rho_{\mathrm{sc}}(x)\mathrm{d}x\right| \prec N^{-1}. \quad (1.2.34)$$

From (1.0.4) it follows that

$$\rho_{\mathrm{sc}}(x) \sim \sqrt{2+x} \quad (1.2.35)$$

for $x \in [-2, 1)$ and therefore

$$\left|\int_{\lambda_{k(E)}}^{E} \rho_{\mathrm{sc}}(x)\mathrm{d}x\right| \sim \left|(E+2)^{\frac{3}{2}} - (\lambda_{k(E)}+2)_+^{\frac{3}{2}}\right|, \quad (1.2.36)$$

where we have used the notation $(x)_+ := \max\{x, 0\}$ for $x \in \mathbb{R}$. First consider $E + 2 \geq N^{-2/3+\varepsilon}$ for a fixed $\varepsilon > 0$. Then we have with high probability

$$\begin{aligned} (E+2)^{\frac{3}{2}} &\sim (E+2)^{\frac{3}{2}} - |(E+2)^{\frac{3}{2}} - (\lambda_{k(E)}+2)_+^{\frac{3}{2}}| \\ &\lesssim (\lambda_{k(E)}+2)_+^{\frac{3}{2}} \lesssim (E+2)^{\frac{3}{2}} + |(E+2)^{\frac{3}{2}} - (\lambda_{k(E)}+2)_+^{\frac{3}{2}}| \sim (E+2)^{\frac{3}{2}}, \end{aligned} \quad (1.2.37)$$

where we have used that with high probability

$$(E+2)^{\frac{3}{2}} \geq N^{-1+\frac{3}{2}\varepsilon} \geq N^{\frac{\varepsilon}{2}}|(E+2)^{\frac{3}{2}} - (\lambda_{k(E)}+2)_+^{\frac{3}{2}}|. \quad (1.2.38)$$

Equation (1.2.37) implies in particular that $(\lambda_{k(E)}+2)_+ = \lambda_{k(E)}+2$ and therefore

$$E + 2 \sim \lambda_{k(E)} + 2 \quad \text{and thus} \quad \rho_{\mathrm{sc}}(E) \sim \rho_{\mathrm{sc}}(\lambda_{k(E)}). \quad (1.2.39)$$

Combining (1.2.34) and (1.2.39) it follows from the mean value theorem that

$$|E - \lambda_{k(E)}| \prec N^{-1}\rho_{\mathrm{sc}}(E)^{-1} \lesssim \frac{1}{N|E-\kappa|}. \tag{1.2.40}$$

Now, let $0 < E + 2 \leq N^{-2/3+\varepsilon}$. Then we have by (1.2.36) that

$$\lambda_{k(E)} + 2 \leq (\lambda_{k(E)} + 2)_+ \lesssim (E+2) + \left| \int_{\lambda_{k(E)}}^{E} \rho_{\mathrm{sc}}(x)\mathrm{d}x \right|^{\frac{2}{3}} \lesssim N^{-\frac{2}{3}+\varepsilon} \tag{1.2.41}$$

again with high probability. This gives the upper bound

$$\lambda_{k(E)} - E \leq N^{-\frac{2}{3}+\varepsilon}. \tag{1.2.42}$$

The lower bound

$$\lambda_{k(E)} - E \geq \lambda_{k(E)} - 2 \geq -N^{-\frac{2}{3}+\varepsilon} \tag{1.2.43}$$

follows directly from Lemma 1.2.4

□

Proof of Corollary 1.2.2. Let $\mathbf{v}_i \in \mathbb{C}^N$, $i \in [\![N]\!]$, be the eigenvectors of $\mathbf{W}$ with eigenvalue λ_i and let $\mathbf{x} \in \mathbb{C}^N$ be a normalized deterministic vector. By (1.1.21) we have with high probability at $z = \lambda_j + \mathrm{i}\eta$ that

$$1 \gtrsim \mathrm{Im}\langle \mathbf{x}, \mathbf{G}(\lambda_j + \mathrm{i}\eta)\mathbf{x}\rangle \sum_{i=1} \frac{\eta}{\eta^2 + (\lambda_j - \lambda_i)^2} |\langle \mathbf{x}, \mathbf{v}_i\rangle|^2 \geq \frac{|\langle \mathbf{x}, \mathbf{v}_j\rangle|^2}{\eta} \tag{1.2.44}$$

for all $j \in [\![N]\!]$, $\eta \geq N^{-1+\varepsilon}$ and $\varepsilon > 0$. Therefore we have with high probability

$$|\langle \mathbf{x}, \mathbf{v}_j\rangle| \lesssim \sqrt{\eta} \leq N^{-1+\frac{\varepsilon}{2}} \tag{1.2.45}$$

and (1.2.3) follows. □

Proof of Theorem 1.0.2. It is straightforward to prove Theorem 1.0.2 as a consequence of eigenvalue rigidity, but it also follows directly from Lemma 1.2.5. Note that we have

$$\left|\mathrm{Spec}(\mathbf{W}) \cap [a, a + N^{-1+\gamma}]\right| = n_N(a + N^{-1+\gamma}) - n_N(a) \tag{1.2.46}$$

and

$$\int_a^{a+N^{-1+\gamma}} \rho_{\mathrm{sc}}(E)\mathrm{d}E = n_{\mathrm{sc}}(a + N^{-1+\gamma}) - n_{\mathrm{sc}}(a), \tag{1.2.47}$$

where the cumulative distribution functions n_N and n_{sc} were introduced in (1.2.8). Taking the difference between the two above equations we find

$$\left| \left|\mathrm{Spec}(\mathbf{W}) \cap [a, a+N^{-1+\gamma}]\right| - \int_a^{a+N^{-1+\gamma}} \rho_{\mathrm{sc}}(E)\mathrm{d}E \right| \leq 2 \sup_{E\in\mathbb{R}} |n_N(E) - N_{\mathrm{sc}}(E)|. \tag{1.2.48}$$

The right-hand side of the above impression is bounded by $N^{-1+\varepsilon}$ with high probability for all $\varepsilon > 0$. Theorem 1.0.2 follows by choosing $\varepsilon = \frac{\gamma}{2}$. □

1.3 Multivariate polynomials of random matrices

A natural continuation to the study of single Wigner matrices is the study of polynomials which have Wigner matrices as their variables. For polynomials q in a single Wigner matrix, this a straight-forward matter, since for a Wigner matrix $\mathbf{W}$ with eigenvalues λ_i, $i \in [\![N]\!]$, the eigenvalues of $q(\mathbf{W})$ are given by $q(\lambda_i)$, $i \in [\![N]\!]$, while the eigenvectors are completely invariant under the polynomial. Hence, analogous results to Theorem 1.1.2, Corollary 1.2.1 and Corollary 1.2.2 can be directly derived from their respective counterpart in the Wigner case, as can the limiting spectral density.

Once we consider polynomials in multiple independent Wigner matrices, however, there is no such shortcut, and studying their spectral statistics becomes a non-trivial endeavour. The strong correlation between the entries of any polynomial of at least degree two makes it impossible to directly apply the resolvent method we have introduced in this chapter. This problem is resolved by introducing linearizations of the polynomial instead. The method of linearizations was first developed in the context of automata theory in [47] and [58] and it was first used in the context of random matrix theory in [43] and [42]. We briefly outline the main idea of the method here. The linearization $\mathbf{L} \in \mathbb{C}^{kN\times kN}$, $k \geq 2$, of a polynomial $q(\mathbf{W}_1, \dots \mathbf{W}_l) \in \mathbb{C}^{N\times N}$ in l independent Wigner matrices is a larger dimensional block matrix that has the individual matrices $\mathbf{W}_1, \dots, \mathbf{W}_l$ as block entries but only in a linear fashion. We define a generalized resolvent $\mathbf{G}$ of $\mathbf{L}$ by

$$\mathbf{G} := (\mathbf{L} - z\mathbf{J})^{-1} \in \mathbb{C}^{kN\times kN}, \tag{1.3.1}$$

where $\mathbf{J} \in \mathbb{C}^{kN\times kN}$ is some projection. The generalized resolvent of $\mathbf{L}$ is connected to the resolvent $\mathbf{g} := (q(\mathbf{W}_1, \dots \mathbf{W}_l) - z)^{-1}$ of $q(\mathbf{W}_1, \dots \mathbf{W}_l)$ by

$$\mathbf{g} = \mathbf{G}_{11} \in \mathbb{C}^{N\times N}, \tag{1.3.2}$$

where the (1,1) entry is understood blockwise. Consider the following example of a polynomial and its linearization. Let

$$q(\mathbf{W}_1, \mathbf{W}_2) = \mathbf{W}_1\mathbf{W}_2 + \mathbf{W}_2\mathbf{W}_1, \tag{1.3.3}$$

i.e. $q(\mathbf{W}_1, \mathbf{W}_2)$ denotes the anti-commutator of two independent Wigner matrices. Also, define the pair $\mathbf{L}$ and $\mathbf{J}$, by

$$\mathbf{L} := \begin{pmatrix} 0 & \mathbf{W}_1 & \mathbf{W}_2 \\ \mathbf{W}_1 & 0 & -\mathbf{I}_N \\ \mathbf{W}_2 & -\mathbf{I}_N & 0 \end{pmatrix} \in \mathbb{C}^{3N\times 3N}, \quad \mathbf{J} := \begin{pmatrix} \mathbf{I}_N & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{C}^{3N\times 3N}. \tag{1.3.4}$$

Using the Schur complement formula to obtain the (1,1) block of $\mathbf{G} = (\mathbf{L} - z\mathbf{J})^{-1} \in \mathbb{C}^{3N\times 3N}$, we find

$$\mathbf{G}_{11} = (\mathbf{W}_1\mathbf{W}_2 + \mathbf{W}_2\mathbf{W}_1 - z)^{-1} = (q(\mathbf{W}_1, \mathbf{W}_2) - z)^{-1} \in \mathbb{C}^{N\times N}. \tag{1.3.5}$$

Due to the much simpler correlation structure of the entries of $\mathbf{L}$, we expect to be able to use the Dyson equation

$$\mathbf{I}_{3N} + (z\mathbf{J} - \mathbb{E}[\mathbf{L}] + \mathcal{S}[\mathbf{M}])\mathbf{M} = 0 \quad \text{with} \quad \mathcal{S}[\mathbf{R}] = \mathbb{E}[(\mathbf{L} - \mathbb{E}[\mathbf{L}])\mathbf{R}(\mathbf{L} - \mathbb{E}[\mathbf{L}])]. \tag{1.3.6}$$

to obtain a good deterministic approximation $\mathbf{M} \in \mathbb{C}^{3N\times 3N}$ for $\mathbf{G}$ and to prove a local law in the form of Theorem 1.1.2 for $\mathbf{G} - \mathbf{M}$. If we are able to obtain a local law for $\mathbf{G} - \mathbf{M}$, then a local law for $\mathbf{g} - \mathbf{M}_{11} = \mathbf{G}_{11} - \mathbf{M}_{11}$ follows immediately as $\mathbf{G}_{11}$ and $\mathbf{M}_{11}$ are submatrices of $\mathbf{G}$ and $\mathbf{M}$ respectively.

In the preprint that is reproduced in the next chapter, we investigate precisely this problem, not only for the anti-commutator but for all self-adjoint quadratic polynomials. We analyse the limiting spectrum close to the edge and characterize its behaviour. We show that, with very few specific exceptions, the polynomials only have regular edges, i.e. edges with a square root growth of their density. This result is presented in Proposition 2.2.7 and Proposition 2.2.8. We also prove an optimal local law around any regular edge. This result is given in Theorem 2.2.9. These two results are then combined to prove a norm convergence rate with optimal exponent in Theorem 2.2.2.

Remark. *In the preprint we do not give a proof of the eigenvalue rigidity statement, Corollary 2.2.12, for reasons of conciseness. However, in addition to the local law Theorem 1.1.2, the proof of eigenvalue rigidity for Wigner matrices, Corollary 1.2.1, used only the square root growth of the limiting density close to its edges. Therefore, the proof of Corollary 2.2.12 is completely analogous to the proof given for Corollary 1.2.1 in Section 1.2.*

Chapter 2

Norm Convergence Rate for Multivariate Quadratic Polynomials of Wigner Matrices

Up to minor revisions, this chapter is a reproduction of the preprint arXiv:2308.16778: [35]. It has been co-authored by the author of this thesis, Torben Krüger and Yuriy Nemish.

arXiv:2308.16778:

Norm Convergence Rate for Multivariate Quadratic Polynomials of Wigner Matrices

Jacob Fronk[1]*, Torben Krüger*[1]*, Yuriy Nemish*

Abstract

We study Hermitian non-commutative quadratic polynomials of multiple independent Wigner matrices. We prove that, with the exception of some specific reducible cases, the limiting spectral density of the polynomials always has a square root growth at its edges and prove an optimal local law around these edges. Combining these two results, we establish that, as the dimension N of the matrices grows to infinity, the operator norm of such polynomials q converges to a deterministic limit with a rate of convergence of $N^{-2/3+o(1)}$. Here, the exponent in the rate of convergence is optimal. For the specific reducible cases, we also provide a classification of all possible edge behaviours.

Keywords: polynomials of random matrices, local laws, Dyson equation, extreme eigenvalues
AMS Subject Classification: 60B20, 15B52

2.1 Introduction

The empirical spectral distribution of a random matrix is typically well approximated by a deterministic measure as its dimension grows to infinity. A clear contender for the most famous example of such a convergence is the celebrated semi-circle law. It states that the spectral measure of a Wigner matrix, a Hermitian $N \times N$-matrix $\mathbf{X}$ with centered i.i.d. entries x_{ij} above the diagonal and $\mathbb{E}|x_{ij}|^2 = N^{-1}$, converges to the semi-circle distribution, supported on the interval $[-2, 2]$ [68]. In particular, the largest and smallest eigenvalues of $\mathbf{X}$ converge to the respective edges of the support, implying the convergence $\|\mathbf{X}\| \to 2$ of the operator norm, provided the fourth moments of the entries of $\sqrt{N}\mathbf{X}$ are finite [11, 10].

For non-commutative Hermitian polynomials $\mathbf{Q} = q(\mathbf{X}_1, \ldots, \mathbf{X}_l)$ in several independent Wigner matrices $\mathbf{X}_i$ an analogous statement holds. In this setup the limit

[1] Partially supported by VILLUM FONDEN research grant no. 29369
E-mail addresses: jf@math.ku.dk (J. Fronk), torben.krueger@fau.de (T. Krüger), ynemish@ucsd.edu (Yu. Nemish)

of the eigenvalue distribution equals the distribution associated to the polynomial $\mathfrak{q} = q(\mathfrak{s}_1, \ldots, \mathfrak{s}_l)$, where the matrices are replaced by free semicircular random variables within a non-commutative probability space, i.e. $\mathfrak{s}_i$ can be interpreted as operators acting on an infinite dimensional Hilbert space. This result was first established for Gaussian random matrices in [67] and extended to Wigner matrices in [25]. Similarly, the convergence of the norms $\|\mathbf{Q}\| \to \|\mathfrak{q}\|$ was first shown by Haagerup-Thorbjørnsen [43] in the Gaussian case and the Wigner case was proven by Anderson in [4]. Such results have also been shown when some $\mathbf{X}_i$ are replaced by non-random matrices [15, 53]. For non-Hermitian polynomials convergence of the spectral measure to the limiting Brown measure, predicted by free probability theory, is known only for very specific cases, e.g. for products of random matrices [41, 55] and for quadratic polynomials [24].

Determining the limiting spectral measure ρ of Hermitian polynomials $\mathbf{Q}$, or equivalently the distribution of $\mathfrak{q}$, becomes a nontrivial task beyond particular computable cases. Several works have been devoted to the analysis of the regularity properties of ρ. It has been shown that ρ has a single interval support [42] and does not contain any atoms [52, 59]. The cumulative distribution function of ρ is Hölder-continuous [12], and if q is a monomial or a homogeneous quadratic polynomial then ρ is even absolutely continuous [22, 28].

In the present paper we consider the case when q is a general polynomial of degree two, i.e. we study self-adjoint polynomials in l independent Wigner matrices $\mathbf{X}_1, \ldots, \mathbf{X}_l$ of the form

$$q(\mathbf{X}_1, \ldots, \mathbf{X}_l) = \sum_{i,j=1}^{l} \mathbf{X}_i A_{ij} \mathbf{X}_j + \sum_{i=1}^{l} b_i \mathbf{X}_i + c, \tag{2.1.1}$$

where $0 \neq A = (A_{ij}) \in \mathbb{C}^{l \times l}$ is a Hermitian matrix, $b = (b_i) \in \mathbb{R}^l$ and $c \in \mathbb{R}$. For these polynomials we classify the edge behaviour of ρ and show (see Proposition 2.2.7 below) that at both edges the limiting spectral measure is absolutely continuous and apart from specific reducible cases its density exhibits a square root growth. The reducible cases are, up to a shift and change in sign, of the form $\mathbf{Y}^*\mathbf{Y}$, where $\mathbf{Y}$ is an affine combination of the underlying Wigner matrices. Such polynomials still have a square root edge at the rightmost point of the spectrum, but have a density blow-up at the leftmost point if it is equal to zero. All these cases are classified in Proposition 2.2.8 below.

The square root growth of the limiting spectral distribution is a well-known phenomenon, that is already present in the semicircle law for Wigner matrices $\mathbf{X}$. In this setup the rate of convergence for the norm of $\mathbf{X}$ is $\|\mathbf{X}\| = 2 + \mathcal{O}(N^{-2/3+o(1)})$ with very high probability [27] and several tail estimates have been established [7, 8, 30]. In fact, for Wigner matrices the distribution of the largest eigenvalue is known to be universal and given by the Tracy-Widom law [32, 60, 62]. This distribution was first identified by Tracy and Widom for the Gaussian ensembles [65, 66] and necessary and sufficient conditions for its universality in the context of Wigner matrices identified in [51].

Edge universality has been extended to many other Hermitian random matrix models, including invariant ensembles [19], covariance matrices [34, 57], deformed Wigner matrices [50], deterministic matrices with Gaussian perturbations [49] and models with correlations [3, 56]. Such universality results often rely on control of the eigenvalue location on mesoscopic scales between $O(1)$ and $O(N^{-1})$, i.e. on a local law.

Local laws arose in the context of Wigner matrices [27, 64] and have subsequently been extended to the more complex models listed above. For non-commuting polynomials of several random matrices local laws are known only in specific cases, starting with Anderson's work on the anti-commutator $\mathbf{X}_1\mathbf{X}_2 + \mathbf{X}_2\mathbf{X}_1$ of Wigner matrices [5] that controls the deviation of bulk eigenvalues from their expected position on the scale $\mathcal{O}(N^{-1/2})$. For general polynomials, that satisfy certain checkable conditions, an optimal local law in the bulk regime was proved in [28]. Related results are [13] and [14], where it was shown that for two random matrices satisfying a local law and whose eigenvectors are in generic direction to each other also their sum satisfies a local law at the edge and in the bulk. This result covers e.g. $\mathbf{X}_1^2 + \mathbf{X}_2^3$ if one of these matrices is a Gaussian unitary ensemble. For non-Hermitian polynomials the results [54] and [40] cover products of independent matrices with i.i.d. entries. and [24] quadratic polynomials.

Currently the best estimate on the convergence rate of the norm for general polynomials of GUE matrices is $-N^{-\varepsilon} \leq \|\mathbf{Q}\| - \|\mathfrak{q}\| \leq CN^{-1/4}$, for some $\varepsilon < \frac{1}{3}$ and $C > 0$ established in [23]. Our work improves this bound for polynomials $\mathbf{Q} = q(\mathbf{X}_1, \ldots, \mathbf{X}_l)$ of the form (2.1.1) to the optimal rate of $N^{-2/3+o(1)}$ with square root growth at the edges of the spectral density and extends the result to Wigner matrices. The main novelty here is a detailed analysis of the Dyson equation, describing a generalized resolvent of the linearization matrix associated with the polynomial in the limit $N \to \infty$ and, consequently, the resolvent of $\mathbf{Q}$ itself. The idea of linearising polynomials of random matrices in this way stems from [42, 43] and has been used in many works since, in particular in [4, 6, 16, 44, 45]. In particular, we perform a comprehensive stability analysis of the Dyson equation that allows us (i) to prove a square root growth of the limiting spectral density ρ and (ii) to establish an optimal bound on the difference between the solution to the Dyson equation and the generalized resolvent by using a modification of the bound on the random error matrix in the Dyson equation from [29]. The main insight is that the matrix Dyson equation for the linearization, which has a linear self-energy term, can be reduced to a scalar equation for a function $m = m(z)$ of the form

$$-\frac{1}{m} = z + \gamma(m), \tag{2.1.2}$$

where the self-energy term $\gamma(m)$ is now a non-linear function of m. This representation allows us to identify the values of the spectral edges in terms of the coefficients of q and study the quadratic singularity at these edge points.

Notations

In this section, we introduce some definitions commonly used throughout the paper. The standard scalar product of vectors $v, w \in \mathbb{C}^n$ will be denoted by $\langle v, w\rangle$ and the standard euclidean norm by $\|v\| = \sqrt{\langle v, v\rangle}$. A vector v is called normalized if $\|v\| = 1$.

Matrices $R \in \mathbb{C}^{k\times n}$ for (fixed) $k, n \in \mathbb{N}$ are usually denoted by non-boldfaced roman letters and matrices $\mathbf{R} \in \mathbb{C}^{kN\times nN}$ are usually denoted by boldfaced roman letters. In particular, we denote the identity matrix on $\mathbb{C}^{k\times k}$ by I_k and the identity matrix on $\mathbb{C}^{kN\times kN}$ by $\mathbf{I}_{kN}$. For any $k, n \in \mathbb{N}$ we embed $\mathbb{C}^{k\times n}$ in $\mathbb{C}^{k\times n} \otimes \mathbb{C}^{N\times N}$ by identifying $R \in \mathbb{C}^{k\times n}$ with $R \otimes \mathbf{I}_N \in \mathbb{C}^{k\times n} \otimes \mathbb{C}^{N\times N} \cong \mathbb{C}^{kN\times nN}$ and write compactly

$$R = R \otimes \mathbf{I}_N \in \mathbb{C}^{kN\times nN}. \tag{2.1.3}$$

Matrices $R \in \mathbb{C}^{k\times n}$, which are embedded into $\mathbb{C}^{kN\times nN}$, get still denoted by non-boldfaced letters.

For $R, T \in \mathbb{C}^{n\times n}$ we denote the normalized trace by $\langle R\rangle = \frac{1}{n}\operatorname{Tr} R$ and define a scalar product by $\langle R, T\rangle := \langle R^*T\rangle$. We use the standard operator norm and the Hilbert-Schmidt norm, which are given by

$$\|R\| = \sup_{\|x\|\le 1} \|Rx\| \quad \text{and} \quad \|R\|_{\mathrm{hs}} = \sqrt{\langle R^*R\rangle}. \tag{2.1.4}$$

For vectors of matrices $V = (V_i)_{i\in[\![l]\!]} \in (\mathbb{C}^{n\times n})^l$ we denote by $\|V\|$ the maximum of the operator norms of the entries, i.e.

$$\|V\| = \max_{i\in[\![l]\!]} \|V_i\|. \tag{2.1.5}$$

For random matrices $\mathbf{R} \in \mathbb{C}^{kN\times kN}$ the isotropic and averaged p-norms are defined by

$$\|\mathbf{R}\|_p := \sup_{\|\mathbf{x}\|,\|\mathbf{y}\|\le 1} (\mathbb{E}|\langle \mathbf{x}, \mathbf{R}\mathbf{y}\rangle|^p)^{\frac{1}{p}} \quad \text{and} \quad \|\mathbf{R}\|_p^{\mathrm{av}} := \sup_{\|\mathbf{B}\|\le 1} (\mathbb{E}|\langle \mathbf{B}\mathbf{R}\rangle|^p)^{\frac{1}{p}}. \tag{2.1.6}$$

For a block matrix $\mathbf{R} \in \mathbb{C}^{kN\times nN}$ with blocks $\mathbf{R}_{ij} \in \mathbb{C}^{N\times N}$, $i \in [\![k]\!]$ and $j \in [\![n]\!]$, we define the blockwise (averaged) trace $\underline{\mathbf{R}} \in \mathbb{C}^{k\times n}$ by

$$\underline{\mathbf{R}}_{ij} = \langle \mathbf{R}_{ij}\rangle. \tag{2.1.7}$$

A matrix R is said to be positive definite if $\langle v, Rv\rangle > 0$ for all $v \in \mathbb{C}^{n\times n}\setminus\{0\}$ and we write $R > 0$. It is called positive semi-definite if $\langle v, Rv\rangle \ge 0$ for all $v \in \mathbb{C}^{n\times n} \setminus \{0\}$ and we write $R \ge 0$. R is called negative (semi-)definite, denoted by $R < 0$ ($R \le 0$), if $-R$ is positive (semi)-definite. For $S \in \mathbb{C}^{n\times n}$ Hermitian we write $S \ge R$ if $S - R \ge 0$ and $S > R$ if $S - R > 0$.

For linear operators acting on matrix spaces, we denote by $\|\cdot\|_{\mathrm{sp}}$ the norm induced by the Hilbert-Schmidt norm, $\|\cdot\|_{\mathrm{hs}}$. The identity map between matrix spaces is denoted by $\mathbb{1}$, i.e. $\mathbb{1}[R] = R$ for all $R \in \mathbb{C}^{k\times n}$.

The upper half-plane will be denoted by $\mathbb{H}$, i.e.

$$\mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im} z > 0\} \tag{2.1.8}$$

and $[\![n]\!] := \{1, 2, \ldots, n\}$ is used for the natural numbers up to n. If X and Y are positive quantities, we use the notation $X \lesssim Y$ if there is some constant $c > 0$ such that $X \leq cY$. The constant will in general depend on the coefficients of q. If it also depends on some other parameters α, we write $X \lesssim_\alpha Y$. In particular, the constant will never depend on the dimension of our random matrices, N. If both $X \lesssim Y$ and $Y \lesssim X$ hold true, we write $X \sim Y$.

Acknowledgements

We thank Peter Henning Thomsen Krarup for his valuable contributions in the early stages of the project.

2.2 Main results

Assumption. *Let ζ_0 be a real-valued and ζ_1 be a complex-valued random variable and let ζ_0 and ζ_1 be independent. For $i = 0, 1$, they are to satisfy*

$$\mathbb{E}[\zeta_i] = 0, \quad \mathbb{E}[|\zeta_i|^2] = 1 \quad \textit{and} \quad \mathbb{E}[|\zeta_i|^p] \leq C_p \tag{2.2.1}$$

for all $p \in \mathbb{N}$ and some constants $C_p > 0$, depending on p. Let $\mathbf{W} \in \mathbb{C}^{N\times N}$ be a Hermitian random matrix characterized by its entry distribution:

1. *The diagonal entries $\{w_{ii} : i \in [\![N]\!]\}$ and off-diagonal entries $\{(w_{ij}, w_{ji}) : i, j \in [\![N]\!],\, i < j\}$ are independent;*
2. *$\{w_{ii} : i \in [\![N]\!]\}$ consists of independent copies of $\frac{1}{\sqrt{N}}\zeta_0$,*
3. *$\{(w_{ij}, w_{ji}) : i, j \in [\![N]\!],\, i < j\}$ consists of independent copies of $\frac{1}{\sqrt{N}}(\zeta_1, \bar{\zeta}_1)$.*

For a fixed $l \in \mathbb{N}$ we define $\mathbf{X} = (\mathbf{X}_i)_{i\in[\![l]\!]} \in (\mathbb{C}^{N\times N})^l$, a vector of random matrices, where each $\mathbf{X}_i$, $i \in [\![l]\!]$, is an independent copy of $\mathbf{W}$.

To present our results we first need to distinguish between shifted reducible and non-reducible quadratic polynomials.

Definition 2.2.1 (Shifted reducible and non-reducible second degree polynomial)**.** *We call any non-commutative quadratic polynomial of the matrices $\mathbf{X} = (\mathbf{X}_i)_{i\in[\![l]\!]}$ which is of the form*

$$q_r(\mathbf{X}) = \alpha(v^*\mathbf{X} - \xi)(v^*\mathbf{X} - \xi)^* - \beta \tag{2.2.2}$$

*for some $\alpha, \beta, \xi \in \mathbb{R}$, $\alpha \neq 0$, $\xi \geq 0$ and $v \in \mathbb{C}^l$ with $\|v\| = 1$ a **shifted reducible quadratic polynomial**. Any polynomial not of this form is called a **non-reducible quadratic polynomial**.*

Remark. *Shifted reducible quadratic polynomials are exactly the polynomials of the form (2.1.1) with coefficients $A = \alpha vv^*$, $b = -\alpha\xi(v + \bar{v})$ and $c = \alpha|\xi|^2 - \beta$ for some $\alpha, \beta, \xi \in \mathbb{R}$ with $\alpha \neq 0$, $\xi \geq 0$ and normalised $v \in \mathbb{C}^l$. Note that our definition of a shifted reducible polynomial also allows for polynomials of the form (2.2.2) with $\xi \in \mathbb{C}$ since $(v^*\mathbf{X} - \xi)(v^*\mathbf{X} - \xi)^* = (e^{\mathrm{i}\varphi}v^*\mathbf{X} - e^{\mathrm{i}\varphi}\xi)(e^{\mathrm{i}\varphi}v^*\mathbf{X} - e^{\mathrm{i}\varphi}\xi)^*$ for all $\varphi \in \mathbb{R}$ and we used this invariance to restrict to $\xi \in \mathbb{R}_{\geq 0}$ w.l.o.g..*

Shifted reducible polynomials are those where the edge characteristics reduce to understanding the singular value statistics of some (not necessarily Hermitian) first order polynomial in $\mathbf{X}$, whereas non-reducible polynomials are those where such a simplification is not possible. The main focus of this work are non-reducible polynomials, but we also characterize the limiting spectral measure of the shifted reducible polynomials.

For non-reducible polynomials, we prove the convergence of the norm of $q(\mathbf{X})$ to a deterministic τ_* in the following sense:

Theorem 2.2.2 (Convergence of the matrix norm)**.** *Let q be a non-reducible quadratic polynomial of the form (2.1.1). There is a deterministic $\tau_* > 0$, only depending on the coefficients A, b, c of q, such that for all $\varepsilon, D > 0$ the operator norm of $q(\mathbf{X})$ satisfies the estimate*

$$\mathbb{P}\left(\left|\|q(\mathbf{X})\| - \tau_*\right| \geq N^{-\frac{2}{3}+\varepsilon}\right) \leq C_{\varepsilon,D}N^{-D}.$$

Remark. *The deterministic value τ_* in Theorem 2.2.2 is the value of the norm as predicted by the limiting spectral measure*

$$\rho := \lim_{N\to\infty} \frac{1}{N} \sum_{\mu\in\mathrm{Spec}(q(\mathbf{X}))} \delta_\mu, \tag{2.2.3}$$

where δ_μ denotes the Dirac measure at point μ and the sum runs over all eigenvalues of $q(\mathbf{X})$ accounting also for multiplicity. That is, we have $\tau_* = \max\{|\tau_+|, |\tau_-|\}$, where $\mathrm{supp}(\rho) = [\tau_-, \tau_+]$ (see Definition 2.2.5 below). The points τ_+ and τ_- can be obtained by solving an explicit polynomial equation, for details see Lemma 2.3.5 below.

To obtain the main theorem we need several intermediate results. They establish that the eigenvalue density of any non-reducible polynomial $q(\mathbf{X})$ approximately shows a square root behaviour around its edge. For reducible polynomials, we classify the different edge behaviours.

The central object of our interest is the Stieltjes transform of the limiting spectral measure ρ, which we denote by m. The function m is uniquely defined by the following proposition.

Proposition 2.2.3 (Existence and uniqueness of the Stieltjes transform)**.** *Let $A \in \mathbb{C}^{l\times l}$, $A = A^*$, $A \neq 0$, $b \in \mathbb{R}^l$ and $c \in \mathbb{R}$. There is a unique function $m : \mathbb{H} \to \mathbb{H}$ such that*

1. *m is complex analytic on all of $\mathbb{H}$;*
2. *$\lim_{z\to\infty} zm(z) = -1$;*
3. *For all $z \in \mathbb{H}$ the equation*

$$-m^{-1} = z + \gamma(m), \tag{2.2.4}$$

with

$$\gamma(m) := -\operatorname{Tr} A(I_l + Am)^{-1} + mb^t \left((I_l + 2m\widehat{A})^{-2}(I_l + m\widehat{A})\right) b - c, \tag{2.2.5}$$

is satisfied for $m = m(z)$. Here, the notation $\widehat{A} = \frac{1}{2}(A + A^t)$ was used.

The proof of Proposition 2.2.3 is deferred to Appendix 2.6.3.

Definition 2.2.4 (Self-consistent density of states)**.** *By Conditions 1 and 2 the function m in Proposition 2.2.3 is the Stieltjes transform of a unique probability measure ρ on the real line, i.e.*

$$m(z) = \int_{\mathbb{R}} \frac{\rho(\mathrm{d}x)}{x - z} \tag{2.2.6}$$

for all $z \in \mathbb{H}$. We call ρ the self-consistent density of states corresponding to $q(\mathbf{X})$.

Remark. *By the global law, [28, Proposition 2.17], the self-consistent density of states is indeed the limiting spectral measure.*

Note that since A is Hermitian, the matrix $\widehat{A} = \frac{1}{2}(A + A^t)$ denotes the entrywise real part of A. This should not be confused with the algebraic definition of the real part of a matrix, $\operatorname{Re} R = \frac{1}{2}(R + R^*)$. Let $I \subset \mathbb{R}$ be an interval and $E \in I$. Due to the Stieltjes inversion formula, the limiting spectral measure and its Stieltjes transform are related by the equation

$$\rho(E) = \lim_{\eta\searrow 0} \frac{1}{\pi} \operatorname{Im} m(E + \mathrm{i}\eta), \tag{2.2.7}$$

whenever that limit exists for all $E \in I$ (see e.g. [36, Equation (1.4)]). It was shown in [59, Theorem 1.1 (3)] that $\operatorname{supp}(\rho)$ is a single compact interval on the real line. In particular, this means that ρ has no internal edges. We use the following definition.

Definition 2.2.5 (Edges of the limiting spectral measure)**.** *Let τ_+ denote the position of the right edge of ρ and let τ_- denote the position of the left edge of ρ, i.e., we have*

$$\operatorname{supp}(\rho) = [\tau_-, \tau_+] \tag{2.2.8}$$

We also introduce the notion of a regular edge.

Definition 2.2.6. *Let $\tau_0 \in \{\tau_-, \tau_+\}$. The limiting spectral measure ρ is said to have a regular edge at τ_0 if $\rho(\mathrm{d}E) = \rho(E)\mathrm{d}E$ has a Lebesgue density (also denoted by ρ) in a neighborhood of τ_0 and*

$$\lim_{\substack{E\in\mathrm{supp}(\rho)\\ E\to\tau_0}} \frac{\rho(E)}{\sqrt{|\tau_0 - E|}} \tag{2.2.9}$$

exists and does not equal 0.

In other words, regular edges are those that show a square root decay of the density ρ. Our next two results concern the edge characterization of reducible and non-reducible polynomials.

Proposition 2.2.7 (Edges of non-reducible polynomials)**.** *Let q be a non-reducible quadratic polynomial and ρ its associated limiting spectral measure defined in (2.2.3). Then the measure ρ has regular edges both at τ_- and at τ_+.*

Next, we consider shifted reducible polynomials of the form

$$q_r(\mathbf{X}) = (v^*\mathbf{X} - \xi)(v^*\mathbf{X} - \xi)^* \tag{2.2.10}$$

for some $\xi \in \mathbb{R}_{\geq 0}$ and normalized $v \in \mathbb{C}^l$.

Remark. *Compared to the general case defined in (2.2.2) we restrict here to $\alpha = 1$ and $\beta = 0$. As these constants only constitute a scaling and a shift respectively, the following result, Proposition 2.2.8, generalizes in a straightforward manner to all $\alpha \neq 0$ and $\beta \in \mathbb{R}$. For $\alpha < 0$ the roles of the left and the right edge are reversed.*

We introduce the quantities $\sigma = \|\operatorname{Re} v\|$, $\mu = \langle \operatorname{Re} v, \operatorname{Im} v\rangle$ and for $\sigma \in (0,1)$ we define $s > 0$ with

$$s^2 := \begin{cases} \big((\sigma^2 + a_+^2(1-\sigma^2) + 2a_+\mu)(\sigma^2 + a_+\mu)\big)^{-1} & \\ \qquad + \big((\sigma^2 + a_-^2(1-\sigma^2) + 2a_-\mu)(\sigma^2 + a_-\mu)\big)^{-1} & \text{if } \mu \neq 0, \\ \sigma^{-4} & \text{if } \mu = 0, \end{cases} \tag{2.2.11}$$

where for $\mu \neq 0$ we also set the constant $a_\pm \in \mathbb{R}$ to

$$a_\pm = \frac{1-2\sigma^2}{2\mu} \pm \sqrt{\left(\frac{1-2\sigma^2}{2\mu}\right)^2 + 1}. \tag{2.2.12}$$

The edges of the shifted reducible polynomial are then characterized as follows.

Proposition 2.2.8 (Edges of shifted reducible polynomials)**.** *Let q_r be a shifted reducible polynomial as in (2.2.10) for some $\xi \in \mathbb{R}_{\geq 0}$ and normalized $v \in \mathbb{C}^l$. Then the limiting spectral measure ρ has a regular edge at $\tau_+ = \max \operatorname{supp}(\rho)$. Furthermore there is a $\kappa > 0$ such that the behaviour of ρ on $(\tau_-, \tau_- + \kappa)$ with $\tau_- = \min \operatorname{supp}(\rho)$ is given by*

$$\rho(E) \sim \begin{cases} (E-\tau_-)^{-\frac{1}{2}} & \text{if } \xi = 0 \\ (E-\tau_-)^{-\frac{1}{2}} & \text{if } v \in \mathbb{R}^l \text{ and } \xi < 2 \\ (E-\tau_-)^{-\frac{1}{4}} & \text{if } v \in \mathbb{R}^l \text{ and } \xi = 2 \\ (E-\tau_-)^{-\frac{1}{2}} & \text{if } v \in \mathbb{C}^l \setminus e^{\mathrm{i}\varphi}\mathbb{R}^l \text{ for all } \varphi \in (-\pi, \pi] \text{ and } s\xi < 2 \\ (E-\tau_-)^{-\frac{1}{3}} & \text{if } v \in \mathbb{C}^l \setminus e^{\mathrm{i}\varphi}\mathbb{R}^l \text{ for all } \varphi \in (-\pi, \pi] \text{ and } s\xi = 2. \end{cases} \tag{2.2.13}$$

In all other cases the left edge is also regular, i.e. $\rho(E) \sim (E-\tau_-)^{1/2}$.

Remark. *The reason behind this result is that the edge behaviour near the left edge follows from the distribution of small singular values of $v^*\mathbf{X} - \xi$. It follows that ρ has singularities precisely if ξ is in the asymptotic spectrum of $v^*\mathbf{X}$, with stronger singularities being observed if ξ is in the bulk of the spectrum and weaker ones if ξ is on the edge of the spectrum.*

Secondly we prove that $m\mathbf{I}_N$, where $\mathbf{I}_N$ is the identity matrix on $\mathbb{C}^{N\times N}$, well approximates the resolvent $\mathbf{g} = (q(\mathbf{X}) - z\mathbf{I}_N)^{-1} \in \mathbb{C}^{N\times N}$ of $q(\mathbf{X})$ around any regular edge and away from the spectrum. More precisely we prove uniform convergence of the quadratic form of $\mathbf{g}$ on the set

$$\mathbb{D}_\gamma^{\kappa_0} := \{E + \mathrm{i}\eta \in \mathbb{H} : |E - \tau_0| \leq \kappa_0,\ N^{-1+\gamma} \leq \eta \leq 1\} \tag{2.2.14}$$

for some $\kappa_0 \sim 1$ and for all $\gamma > 0$, as well as on the set

$$\mathbb{G}_\gamma^{C,\eta_0} := \{E + \mathrm{i}\eta \in \mathbb{H} : C^{-1} \leq \operatorname{dist}(E, \operatorname{supp}(\rho)) \leq C,\ N^{-1+\gamma} \leq \eta \leq \eta_0\} \tag{2.2.15}$$

for all $C, \gamma > 0$ and some η_0 depending on C.

Theorem 2.2.9 (Local law for regular edges of polynomials)**.** *Let q be a polynomial of the form (2.1.1) and let ρ have a regular edge at $\tau_0 \in \{\tau_-, \tau_+\}$. There is a $\kappa_0 > 0$, depending only on the coefficients of q, such that for all $\varepsilon, \gamma, D > 0$ and $z = E + \mathrm{i}\eta \in \mathbb{D}_\gamma^{\kappa_0}$ the isotropic local law*

$$\mathbb{P}\left(|\langle \mathbf{x}, (\mathbf{g} - m)\mathbf{y}\rangle| > \|\mathbf{x}\|\|\mathbf{y}\| N^\varepsilon \left(\frac{\operatorname{Im} m}{\sqrt{N\eta}} + \frac{1}{N\eta}\right)\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{2.2.16}$$

holds for all deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$. Moreover, the averaged local law

$$\mathbb{P}\left(\left|\frac{1}{N}\operatorname{Tr}(\mathbf{B}(\mathbf{g} - m))\right| > \|\mathbf{B}\| \frac{N^\varepsilon}{N\eta}\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{2.2.17}$$

holds for all deterministic $\mathbf{B} \in \mathbb{C}^{N\times N}$. *If additionally* $E \notin \operatorname{supp}(\rho)$, *an improved averaged local law of the form*

$$\mathbb{P}\left(\left|\frac{1}{N}\operatorname{Tr}(\mathbf{B}(\mathbf{g}-m))\right| > \|\mathbf{B}\|\left(\frac{N^{\varepsilon}}{N|z-\tau_0|}+\frac{N^{\varepsilon}}{(N\eta)^2\sqrt{|z-\tau_0|}}\right)\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{2.2.18}$$

holds.

Away from the spectrum, we will also prove the following form of an averaged local law

Proposition 2.2.10 (Local law away from the spectrum)**.** *Let* q *be a polynomial of the form* (2.1.1). *For all* $C > 0$ *there is an* $\eta_0 > 0$, *depending only on the coefficients of* q, *such that an averaged local law of the form*

$$\mathbb{P}\left(\left|\frac{1}{N}\operatorname{Tr}(\mathbf{B}(\mathbf{g}-m))\right| > \|\mathbf{B}\|N^{\varepsilon}\left(\frac{1}{N}+\frac{1}{(N\eta)^2}\right)\right) \lesssim_{\varepsilon,\gamma,D,C} N^{-D} \tag{2.2.19}$$

holds true for all deterministic $\mathbf{B} \in \mathbb{C}^{N\times N}$, $\gamma, \varepsilon, D > 0$ *and* $z \in \mathbb{G}_{\gamma}^{C,\eta_0}$.

The local laws, Theorem 2.2.9 and Proposition 2.2.10, are proved in Section 2.5. They give us very precise control over the spectral properties of $q(\mathbf{X})$ close to any regular edge as can be seen from the following corollaries,

Corollary 2.2.11 (Edge eigenvector delocalization)**.** *Let the assumptions of Theorem 2.2.9 be satisfied and* $\mathbf{v}$ *be a normalized eigenvector of* $q(\mathbf{X})$ *with eigenvalue* λ. *Then there is a* $\kappa_0 > 0$, *only depending on the coefficients of* q, *such that if* $|\lambda - \tau_0| \le \kappa_0$ *then we have for all* $\varepsilon, D > 0$ *that*

$$\sup_{\mathbf{x}\in\mathbb{C}^{N\times N},\|\mathbf{x}\|=1} \mathbb{P}\left(|\langle \mathbf{x}, \mathbf{v}\rangle| \ge N^{-\frac{1}{2}+\varepsilon}\right) \lesssim_{D,\varepsilon} N^{-D} \tag{2.2.20}$$

Corollary 2.2.12 (Eigenvalue rigidity)**.** *Denote the classical index of the eigenvalue close to energy* $E \in \operatorname{supp}(\rho)$ *by*

$$k(E) := \left\lceil N\int_{-\infty}^{E}\rho(E')\mathrm{d}E'\right\rceil, \tag{2.2.21}$$

with $\lceil\cdot\rceil$ *being the ceiling function. Let* ρ *have a regular edge at* τ_0. *All eigenvalues around* τ_0 *are close to their classical position in the following sense. There is a* $\kappa_0 > 0$, *only depending on the coefficients of* q, *such that*

$$\mathbb{P}\left(\sup_E|\lambda_{k(E)} - E| \ge \min\left\{\frac{N^{\varepsilon}}{N|E-\tau_0|}, \frac{N^{\varepsilon}}{N^{\frac{2}{3}}}\right\}\right) \lesssim_{D,\varepsilon} N^{-D}. \tag{2.2.22}$$

holds for all $\varepsilon, D > 0$ *and* $E \in \operatorname{supp}(\rho)$ *with* $|E - \tau_0| \le \kappa_0$.

Proof of Theorem 2.2.2. As the norm of any Hermitian matrix H with non-decreasing eigenvalues λ_i, $i \in [\![n]\!]$, is given by $\|H\| = \max\{|\lambda_1|, |\lambda_n|\}$, Theorem 2.2.2 follows as a special case of Corollary 2.2.12 in conjunction with Proposition 2.2.7, which states that all edges of non-reducible polynomials are regular. □

2.3 Properties of the spectral density

We first state two propositions which describe the behaviour of the Stieltjes transform in the upper half-plane close to the edges of the spectrum. The first proposition concerns non-reducible polynomials, whereas the second one covers shifted reducible polynomials. Subsequently, we conclude Propositions 2.2.7 and 2.2.8 from Propositions 2.3.1 and 2.3.2 and state Corollary 2.3.3. It summarizes important properties of the Stieltjes transform close to any regular edge and is used to prove the local law, Theorem 2.2.9. Afterwards, we prove Propositions 2.3.1 and 2.3.2 in the remainder of the section.

Proposition 2.3.1 (Stieltjes transform for non-reducible polynomials)**.** *Let q be a non-reducible quadratic polynomial. Then we have the following behaviour of m close to the edges.*

1. *There are $m_+ < 0$, $c_+ > 0$ and $u > 0$ such that for all $z \in \mathbb{H}$ with $|z - \tau_+| \leq u$ the Stieltjes transform of the limiting spectral density $m = m(z)$ satisfies*

$$m - m_+ = c_+\sqrt{z - \tau_+} + \mathcal{O}(|z - \tau_+|). \tag{2.3.1}$$

2. *There are $m_- > 0$, $c_- > 0$ and $u > 0$ such that for all $z \in \mathbb{H}$ with $|z - \tau_+| \leq u$ the Stieltjes transform of the limiting spectral density $m = m(z)$ satisfies*

$$m - m_- = -c_-\sqrt{\tau_- - z} + \mathcal{O}(|\tau_- - z|). \tag{2.3.2}$$

Here, $\sqrt{\cdot}$ denotes the square root function that maps the positive real axis to itself and with a branch cut along the negative real axis.

Proposition 2.3.2 (Stieltjes transform for shifted reducible polynomials)**.** *Let q be a shifted reducible polynomial of the form (2.2.10). Then we have the following behaviour of m close to the edges.*

1. *There are $m_+ < 0$, $c_+ > 0$ and $u > 0$ such that for all $z \in \mathbb{H}$ with $|z - \tau_+| \leq u$ the function $m = m(z)$ satisfies*

$$m - m_+ = c_+\sqrt{z - \tau_+} + \mathcal{O}(|z - \tau_+|). \tag{2.3.3}$$

2. *There are $c_- > 0$ and $u > 0$ such that for all $z \in \mathbb{H}$ with $|z - \tau_-| \leq u$ the function $m = m(z)$ satisfies*

$$m = \begin{cases} c_-(\tau_- - z)^{-\frac{1}{2}} + \mathcal{O}(1) & \text{if } \xi = 0 \\ c_-(\tau_- - z)^{-\frac{1}{2}} + \mathcal{O}(1) & \text{if } v \in \mathbb{R}^l, \xi < 2 \\ c_-(\tau_- - z)^{-\frac{1}{4}} + \mathcal{O}(1) & \text{if } v \in \mathbb{R}^l, \xi = 2 \\ c_-(\tau_- - z)^{-\frac{1}{2}} + \mathcal{O}(1) & \text{if } v \in \mathbb{C}^l \setminus e^{\mathrm{i}\varphi}\mathbb{R}^l \text{ for all } \varphi \in (-\pi, \pi], s\xi < 2 \\ c_-(\tau_- - z)^{-\frac{1}{3}} + \mathcal{O}(1) & \text{if } v \in \mathbb{C}^l \setminus e^{\mathrm{i}\varphi}\mathbb{R}^l \text{ for all } \varphi \in (-\pi, \pi], s\xi = 2, \end{cases} \tag{2.3.4}$$

where s was defined in (2.2.11). If none of the above conditions is satisfied, there is additionally an $m_- > 0$ such that for all $z \in \mathbb{H}$ with $|z - \tau_-| \leq u$ we have

$$m - m_- = -c_-\sqrt{\tau_- - z} + \mathcal{O}(|\tau_- - z|). \tag{2.3.5}$$

The function $\zeta \mapsto \zeta^p$ is chosen such that the positive real axis is mapped to itself and with a branch cut along the negative real axis for all $p \in \mathbb{R} \setminus \mathbb{Z}$.

Corollary 2.3.3. *Let $q(\mathbf{X})$ have a regular edge at τ_0 and $m_0 = m(\tau_0)$. There is a $u > 0$ such that the function $m = m(z)$ satisfies*

$$|z - \tau_0| \sim |m - m_0|^2 \tag{2.3.6}$$

and

$$\operatorname{Im} m \sim \begin{cases} \sqrt{|E - \tau_0| + \eta} & \text{if } E \in \operatorname{supp}(\rho) \\ \frac{\eta}{\sqrt{|E-\tau_0|+\eta}} & \text{if } E \notin \operatorname{supp}(\rho) \end{cases} \tag{2.3.7}$$

for all $z = E + \mathrm{i}\eta \in \mathbb{H}$ with $|z - \tau_0| \leq u$.

Proof of Proposition 2.2.7 and Proposition 2.2.8. By (2.2.7) we have

$$\rho(E) = \lim_{\eta \searrow 0} \pi^{-1} m(E + \mathrm{i}\eta) \tag{2.3.8}$$

on any interval $I \subset \mathbb{R}$ on which the limit exists for all $E \in I$. For the cases in Proposition 2.3.1 and Proposition 2.3.2 where $m - m_0$ shows a square root behaviour around an edge τ_0, we take the limit on $(\tau_0 - u, \tau_0 + u)$ to prove that the edge is regular. For the cases in Proposition 2.3.2 where m diverges at τ_-, we take the limit on $(\tau_- - u, \tau_- + u) \setminus \{\tau_-\}$ to obtain the respective asymptotic behaviour. □

Let $\mu_i \in \mathbb{R}$ denote the eigenvalues of A, $\widehat{\mu}_i \in \mathbb{R}$ the eigenvalues of $\widehat{A}$ and let $w_i \in \mathbb{R}^l$ be an orthonormal set of eigenvectors of $\widehat{A}$ corresponding to the $\widehat{\mu}_i$. By (2.2.5) the function γ is defined in terms of these quantities as

$$\gamma(m) = -\sum_{i=1}^{l} \frac{\mu_i}{1 + m\mu_i} + \sum_{i=1}^{l} \frac{|\langle w_i, b\rangle|^2 (1 + m\widehat{\mu}_i)}{(1 + 2m\widehat{\mu}_i)^2} m - c. \tag{2.3.9}$$

From now on we consider γ to be defined on its maximal domain, $\mathbb{C} \setminus \mathscr{S}(\gamma)$, where

$$\mathscr{S}(\gamma) := \{-\mu_i^{-1} \in \mathbb{R} : \mu_i \neq 0\} \cup \{-(2\widehat{\mu}_i)^{-1} \in \mathbb{R} : \widehat{\mu}_i \neq 0 \text{ and } \langle w_i, b\rangle \neq 0\} \tag{2.3.10}$$

denotes the poles of γ. We also require the function h, which we define as follows.

Definition 2.3.4. *Let $h : \mathbb{C} \setminus \mathscr{S}(h)$ be given by*

$$h(m) := \frac{1}{m^2} - \gamma'(m) = \frac{1}{m^2} - \sum_{i=1}^{l} \frac{\mu_i^2}{(1 + m\mu_i)^2} - \sum_{i=1}^{l} \frac{|\langle w_i, b\rangle|^2}{(1 + 2m\widehat{\mu}_i)^3}. \tag{2.3.11}$$

Here, $\mathscr{S}(h) := \mathscr{S}(\gamma) \cup \{0\}$ denotes the set of poles of h and $\gamma'(m)$ is the derivative of γ with respect to m.

Proposition 2.2.3 asserts that m is a Stieltjes transform of the measure ρ, which has real and compact support. Thus m is analytic and has positive imaginary part on $\mathbb{H}$. We take the derivative of (2.2.4) with respect to z on $\mathbb{H}$ to find

$$m' = \frac{1}{h(m)}. \tag{2.3.12}$$

As m is a Stieltjes transform of ρ, its analyticity extends to $\mathbb{C} \setminus \operatorname{supp}(\rho)$ and the above equation holds on $\overline{\mathbb{H}} \setminus \operatorname{supp}(\rho)$, where $\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{R}$. Next, we define

$$m_-^* = \begin{cases} \min((0,\infty) \cap \mathscr{S}(h)) & \text{if } (0,\infty) \cap \mathscr{S}(h) \neq \emptyset \\ \infty & \text{otherwise,} \end{cases} \tag{2.3.13}$$

$$m_+^* = \begin{cases} \max((-\infty,0) \cap \mathscr{S}(h)) & \text{if } (-\infty,0) \cap \mathscr{S}(h) \neq \emptyset \\ -\infty & \text{otherwise.} \end{cases} \tag{2.3.14}$$

In other words, $(m_+^*, 0)$ and $(0, m_-^*)$ are the maximal intervals to the left and the right of the origin, where h is continuous. The following lemmata describe the existence and characterization of roots of h both for the shifted reducible and the non-reducible case on $(m_+^*, 0)$ and $(0, m_-^*)$.

Lemma 2.3.5. *Let q be a non-reducible quadratic polynomial. Then we have the following.*

1. *The function h has a unique root m_- in $(0, m_-^*)$ and h has a unique root m_+ in $(m_+^*, 0)$. Both of them are of first order. More precisely, they satisfy*

$$\pm h'(m_\pm) > 0. \tag{2.3.15}$$

2. *The positions of the edges $\tau_\pm$, defined in Definition 2.2.5, are given in terms of $m_\pm$ by*

$$\tau_\pm = -m_\pm^{-1} - \gamma(m_\pm). \tag{2.3.16}$$

Lemma 2.3.6. *Let h be the function introduced in Definition 2.3.4 for a reducible polynomial $q = q_r$ of the form (2.2.10) with normalized $v \in \mathbb{C}^l$ and $\xi \geq 0$. Then the following holds true.*

1. *The function h has a unique root m_+ in $(m_+^*, 0)$, which is of first order and satisfies*

$$h'(m_+) > 0. \tag{2.3.17}$$

2. *The function h is continuous on $(0,\infty)$. It has no root in $(0,\infty)$ if and only if one of the following hold true:*

   a) $\xi = 0$;

   b) or $v \in \mathbb{R}^l$ and $\xi \leq 2$;

c) or $v \in \mathbb{C}^l \setminus e^{\mathrm{i}\varphi}\mathbb{R}^l$ *for all* $\varphi \in (-\pi, \pi]$ *and* $s\xi \leq 2$.

If none of the above conditions are satisfied, h *does have a unique root* $m_- \in (0, \infty)$*, which is of first order and satisfies*

$$h'(m_-) < 0. \tag{2.3.18}$$

If h *does not have any roots in* $(0, \infty)$*, the following asymptotic behaviour is observed for* $m \to \infty$*:*

$$h(m) \sim \begin{cases} m^{-3} & \text{if } \xi = 0 \\ m^{-3} & \text{if } v \in \mathbb{R}^l \text{ and } \xi < 2 \\ m^{-5} & \text{if } v \in \mathbb{R}^l \text{ and } \xi = 2 \\ m^{-3} & \text{if } v \in \mathbb{C}^l \setminus e^{\mathrm{i}\varphi}\mathbb{R}^l \text{ for all } \varphi \in (-\pi, \pi] \text{ and } s\xi < 2 \\ m^{-4} & \text{if } v \in \mathbb{C}^l \setminus e^{\mathrm{i}\varphi}\mathbb{R}^l \text{ for all } \varphi \in (-\pi, \pi] \text{ and } s\xi = 2. \end{cases} \tag{2.3.19}$$

Here s *was defined in* (2.2.11).

3. *The positions of the edges* $\tau_\pm$*, defined in Definition 2.2.5, are given by*

$$\tau_+ = -m_+^{-1} - \gamma(m_+) \tag{2.3.20}$$

and

$$\tau_- = \begin{cases} -m_-^{-1} - \gamma(m_-) & \text{if } m_- > 0 \text{ with } h(m_-) = 0 \text{ exists;} \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.21}$$

The proof of the lemmata is deferred until after the proof of Propositions 2.3.1 and 2.3.2.

Proof of Propositions 2.3.1 and 2.3.2. In both the non-reducible and the shifted reducible cases the function h has a unique root m_+ in $(m_+^*, 0)$. For all $m \in (m_+, 0)$ the derivative of the inverse $z = z(m)$ of $m = m(z)$ with respect to m is given by

$$\frac{\mathrm{d}}{\mathrm{d}m} z = h(m). \tag{2.3.22}$$

Then, by Lemma 2.3.5 and Lemma 2.3.6, there is an $r \sim 1$ and a $c_+ > 0$, only depending on the coefficients of q, such that for all $m \in (m_+, m_+ + r)$ we have

$$\frac{\mathrm{d}}{\mathrm{d}m} z = h(m) = 2c_+^2(m - m_+)(1 + \mathcal{O}(m - m_+)) \tag{2.3.23}$$

and thus

$$z - \tau_+ = c_+^2(m - m_+)^2(1 + \mathcal{O}(m - m_+)). \tag{2.3.24}$$

Taking the square root of the above equation, we obtain

$$m - m_+ = c_+\sqrt{z - \tau_+} + \mathcal{O}(|z - \tau_+|) \tag{2.3.25}$$

for all $z \in (\tau_+, \tau_+ + u)$ for some $u > 0$. As m is holomorphic on $\mathbb{H} \cup (\tau_+, \infty)$ and has positive imaginary part on $\mathbb{H}$, the relation (2.3.25) also holds for all $z \in \mathbb{H}$ with $|z - \tau_+| \leq u$ for some $u > 0$.

In all cases, where h has a root m_- in $(0, m_-^*)$, we find by the same argument that

$$m - m_- = -c_- \sqrt{\tau_- - z} + \mathcal{O}(|\tau_- - z|) \tag{2.3.26}$$

for all $z \in \mathbb{H}$ with $|z - \tau_+| \leq u$ for some $u, c_- > 0$.

Finally, in the cases where h does not have a root in $(0, \infty)$, it is continuous on the whole interval and we have that for all $m \in (0, \infty)$ the derivative of z with respect to m is given by

$$\frac{\mathrm{d}}{\mathrm{d}m} z = h(m). \tag{2.3.27}$$

Then, by Lemma 2.3.6, there is an $R \sim 1$ and a $c_0 > 0$, only depending on the coefficients of q, such that for all $m \in (R, \infty)$ we have

$$\frac{\mathrm{d}}{\mathrm{d}m} z = h(m) = c_0 m^{-p}(1 + \mathcal{O}(m^{-1})) \quad \text{and thus} \quad z = -\frac{c_0 m^{-(p-1)}}{p-1}(1 + \mathcal{O}(m^{-1})), \tag{2.3.28}$$

with $p \in \{3, 4, 5\}$. Inverting the relation, we obtain

$$m = c_-(-z)^{-\frac{1}{p-1}} + \mathcal{O}(1) \tag{2.3.29}$$

for all $z \in (-u, 0)$ and some $u, c_- > 0$. Again, as m is holomorphic on $\mathbb{H} \cup (-\infty, 0)$ and has positive imaginary part on $\mathbb{H}$, the relation (2.3.29) also holds for all $z \in \mathbb{H}$ with $|z| \leq u$ for some $u > 0$.

□

Proof of Lemma 2.3.5 and Lemma 2.3.6. We prove both lemmata simultaneously and remark where it becomes necessary to distinguish the different polynomials q. We first investigate the existence and order of roots of h. Let h_q denote the rational function (2.3.11) with the dependence on q being made explicit. We have $h_q(-m) = h_{-q}(m)$. Therefore any root of h_q on $(0, \infty)$ corresponds to a root of h_{-q} on $(-\infty, 0)$. For simplicity, the proofs are thus only formulated for $m \in (-\infty, 0)$. To obtain the corresponding result for $m \in (0, \infty)$ for the function h_q, we consider the function h_{-q} on $(-\infty, 0)$ instead.

We now prove for all polynomials q as in (2.1.1) that $h'(m) > 0$ for all $m \in (m_+^*, 0)$ with $h(m) \geq 0$ and thus h can only have first-order roots on the interval.

Note that on $(m_+^*, 0)$ the inequality $h \geq 0$ is equivalent to

$$\sum_{i=1}^{l} \frac{(m\mu_i)^2}{(1 + m\mu_i)^2} + \sum_{\substack{i=1, \\ \langle w_i, b\rangle \neq 0}}^{l} (m|\langle w_i, b\rangle|)^2 \frac{1}{(1 + 2m\widehat{\mu}_i)^3} \leq 1 \tag{2.3.30}$$

and that on $(m_+^*, 0)$ the inequality $h' > 0$ is equivalent to

$$\sum_{i=1}^{l} \frac{(m\mu_i)^3}{(1+m\mu_i)^3} + \sum_{\substack{i=1,\\ \langle w_i,b\rangle\neq 0}}^{l} (m|\langle w_i, b\rangle|)^2 \frac{3(m\widehat{\mu}_i)}{(1+2m\widehat{\mu}_i)^4} < 1. \tag{2.3.31}$$

For $m \in (m_+^*, 0)$ we have also that $m\mu_i > -1$ for all $i \in [\![l]\!]$ and $m\widehat{\mu}_i > -\frac{1}{2}$ for all $i \in [\![l]\!]$ such that $\langle w_i, b\rangle \neq 0$ by the definition of m_+^* in (2.3.14). Furthermore, Lemma 2.6.1 ensures that $\max_{i\in[\![l]\!]}\{m\mu_i\} \geq \max_{i\in[\![l]\!]}\{m\widehat{\mu}_i\}$. Thus the desired implication, $h \geq 0$ implies $h' > 0$, follows directly from the following lemma by identifying (2.3.32) and (2.3.33) right below with (2.3.30) and (2.3.31) respectively.

Lemma 2.3.7. *Let $n, k \in \mathbb{N}$ with $n \geq k$. Let y_i, $i \in [\![n]\!]$ and let $\widehat{y}_j$, $j \in [\![k]\!]$ be collections of real numbers, both of them sorted in non-increasing order, $y_1 \geq y_2 \geq \ldots \geq y_n$ and $\widehat{y}_1 \geq \widehat{y}_2 \geq \ldots \geq \widehat{y}_k$. Suppose they satisfy*

1. *$y_1 \geq \widehat{y}_1$;*
2. *$y_n > -1$ and $\widehat{y}_k > -\frac{1}{2}$.*

Furthermore let $c_j > 0$, $j \in [\![k]\!]$. Then the inequality

$$\sum_{i=1}^{n} \frac{y_i^2}{(y_i+1)^2} + \sum_{j=1}^{k} c_j^2 \frac{1}{(2\widehat{y}_j+1)^3} \leq 1 \tag{2.3.32}$$

implies

$$\sum_{i=1}^{n} \frac{y_i^3}{(y_i+1)^3} + 3\sum_{j=1}^{k} c_j^2 \frac{\widehat{y}_j}{(2\widehat{y}_j+1)^4} \leq 1-\nu \tag{2.3.33}$$

for some $\nu > 0$, depending only on y_1.

The proof of the lemma is deferred to the end of the section.

To characterize the existence of roots of h we introduce the following notions. If A is a rank one matrix with real entries, then $\widehat{A} = A$ and $\widehat{A}$ is a rank one matrix as well. We will denote its non-zero eigenvalue by μ and the corresponding normalized eigenvector by w. If, on the other hand, A is a rank one matrix with $A \in \mathbb{C}^{l\times l} \setminus \mathbb{R}^{l\times l}$, then $\widehat{A}$ is a rank two matrix by Lemma 2.6.1 in the appendix. We denote its non-zero eigenvalues by $\mu_\pm$ and the corresponding normalized eigenvectors by $w_\pm$.

Lemma 2.3.8. *Let m_+^* be defined as in (2.3.14). The function h has no root in $(m_+^*, 0)$ if and only if A is a negative semi-definite rank one matrix, $b \in \mathrm{Image}(\widehat{A})$ and either*

1. *$A \in \mathbb{R}^{l\times l}$ and $\|b\| \leq 4\|A\|$;*

2. *or* $A \in \mathbb{C}^{l\times l} \setminus \mathbb{R}^{l\times l}$ *and*

$$\frac{|\langle b, w_+\rangle|^2}{r_+^3} + \frac{|\langle b, w_-\rangle|^2}{r_-^3} \leq (4\|A\|)^2, \tag{2.3.34}$$

with $r_\pm := -\|A\|^{-1}\mu_\pm$.

If h *has no root in* $(m_+^*, 0)$*, then we have* $m_+^* = -\infty$ *and the asymptotic behaviour of* h *for* $m \to -\infty$ *is given by*

$$h(m) \sim \begin{cases} \frac{1}{(-m)^3} & A \in \mathbb{R}^{l\times l} \text{and } \|b\| < 4\|A\| \\ \frac{1}{(-m)^5} & A \in \mathbb{R}^{l\times l} \text{and } \|b\| = 4\|A\| \\ \frac{1}{(-m)^3} & A \in \mathbb{C}^{l\times l} \setminus \mathbb{R}^{l\times l} \text{ and } \frac{|\langle b,w_+\rangle|^2}{|r_+|^3} + \frac{|\langle b,w_-\rangle|^2}{|r_-|^3} < (4\|A\|)^2 \\ \frac{1}{(-m)^4} & A \in \mathbb{C}^{l\times l} \setminus \mathbb{R}^{l\times l} \text{ and } \frac{|\langle b,w_+\rangle|^2}{|r_+|^3} + \frac{|\langle b,w_-\rangle|^2}{|r_-|^3} = (4\|A\|)^2. \end{cases} \tag{2.3.35}$$

The proof of the lemma is also deferred to the end of the section.

Note that h can have at most one root on $(m_+^*, 0)$. To see this, recall that $h(m) \geq 0$ implies $h'(m) > 0$. Therefore, if m_+ is a root of h on $(m_+^*, 0)$, we know that h is strictly monotonously increasing on $(m_+, 0)$. Thus, any root of h on $(m_+^*, 0)$ is also the largest one on the interval; therefore it must be the unique root.

Now let q be a non-reducible polynomial. If $\operatorname{rank} A \geq 2$, then h has a root in $(-\infty, 0)$ by Lemma 2.3.8. If $\operatorname{rank} A = 1$, we can express A as $A = \alpha ww^*$ for some $\alpha \in \mathbb{R} \setminus \{0\}$ and normalized $w \in \mathbb{C}^l$. The vector b cannot be of the form $b = \alpha_1 \operatorname{Re} w + \alpha_2 \operatorname{Im} w$ for any $\alpha_1, \alpha_2 \in \mathbb{R}$ as q would be reducible otherwise. Thus $b \notin \operatorname{Image} A$ and again h has a root in $(-\infty, 0)$ by Lemma 2.3.8, completing the proof of Statement 1 of Lemma 2.3.5.

Next, consider a shifted reducible polynomial $q = q_r$ for q_r as in (2.2.10). If we write q in terms of (2.1.1) we have $A = vv^* \geq 0$. Applying Lemma 2.3.5 we obtain that h has a root in $(-\infty, 0)$ and thereby prove Statement 1 of Lemma 2.3.6.

Finally, consider $q = -q_r$ for q_r as in (2.2.10). Then the coefficients A and b associated with q are given by $A = -vv^*$ and $b = 2\xi \operatorname{Re} v$. The matrix A is a negative semi-definite rank one matrix. For $\mu = \langle \operatorname{Re} v, \operatorname{Im} v\rangle \neq 0$ the non-zero eigenvalues of $\widehat{A} = -\frac{1}{2}(v^*v + vv^*) = -(\operatorname{Re} v)^* \operatorname{Re} v - (\operatorname{Im} v)^* \operatorname{Im} v$ are given by

$$\mu_\pm = \sigma^2 + \mu a_\pm, \tag{2.3.36}$$

where $\sigma = \|\operatorname{Re} v\|$ and $a_\pm$ was defined in (2.2.12). The corresponding normalized eigenvectors are

$$w_\pm = \frac{1}{\sigma^2 + a_\pm(1-\sigma^2) + 2a_\pm\mu} (\operatorname{Re} v + a_\pm \operatorname{Im} v). \tag{2.3.37}$$

For $\mu = 0$ the eigenvector-eigenvalue pairs are given by

$$(w_+, \mu_+) = \left(\frac{\operatorname{Im} v}{1-\sigma^2}, 1-\sigma^2 \right), \quad (w_-, \mu_-) = \left(\frac{\operatorname{Re} v}{\sigma^2}, \sigma^2 \right). \tag{2.3.38}$$

Now recall that $h_q(m) = h_{-q}(-m) = h_{q_r}(-m)$ and Statement 2 of Lemma 2.3.6 now follows from applying Lemma 2.3.8 to h_q.

It remains to prove the relations between $m_\pm$ and $\tau_\pm$ stated in (2.3.16), (2.3.20) as well as (2.3.21). Let q be either a non-reducible polynomial or a reducible polynomial of the form (2.2.10). For $\widetilde{m} \in (m_+^*, 0)$ define

$$z(\widetilde{m}) := -\frac{1}{\widetilde{m}} - \gamma(\widetilde{m}). \tag{2.3.39}$$

The function z is real analytic on $(m_+^*, 0)$ with derivative $z'(\widetilde{m}) = h(\widetilde{m})$. By Lemma 2.3.5, Part 1 and Lemma 2.3.6, Part 1, the function h has a unique root m_+ in $(m_+^*, 0)$ and $h(\widetilde{m}) > 0$ for all $\widetilde{m} \in (m_+, 0)$. Therefore,the function z is invertible on $(m_+, 0)$ and its inverse function $\widetilde{m} : (z_+, \infty) \to (m_+, 0)$, with $z_+ := z(m_+)$, is also real analytic and monotonically increasing. Since $z'(m_+) = h(m_+) = 0$, the analyticity of $\widetilde{m}$ cannot be extended onto any neighbourhood of z_+.

By (2.3.39), the function $\widetilde{m}$ satisfies (2.2.4) and has the asymptotic behaviour $\widetilde{m} = -z^{-1} + \mathcal{O}(z^2)$ for $z \to \infty$. At the same time, m, the function uniquely defined by Proposition 2.2.3, is a Stieltjes transform of a probability measure with support $[\tau_-, \tau_+]$. As such, it can be analytically extended to $\mathbb{C} \setminus [\tau_-, \tau_+]$ but not to a neighbourhood of τ_+ and the extension is real-valued on $\mathbb{R} \setminus [\tau_-, \tau_+]$. As an extension it also satisfies (2.2.4) on $\mathbb{C} \setminus [\tau_-, \tau_+]$ and has the asymptotic behaviour $m = -z^{-1} + \mathcal{O}(z^2)$. In particular, the restriction of the extension to $\mathbb{R} \setminus [\tau_-, \tau_+]$, called $m_{\mathbb{R}}$, is a real analytic function that also satisfies (2.2.4), has the asymptotic behaviour $m_{\mathbb{R}} = -z^{-1} + \mathcal{O}(z^2)$ and cannot be analytically extended to a neighbourhood of τ_+. As satisfying (2.2.4) and having the asymptotic behaviour $m = -z^{-1} + \mathcal{O}(z^2)$ uniquely define an analytic function on a neighbourhood of ∞, we have $\widetilde{m} = m_{\mathbb{R}}$ on (C, ∞) for some $C > 0$. If we assume that $\tau_+ > z_+$, then $\widetilde{m}$ would be an analytical continuation of $m_{\mathbb{R}}$ to some neighbourhood of τ_+, which is a contradiction and vice versa. Therefore we must have $z_+ = \tau_+$. In particular, we have

$$\tau_+ = z(m_+) = -\frac{1}{m_+} - \gamma(m_+). \tag{2.3.40}$$

By an analogous argument to the left of the spectrum we find

$$\tau_- = z(m_-) = -\frac{1}{m_-} - \gamma(m_-) \tag{2.3.41}$$

if m_- exists. Now, let $q = q_r$ be a shifted reducible polynomial of the form (2.2.10) such that h has no root on $(0, \infty)$. Then the function $z(\widetilde{m})$ defined on $(0, \infty)$ by (2.3.39) is a monotonously increasing analytic function on the entirety of its domain. It is therefore invertible and its inverse function, $\widetilde{m}$, is analytic on $(-\infty, z_\infty)$ with

$z_\infty := \lim_{x\to\infty} z(x)$ but cannot be analytically extended to a neighbourhood of z_∞. Along the lines of the above argument, $m_{\mathbb{R}}$ would be an analytical continuation of $\widetilde{m}$ in a neighbourhood of z_∞ if $\tau_- > z_\infty$ and $\widetilde{m}$ would be an analytical continuation of $m_{\mathbb{R}}$ in a neighbourhood of τ_- if $z_\infty > \tau_-$. Thus we have $\tau_- = z_\infty$ and in particular

$$\tau_- = \lim_{x\to\infty} -\frac{1}{x} - \gamma(x). \tag{2.3.42}$$

For reducible q_r as in (2.2.10), we have $A = vv^*$, $b = -\xi(v+\bar{v})$ and $c = \xi^2$. Taking the limit (2.3.42) with these parameters we obtain $\tau_- = 0$. □

Proof of Lemma 2.3.7. Let $y \in (-1,\infty)$ and $\widehat{y} \in (-\frac{1}{2},\infty)$. We define

$$h_1(y) = \frac{y}{y+1} \quad \text{and} \quad h_2(\widehat{y}) = \frac{1}{2\widehat{y}+1}. \tag{2.3.43}$$

Using this notation, (2.3.32) can be expressed as

$$g(\mathbf{y},\widehat{\mathbf{y}},\mathbf{c}) := \sum_{i=1}^{n} h_1(y_i)^2 + \sum_{j=1}^{k} c_j^2 h_2^3(\widehat{y}_j) \le 1 \tag{2.3.44}$$

and (2.3.33) as

$$f(\mathbf{y},\widehat{\mathbf{y}},\mathbf{c}) := \sum_{i=1}^{n} h_1(y_i)^3 + 3\sum_{j=1}^{k} c_j^2 \widehat{y}_j h_2^4(\widehat{y}_j) \le 1-\nu \tag{2.3.45}$$

To prove the lemma it is then sufficient to show that $g \le 1$ implies $f \le 1-\nu$.

First assume $g = 0$. Since all summands of g are non-negative, it follows that $y_i = 0$ for all $i \in [\![n]\!]$ and $c_j = 0$ for all $j \in [\![k]\!]$ and so f vanishes as well. Thus we assume $g \neq 0$ from now on and we will prove that $g \le 1$ implies $f < g$.

We will prove the cases $\widehat{y}_1 \le \frac{1}{2}$, $\frac{1}{2} < \widehat{y}_1 < 1$ and $\widehat{y}_1 \ge 1$ separately and start with $\widehat{y}_1 \le \frac{1}{2}$. Then $\widehat{y}_j \le \frac{1}{2}$ for all j. Note that $\widehat{y}_j h_2^4(\widehat{y}_j) \le \widehat{y}_1 h_2(\widehat{y}_1) h_2^3(\widehat{y}_j)$ and $h_1^3(y_i) \le h_1(y_1)h_1^2(y_i)$ hold for all i and j since $\widehat{y} \mapsto \widehat{y}h_2(\widehat{y})$ and h_1 are monotonously increasing and the $\widehat{y}_j$ and y_i are sorted in non-increasing order. Thus

$$\begin{aligned} f(\mathbf{y},\widehat{\mathbf{y}},\mathbf{c}) &\le h_1(y_1)\sum_{i=1}^{n} h_1^2(y_i) + 3\widehat{y}_1 h_2(\widehat{y}_1)\sum_{j=1}^{k} c_j^2 h_2^3(\widehat{y}_j) \\ &\le h_1(y_1)\sum_{i=1}^{n} h_1^2(y_i) + \frac{3}{4}\sum_{j=1}^{k} c_j^2 h_2^3(\widehat{y}_j) \le \max\left\{\frac{y_1}{1+y_1}, \frac{3}{4}\right\} g(\mathbf{y},\widehat{\mathbf{y}},\mathbf{c}). \end{aligned} \tag{2.3.46}$$

The second inequality holds since $3\widehat{y}_1 h_2(\widehat{y}_1) \le \frac{3}{4}$ due to $\widehat{y}_1 \le \frac{1}{2}$ and the third one because all summands in both sums are non-negative. The relation $f \le 1-\nu$ for $g \le 1$ then follows in this regime.

For $\frac{1}{2} < \widehat{y}_1 < 1$, we estimate every summand in f but the y_1 term by the corresponding term in g and we find

$$f(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) \leq g(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) - h_1(y_1)^2 + h_1(y_1)^3. \tag{2.3.47}$$

This upper bound is valid since $\widehat{y}_j \leq \widehat{y}_1 < 1$ for all j. Thus for $g \leq 1$ we find

$$f(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) \leq g(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) - \left(\frac{y_1}{y_1+1}\right)^2 \frac{1}{1+y_1} < 1 - \frac{2}{27}, \tag{2.3.48}$$

where in the last step we used $y_1 \geq \widehat{y}_1 > \frac{1}{2}$.

Now let $\widehat{y}_1 \geq 1$. Let k_0 be the largest integer such that $\widehat{y}_{k_0} \geq 1$. Then

$$\begin{aligned} f_0(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) &:= \sum_{i=2}^{n} h_1^3(y_i) + \sum_{j=k_0+1}^{k} 3c_j^2 \widehat{y}_j h_2^4(\widehat{y}_j) \\ &\leq \sum_{i=2}^{n} h_1^2(y_i) + \sum_{j=k_0+1}^{k} c_j^2 h_2^3(\widehat{y}_j) =: g_0(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) \end{aligned} \tag{2.3.49}$$

is concluded by again estimating each term individually (either of the sums might be empty). Note that $g_0 \geq 0$. Therefore we only need to prove that

$$g_1(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) := (g - g_0)(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) = h_1^2(y_1) + \sum_{j=1}^{k_0} c_j^2 h_2^3(\widehat{y}_j) \leq 1 \tag{2.3.50}$$

implies

$$f_1(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) := (f - f_0)(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) = h_1^3(y_1) + \sum_{j=1}^{k_0} 3c_j^2 \widehat{y}_j h_2^4(\widehat{y}_j) \leq g_1(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) - \nu \tag{2.3.51}$$

for some $\nu > 0$, only depending on y_1, to conclude the proof of the lemma.

For $\widehat{y} \in [1, \infty)$ the inequality $h_2^3(\widehat{y}) \leq 3\widehat{y} h_2^4(\widehat{y})$ is satisfied and consequently

$$\sum_{j=1}^{k_0} h_2^3(\widehat{y}_j) \leq \sum_{j=1}^{k_0} 3\widehat{y}_j h_2^4(\widehat{y}_j) \tag{2.3.52}$$

holds. Therefore for any fixed $(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c})$ and $r \in \mathbb{R}_+$ the difference $(g_1 - f_1)(\mathbf{y}, \widehat{\mathbf{y}}, r\mathbf{c})$ decreases monotonically in r. Hence we only need to prove $(g_1 - f_1)(\mathbf{y}, \widehat{\mathbf{y}}, r^*\mathbf{c}) \geq \nu$ for

$$r^* := \sup\{r \geq 0 |\, g_1(\mathbf{y}, \widehat{\mathbf{y}}, r\mathbf{c}) \leq 1\}. \tag{2.3.53}$$

On the other hand $g_1(\mathbf{y}, \widehat{\mathbf{y}}, r\mathbf{c})$ increases monotonically in r and therefore

$$g_1(\mathbf{y}, \widehat{\mathbf{y}}, r^*\mathbf{c}) = 1.$$

It is thus sufficient to prove that $g_1 = 1$ implies $f_1 \leq 1 - \nu$. Hence we assume $g_1 = 1$. Then

$$\sum_{j=1}^{k_0} c_j^2 h_2^3(\widehat{y}_j) = 1 - h_1^2(y_1) \tag{2.3.54}$$

and

$$\begin{aligned} f_1(\mathbf{y}, \widehat{\mathbf{y}}, \mathbf{c}) &= h_1^3(y_1) + \sum_{j=1}^{k_0} 3c_j^2 \widehat{y}_j h_2^4(\widehat{y}_j) \leq h_1^3(y_1) + \max_{j \in [\![k_0]\!]} \{3\widehat{y}_j h_2(\widehat{y}_j)\} \sum_{j=1}^{k_0} c_j^2 h_2^3(\widehat{y}_j) \\ &\leq h_1^3(y_1) + 3\widehat{y}_1 h_2(\widehat{y}_1)(1 - h_1^2(y_1)) \leq h_1^3(y_1) + 3y_1 h_2(y_1)(1 - h_1^2(y_1)) \\ &= \frac{1}{(1+y_1)^3}(y_1^3 + 3y_1^2 + 3y_1) = 1 - \frac{1}{(1+y_1)^3}. \end{aligned} \tag{2.3.55}$$

The inequalities holds because $yh_2(y)$ increases monotonically in y and $y_1 \geq \widehat{y}_1 \geq \widehat{y}_j$ for all j. Combining this with (2.3.46) and (2.3.48) concludes the proof of the lemma. □

Proof of Lemma 2.3.8. Note that $h(m) > 0$ for m close to 0, since the m^{-2} term is dominating h from (2.3.11) in this regime. First, we consider the situation, where h has a pole in $(-\infty, 0)$. In other words, we have $m_+^* > -\infty$, where m_+^* was defined in (2.3.14). All poles of h on $(-\infty, 0)$ are either of order two or of order three so the pole at m_+^* will be as well. If the pole is of order two, then the behaviour around m^* is given by $\lim_{m \to m_+^*} h(m) = -\infty$ and if it is of order three h behaves like $\lim_{m \searrow m_+^*} h(m) = -\infty$ and $\lim_{m \nearrow m_+^*} h(m) = \infty$. Since h is continuous outside of its poles, the existence of a pole thus implies the existence of a root in $(m_+^*, 0)$. Now let h have no poles in $(-\infty, 0)$, i.e. $m_+^* = -\infty$. Then h is analytic on $(-\infty, 0)$ and h has a root in $(-\infty, 0)$ if and only if $h(m)$ is negative for some $m \in (-\infty, 0)$.

We separate multiple cases.

Case 1 ($\operatorname{rank}(A) \geq 2$)**.** *Either h has a pole in $(-\infty, 0)$ or $b^t(1 + 2m\widehat{A})^{-3}b \geq 0$ for all $m \in (-\infty, 0)$, thus*

$$h(m) \leq \frac{1}{m^2} - \operatorname{Tr} A^2(1 + mA)^{-2} = -\frac{\operatorname{rank}(A) - 1}{m^2} + \mathcal{O}(m^{-3}) < 0 \tag{2.3.56}$$

for sufficiently large $-m$ and h has a root either way.

Case 2 ($\operatorname{rank}(A) = 1$ and $A \geq 0$)**.** *The matrices A and $\widehat{A}$ have a positive eigenvalue, therefore h will have a pole in $(-\infty, 0)$ and thus also a root.*

Case 3 ($\operatorname{rank}(A) = 1$, $A \leq 0$ and $b \notin \operatorname{Image} \widehat{A}$)**.** *In this case, h has no pole on $(-\infty, 0)$ and there is a vector w_0 in the kernel of $\widehat{A}$ such that $\langle w_0, b \rangle \neq 0$. From (2.3.11) we have that h is estimated*

$$h(m) \leq \frac{1}{m^2} - |\langle w_0, b \rangle|^2 < 0, \tag{2.3.57}$$

where the last inequality holds for $-m$ sufficiently large and so h has a root as well.

Case 4 ($\operatorname{rank}(A) = 1$, $A \leq 0$, $b \in \operatorname{Image} \widehat{A}$ and $A \in \mathbb{R}^{l\times l}$)**.** *Since A is real-symmetric, we have $A = \widehat{A}$ and $\widehat{A}$ is also a rank one matrix and h has no poles in $(-\infty, 0)$. Either $b = 0$ or b is an eigenvector of $\widehat{A}$ with eigenvalue $\mu = -\|A\|$ since $A = \widehat{A}$, $\operatorname{rank} A = 1$ and $A \leq 0$. In both cases h becomes*

$$h(m) = \frac{1}{m^2}\left(1 - \frac{(\mu m)^2}{(1+\mu m)^2}\right) - \|b\|^2 \frac{1}{(1+2\mu m)^3} \tag{2.3.58}$$

To find the asymptotics for $m \to -\infty$ consider $0 < \zeta < 1$, with $\zeta = (\mu m)^{-1}$. We find

$$\begin{aligned} &h(m) \\ =&\|A\|^2\zeta^2\left(1 - \frac{1}{(\zeta+1)^2}\right) - \frac{\|b\|^2}{8}\frac{\zeta^3}{(\frac{1}{2}\zeta+1)^3} \\ =&\zeta^3\left(\|A\|^2\sum_{k=0}^{\infty}(k+2)(-\zeta)^k - \frac{\|b\|^2}{8}\sum_{k=0}^{\infty}\frac{(k+1)(k+2)}{2}(-\frac{1}{2}\zeta)^k\right) \\ =&\left(2\|A\|^2 - \frac{\|b\|^2}{8}\right)\zeta^3 + 3\left(\frac{\|b\|^2}{16} - \|A\|^2\right)\zeta^4 + \left(4\|A\|^2 - \frac{3\|b\|^2}{8}\right)\zeta^5 + \mathcal{O}(\zeta^6). \end{aligned} \tag{2.3.59}$$

Therefore h has a root if $\|b\| > 4\|A\|$. If $\|b\| \leq 4\|A\|$ then the leading order coefficient of the expansion is positive. Since $h(m) \geq 0$ implies $h'(m) > 0$ this implies $h(m) > 0$ for all $m \in (-\infty, 0)$. Therefore there is no root and the first two cases in (2.3.35) follow.

Case 5 ($\operatorname{rank}(A) = 1$, $A \leq 0$, $b \in \operatorname{Image} \widehat{A}$ and $A \in \mathbb{C}^{l\times l} \setminus \mathbb{R}^{l\times l}$)**.** *Since A is not real-symmetric, we have $\operatorname{rank} \widehat{A} = 2$ by Lemma 2.6.1 and h has no poles in $(-\infty, 0)$. Since $b \in \operatorname{Image} \widehat{A}$, the function h can be expressed as*

$$h(m) = \frac{1}{m^2}\left(1 - \frac{(\mu m)^2}{(1+\mu m)^2}\right) - |\langle b, w_+\rangle|^2 \frac{1}{(1+2\mu_+ m)^3} - |\langle b, w_-\rangle|^2 \frac{1}{(1+2\mu_- m)^3} \tag{2.3.60}$$

To find the asymptotics for $m \to -\infty$ consider $0 < \zeta < 1$, with $\zeta = (-\|A\|m)^{-1}$. We find

$$\begin{aligned} h(m) =&\|A\|^2\zeta^2\left(1 - \frac{1}{(\zeta+1)^2}\right) - \frac{|\langle b, w_+\rangle|^2}{8r_+^3}\frac{\zeta^3}{(\frac{1}{2r_+}\zeta+1)^3} - \frac{|\langle b, w_-\rangle|^2}{8r_-^3}\frac{\zeta^3}{(\frac{1}{2r_+}\zeta+1)^3} \\ =&\left(2\|A\|^2 - \frac{|\langle b, w_+\rangle|^2}{8r_+^3} - \frac{|\langle b, w_-\rangle|^2}{8r_-^3}\right)\zeta^3 \\ &+ 3\left(\frac{|\langle b, w_+\rangle|^2}{16r_+^4} - \frac{|\langle b, w_-\rangle|^2}{16r_-^4} - \|A\|^2\right)\zeta^4 + \mathcal{O}(\zeta^5). \end{aligned} \tag{2.3.61}$$

By an analogous consideration to the above case, h has a root if and only if

$$\frac{|\langle b, w_+\rangle|^2}{r_+^3} + \frac{|\langle b, w_-\rangle|^2}{r_-^3} > (4\|A\|)^2 \tag{2.3.62}$$

and the remaining two cases in (2.3.35) *follow.*

□

2.4 Linearization

Throughout the remainder of the paper, we suppress the index of the identity matrix if it is in dimension $l+1$ or in $(l+1)N$, i.e. $I := I_{l+1} \in \mathbb{C}^{(l+1)\times(l+1)}$ and $\mathbf{I} := \mathbf{I}_{(l+1)N} \in \mathbb{C}^{(l+1)N\times(l+1)N}$. In all other cases, we still write out the dimension in the index.

To prove that $\mathbf{g}$ converges towards m in the sense laid out in Theorem 2.2.9, we use the linearization method, which we briefly introduce here. For more details on the construction of linearizations in the context of random matrices, see e.g. [28, 42, 43]. First, assume that A is invertible. We define the linearization $\mathbf{L}$ of q as

$$\mathbf{L} = K_0 + \sum_{j=1}^{l} K_j \otimes \mathbf{X}_j \in \mathbb{C}^{(l+1)N\times(l+1)N}, \tag{2.4.1}$$

where

$$K_0 = \begin{pmatrix} c & 0 \\ 0 & -A^{-1} \end{pmatrix} \in \mathbb{C}^{(l+1)\times(l+1)} \quad \text{and} \quad K_j = \begin{pmatrix} b_j & e_j^t \\ e_j & 0 \end{pmatrix} \in \mathbb{C}^{(l+1)\times(l+1)} \tag{2.4.2}$$

for $j \in [\![l]\!]$ and e_j being the j^{th} standard euclidean base vector. Here we made use of our convention to identify any matrix $R \in \mathbb{C}^{k\times n}$, $k, n \in \mathbb{N}$ with $R \otimes \mathbf{I}_N \in \mathbb{C}^{kN\times nN}$ introduced in Section 2.1. Let $J \in \mathbb{C}^{(l+1)\times(l+1)}$ be the orthogonal projection onto the first entry. For $\delta \in [0,1]$ and $z = E + \mathrm{i}\eta \in \mathbb{H}$ we define

$$\mathbf{G}_\delta = (\mathbf{L} - zJ - \mathrm{i}\eta\delta(\mathbf{I} - J))^{-1} \in \mathbb{C}^{(l+1)N\times(l+1)N} \tag{2.4.3}$$

$\mathbf{G}_\delta$ is a generalized resolvent and using the Schur complement formula we obtain

$$\mathbf{G}_\delta = \begin{pmatrix} \mathbf{g}_\delta & \mathbf{g}_\delta \mathbf{X}^t A_\delta \\ A_\delta \mathbf{X} \mathbf{g}_\delta & -A_\delta + A_\delta \mathbf{X} \mathbf{g}_\delta \mathbf{X}^t A_\delta \end{pmatrix}, \tag{2.4.4}$$

where

$$\begin{aligned} A_\delta &= A(I_l + \mathrm{i}\delta\eta A)^{-1} \in \mathbb{C}^{l\times l}, \\ \mathbf{g}_\delta &= \left(\mathbf{X}^t A(I_l + \mathrm{i}\delta\eta A)^{-1}\mathbf{X} + b^t\mathbf{X} + c - z\right)^{-1} \in \mathbb{C}^{N\times N}. \end{aligned} \tag{2.4.5}$$

Note that $I_l + \mathrm{i}\delta\eta A$ is invertible for all $\eta > 0$ and $\delta \in [0,1]$ as A is Hermitian. In particular we find

$$(\mathbf{G}_0)_{11} = \mathbf{g} \in \mathbb{C}^{N\times N}. \tag{2.4.6}$$

This justifies why $\mathbf{L}$ is called the linearization of q. The $\delta \neq 0$ case adds an additional regularization, which we use to prove the local law, Proposition 2.4.3, for $\delta = 0$. $\mathbf{G}_\delta$ satisfies the equation

$$\mathbf{I} + (zJ + \mathrm{i}\eta\delta(\mathbf{I} - J) - K_0 + \mathcal{S}[\mathbf{G}_\delta])\mathbf{G}_\delta = \mathbf{D}, \tag{2.4.7}$$

where

$$\mathcal{S}[\mathbf{R}] = \mathbb{E}[(\mathbf{L} - \mathbb{E}[\mathbf{L}])\mathbf{R}(\mathbf{L} - \mathbb{E}[\mathbf{L}])] \tag{2.4.8}$$

is called the self-energy operator and $\mathbf{D}$ is the error term. It is defined by

$$\mathbf{D} := \mathcal{S}[\mathbf{G}_\delta]\mathbf{G}_\delta + (\mathbf{L} - K_0)\mathbf{G}_\delta. \tag{2.4.9}$$

We split the self-energy term into $\widetilde{\Gamma}$ and $\mathcal{S}_\mathrm{o}$, i.e. $\mathcal{S} = \widetilde{\Gamma} + \mathcal{S}_\mathrm{o}$. The first part, $\widetilde{\Gamma}$, depends only on the averaged trace of its argument, i.e. for all $R \in \mathbb{C}^{(l+1)N\times(l+1)N}$ we have

$$\widetilde{\Gamma}[\mathbf{R}] = \Gamma[\underline{\mathbf{R}}]. \tag{2.4.10}$$

The blockwise trace, $\underline{\mathbf{R}}$, was introduced in (2.1.7) and $\Gamma : \mathbb{C}^{(l+1)\times(l+1)} \to \mathbb{C}^{(l+1)\times(l+1)}$ is given by

$$\Gamma\left[\begin{pmatrix} \omega & v^t \\ w & T \end{pmatrix}\right] = \begin{pmatrix} \omega\|b\|^2 + b^t(v+w) + \operatorname{Tr} T & \omega b^t + w^t \\ \omega b + v & \omega I_l \end{pmatrix}, \tag{2.4.11}$$

where $\omega \in \mathbb{C}$, $v, w \in \mathbb{C}^l$ and $T \in \mathbb{C}^{l\times l}$. In (2.4.10) we have made use of our convention to identify $\Gamma[\underline{\mathbf{R}}]$ with $\Gamma[\underline{\mathbf{R}}] \otimes \mathbf{I}_N$

The second term $\mathcal{S}_\mathrm{o} : \mathbb{C}^{(l+1)N\times(l+1)N} \to \mathbb{C}^{(l+1)N\times(l+1)N}$ is given by

$$\mathcal{S}_\mathrm{o}\left[\begin{pmatrix} \boldsymbol{\omega} & \mathbf{V}^t \\ \mathbf{W} & \mathbf{T} \end{pmatrix}\right] = \frac{1}{N}\mathbb{E}\zeta_1^2 \begin{pmatrix} \|b\|^2\boldsymbol{\omega}^{(o)} + b^t(\mathbf{V}^{(o)} + \mathbf{W}^{(o)}) + \operatorname{Tr}_\mathrm{b}\mathbf{T}^{(o)} & \boldsymbol{\omega}^{(o)}b^t + \mathbf{W}^{(o)} \\ \boldsymbol{\omega}^{(o)}b + (\mathbf{V}^t)^{(o)} & I_l \otimes \boldsymbol{\omega}^{(o)} \end{pmatrix}, \tag{2.4.12}$$

where $\boldsymbol{\omega} \in \mathbb{C}^{N\times N}$, $\mathbf{V} = (\mathbf{V}_i)_{i\in[\![l]\!]} \in (\mathbb{C}^{N\times N})^l$, $\mathbf{W} = (\mathbf{W}_i)_{i\in[\![l]\!]} \in (\mathbb{C}^{N\times N})^l$ and $\mathbf{T} \in \mathbb{C}^{lN\times lN}$. For any $\mathbf{R} = (\mathbf{R}_{ij})_{i\in[\![k]\!],j\in[\![n]\!]} \in \mathbb{C}^{kN\times nN}$, $n, k \in \mathbb{N}$, we define $\mathbf{R}^{(o)}$ by $(\mathbf{R}^{(o)})_{ij} := \mathbf{R}_{ji}^t - \operatorname{diag}(\mathbf{R}_{ji}) \in \mathbb{C}^{N\times N}$. For $\mathbf{R} = (\mathbf{R}_{ij})_{i,j\in[\![l]\!]} \in \mathbb{C}^{lN\times lN}$ the block-trace $\operatorname{Tr}_\mathrm{b}$ is defined by

$$\operatorname{Tr}_\mathrm{b}\mathbf{R} = \sum_{i=1}^{l} \mathbf{R}_{ii} \in \mathbb{C}^{N\times N}. \tag{2.4.13}$$

Equation (2.4.7) without the error term and $\mathcal{S}_\mathrm{o}$ is called the Dyson equation (DE) and its solution is denoted by $M_\delta = M_\delta(z) \in \mathbb{C}^{(l+1)\times(l+1)}$, i.e.

$$I + (zJ + \mathrm{i}\eta\delta(I - J) - K_0 + \Gamma[M_\delta])M_\delta = 0 \in \mathbb{C}^{(l+1)\times(l+1)}. \tag{2.4.14}$$

In the next subsection, we will lay out in what sense $\mathbf{G}_\delta$ is close to M_δ. [28, Lemma 2.6] asserts the existence of a unique analytic solution to (2.4.14) for $\delta = 0$ and [46, Theorem 2.1] guarantees a unique solution for $\delta \in (0, 1]$. For $m_\delta \in \mathbb{C}$, $v_\delta, w_\delta \in \mathbb{C}^l$ and $\widehat{M}_\delta \in \mathbb{C}^{l\times l}$ we partition M_δ as

$$M_\delta := \begin{pmatrix} m_\delta & v_\delta^t \\ w_\delta & \widehat{M}_\delta \end{pmatrix}. \tag{2.4.15}$$

Since M_δ solves (2.4.14) it is invertible and by the Schur complement formula its inverse is given by

$$M_\delta^{-1} = \begin{pmatrix} m^{-1} + m^{-2} v_\delta^t \left(\widehat{M}_\delta - w_\delta m_\delta^{-1} v_\delta^t\right)^{-1} w_\delta & -m^{-1} v_\delta^t \left(\widehat{M}_\delta - w_\delta m_\delta^{-1} v_\delta^t\right)^{-1} \\ -m^{-1} \left(\widehat{M}_\delta - w_\delta m_\delta^{-1} v_\delta^t\right)^{-1} w_\delta & \left(\widehat{M}_\delta - w_\delta m_\delta^{-1} v_\delta^t\right)^{-1} \end{pmatrix} \tag{2.4.16}$$

At the same time by (2.4.14) the inverse of M_δ can be expressed as

$$M_\delta^{-1} = \begin{pmatrix} -z - m_\delta b^t b - b^t(v_\delta + w_\delta) - \operatorname{Tr} \widehat{M}_\delta & -m_\delta b^t - w_\delta^t \\ -m_\delta b - v & -A^{-1} - (m_\delta + \mathrm{i}\eta\delta) I_l \end{pmatrix}. \tag{2.4.17}$$

Comparing the two expressions for M_δ^{-1}, we obtain a set of equations for $m_\delta, v_\delta, w_\delta$ and $\widehat{M}_\delta$. Solving them, we find explicit expressions for v_δ, w_δ and $\widehat{M}_\delta$ in terms of m_δ. In other words, $M_\delta = M_\delta[m_\delta]$ can be expressed purely in terms of its (1,1) entry with $M_\delta[x]$ given by

$$M_\delta[x] := \begin{pmatrix} x & -x b^t V_\delta(x) A_\delta \\ -x A_\delta V_\delta(x) b & -A_\delta (I_l + x A_\delta)^{-1} + x A_\delta V_\delta(x) b b^t V_\delta(x) A_\delta \end{pmatrix} \tag{2.4.18}$$

with

$$V_\delta(x) := x(I_l + 2x\widehat{A}_\delta)^{-1}. \tag{2.4.19}$$

Note that we use square brackets to denote M_δ as a function of its (1,1) entry. This is done to avoid confusion with $M_\delta(z)$, which denotes M_δ as a function of the spectral parameter z. The two functions are related by $M_\delta[m_\delta(z)] = M_\delta(z)$. The entry m_δ satisfies the equation

$$-m_\delta^{-1} = z + \gamma_\delta(m_\delta), \tag{2.4.20}$$

where

$$\gamma_\delta(x) := -\operatorname{Tr} A_\delta (I_l + A_\delta x)^{-1} + b^t (V_\delta(x) - V_\delta(x) \widehat{A}_\delta V_\delta(x)) b - c. \tag{2.4.21}$$

and $\widehat{A}_\delta = \frac{1}{2}(A_\delta + A_\delta^t)$. For $\delta = 0$, Equation (2.4.20) corresponds precisely to (2.2.4). From here on we no longer assume that A is invertible and instead, we use the Equations (2.4.4), (2.4.9) and (2.4.18) as the definition for $\mathbf{G}_\delta$, $\mathbf{D}$ and $M_\delta = M_\delta[m_\delta]$ respectively. Note that this is also well defined for (2.4.9) as K_0 and $\mathbf{L}$, defined in (2.4.2) and (2.4.1) respectively, depend on A^{-1}, but $\mathbf{L} - K_0$ does not. The function m_δ is uniquely defined by the following lemma.

Lemma 2.4.1. *Let ρ have a regular edge at $\tau_\pm$ and let $\delta \in (0, 1]$. The function m_δ is uniquely defined by the following criteria around the edge and away from the spectrum.*

1. *There is a $u > 0$ only depending on the coefficients of q such that for all $z \in \mathbb{H}$ with $|z - \tau_\pm| < u$ there is a unique function $m_\delta = m_\delta(z)$ that solves (2.4.20)*

and satisfies

$$\operatorname{Im} m_\delta \sim \begin{cases} \sqrt{\kappa+\eta} & \text{if } E \in \operatorname{supp}(\rho) \\ \frac{\eta}{\sqrt{\kappa+\eta}} & \text{if } E \notin \operatorname{supp}(\rho) \end{cases} \quad \text{as well as} \quad |m_\delta - m_\pm|^2 \sim |z - \tau_\pm|, \tag{2.4.22}$$

with $\kappa = |E - \tau_\pm|$ *and* $m_\pm = \lim_{\eta \searrow 0} m(\tau_\pm + \mathrm{i}\eta)$.

2. *For all* $C > 0$ *there is an* η_0 *such that for all* $z = E + \mathrm{i}\eta \in \mathbb{H}$ *with* $C^{-1} < \operatorname{dist}(E, \operatorname{supp}(\rho)) < C$ *and* $\eta < \eta_0$ *there is a unique function* $m_\delta = m_\delta(z)$ *that solves* (2.4.20) *and satisfies*

$$\operatorname{Im} m_\delta \sim_C \eta \quad \text{as well as} \quad |m_\delta(z) - m(E)| \sim \eta, \tag{2.4.23}$$

with $m(E) = \lim_{\eta \searrow 0} m(E + \mathrm{i}\eta)$.

In each case $M_\delta = M_\delta[m_\delta]$ *is also the unique solution to* (2.4.14) *with* $\operatorname{Im} M_\delta > 0$ *if the coefficient matrix* A *is invertible.*

The lemma is proven in Appendix 2.6.3.

2.4.1 Local law for the linearization

Definition 2.4.2 (Shifted square of a Wigner matrix)**.** *A polynomial* q *as in* (2.1.1) *with* $A \in \mathbb{R}^{l \times l}$, $\operatorname{rank} A = 1$ *and* $b = 0$ *is called a* ***shifted square of a Wigner matrix.***

Remark. *If a polynomial* q *satisfies the above definition, then there is some* $v \in \mathbb{R}^l$ *with* $v \neq 0$ *and*

$$q(\mathbf{X}) = \pm(v^t \mathbf{X})^2 + c. \tag{2.4.24}$$

Here, $\mathbf{W} := \frac{1}{\|v\|} v^t \mathbf{X}$ *is a Wigner matrix, normalised such that* $\mathbb{E}|w_{ij}|^2 = \frac{1}{N}$ *and we have*

$$q(\mathbf{X}) = \pm\|v\|^2 \mathbf{W}^2 + c. \tag{2.4.25}$$

This justifies the terminology in the above definition.

In case q is not a shifted square of a Wigner matrix, we will prove that M_0 and $\mathbf{G}_0$ are close to each other around any regular edge. As m and $\mathbf{g}$ are sub-matrices of M_0 and $\mathbf{G}_0$ respectively, this also implies closeness of $\mathbf{g}$ and m. This result, a local law for the linearization, is presented below in Propositions 2.4.3 and 2.4.4. For q being a shifted square of a Wigner matrix, we will prove the local law, Theorem 2.2.9 and Proposition 2.2.10, directly without the use of a linearization. The reason why we prove this case separately is that the stability operator $\mathscr{L}$, defined below in (2.4.33), does have an additional unstable direction.

Proposition 2.4.3 (Edge local law for the linearization)**.** *Let q be a polynomial of the form (2.1.1) that is not a shifted square of a Wigner matrix and let the corresponding ρ have a regular edge at τ_0. There is a $\kappa_0 > 0$, depending only on the parameters of q, such that for all $\varepsilon, \gamma, D > 0$, $\delta \in \{0, 1\}$ and $z = E + \mathrm{i}\eta \in \mathbb{D}_\gamma^{\kappa_0}$, the isotropic local law*

$$\mathbb{P}\left(|\langle \mathbf{x}, (\mathbf{G}_\delta - M_\delta)\mathbf{y}\rangle| > \|\mathbf{x}\|\|\mathbf{y}\| N^\varepsilon \left(\frac{\operatorname{Im} m}{\sqrt{N\eta}} + \frac{1}{N\eta}\right)\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{2.4.26}$$

holds for all deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{(l+1)N}$. Moreover, the averaged local law

$$\mathbb{P}\left(|\langle \mathbf{B}(\mathbf{G}_\delta - M_\delta)\rangle| > \|\mathbf{B}\| \frac{N^\varepsilon}{N\eta}\right) \lesssim_{\varepsilon,\gamma,D} N^{-D} \tag{2.4.27}$$

holds for all deterministic $\mathbf{B} \in \mathbb{C}^{(l+1)N\times(l+1)N}$. For $E \notin \operatorname{supp}(\rho)$, an improved averaged local law of the form

$$\mathbb{P}\left(|\langle \mathbf{B}(\mathbf{G}_\delta - M_\delta)\rangle| > \|\mathbf{B}\| N^\varepsilon \left(\frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}\right)\right) \lesssim_{\varepsilon,\gamma,D} N^{-D}, \tag{2.4.28}$$

with $\kappa = |E - \tau_0|$, is obtained.

Proposition 2.4.4 (Local law for the linearization away from the spectrum)**.** *Let q be a polynomial of the form (2.1.1) that is not a shifted square of a Wigner matrix. For all $C > 0$ there is an $\eta_0 > 0$, depending only on the coefficients of q, such that an averaged local law of the form*

$$\mathbb{P}\left(|\langle \mathbf{B}(\mathbf{G}_\delta - M_\delta)\rangle| > \|\mathbf{B}\| N^\varepsilon \left(\frac{1}{N} + \frac{1}{(N\eta)^2}\right)\right) \lesssim_{\varepsilon,\gamma,D,C} N^{-D} \tag{2.4.29}$$

holds true for all deterministic $\mathbf{B} \in \mathbb{C}^{(l+1)N\times(l+1)N}$, $\varepsilon, \gamma, D > 0$ and $z \in \mathbb{G}_\gamma^{C,\eta_0}$.

To obtain Theorem 2.2.9, we only require the $\delta = 0$ case, but to prove it we will require the $\delta = 1$ case, thus we state both cases together in the proposition. The proof of Proposition 2.4.3 has two major ingredients. For one we show that the error term $\mathbf{D}$ in (2.4.7) is indeed small, this is done in Proposition 2.4.5, which we import from [29, Theorem 4.1] and adjust to our setting. Additionally, we need to prove that (2.4.14) is stable under small perturbations. This is done in Proposition 2.4.6.

Proposition 2.4.5. *Let $\varepsilon > 0$, $p \in \mathbb{N}$ and $\delta \in [0, 1]$. Then there is a $c > 0$ such that*

$$\|\mathbf{D}\|_p \lesssim_{p,\varepsilon} N^\varepsilon \sqrt{\frac{\|\mathbf{G}_\delta \mathbf{G}_\delta^*\|_{p_0}}{N}} (1 + \|\mathbf{G}_\delta\|_{p_0})^c \left(1 + N^{-\frac{1}{4}}\|\mathbf{G}_\delta\|_{p_0}\right)^{cp} \tag{2.4.30}$$

and

$$\|\mathbf{D}\|_p^{\mathrm{av}} \lesssim_{p,\varepsilon} N^\varepsilon \frac{\|\mathbf{G}_\delta^* \mathbf{G}_\delta\|_{p_0}}{N} (1 + \|\mathbf{G}_\delta\|_{p_0})^c \left(1 + N^{-\frac{1}{4}}\|\mathbf{G}_\delta\|_{p_0}\right)^{cp} \tag{2.4.31}$$

where we defined $p_0 = c\frac{p^4}{\varepsilon}$.

Proof. First, consider the case of A being invertible. Then $\mathbf{G}_\delta = (\mathbf{L} - zJ - \mathrm{i}\eta\delta(\mathbf{I} - J))^{-1}$ and our proof follows the proof of [29, Theorem 4.1] line by line with the exception that the Ward identities, (51a) and (51b), do not apply for $\delta = 0$. Thus, we cannot replace the $\mathbf{G}_\delta^*\mathbf{G}_\delta$ terms by $\eta^{-1}\operatorname{Im}\mathbf{G}_\delta$ and we are instead left with the upper bounds (2.4.30) and (2.4.31).

Now consider A being non-invertible. Then $A + \varepsilon$ is invertible for all $\varepsilon \in (0, u)$ for some $u \sim 1$. We denote the $\mathbf{G}_\delta$ associated with $A + \varepsilon$ by $\mathbf{G}_\delta^\varepsilon$ and we have $\lim_{\varepsilon\to 0}\mathbf{G}_\delta^\varepsilon = \mathbf{G}_\delta$. We thus only need to prove that the constants in (2.4.30) and (2.4.31) are uniform in ε to obtain the proposition in this case. This is non-trivial as [29, Theorem 4.1] states as a condition that $\mathbb{E}[\mathbf{L}]$ is bounded and this is clearly not the case in the $\varepsilon \to 0$ limit. The assumption is only used, however, to ensure that $\mathbf{G}_\delta^*\mathbf{G}_\delta$ satisfies the lower bound $\|\mathbf{G}_\delta^*\mathbf{G}_\delta\|_{p_0} \gtrsim 1$. In our case, this follows instead from $\|\mathbf{G}_\delta^*\mathbf{G}_\delta\|_{p_0} \geq \|\mathbf{g}_\delta^*\mathbf{g}_\delta\|_{p_0} \gtrsim_{p_0} \|\mathbf{g}^*\mathbf{g}\|_{p_0}$. Since $\mathbf{g}$ is the resolvent of $q(\mathbf{X})$ we have $\|\mathbf{g}^*\mathbf{g}\|_q \geq \mathbb{E}[\|q(\mathbf{X})\|^{-1}] \gtrsim 1$ uniformly in z for bounded z. Here, the last inequality holds true since $q(\mathbf{X})$ satisfies the inequality

$$\|q(\mathbf{X})\| \leq \sum_{i,j=1}^{l} |A_{ij}|\|\mathbf{X}_i\|\|\mathbf{X}_j\| + \sum_{i=1}^{l} |b_i|\|\mathbf{X}_i\| + |c| \tag{2.4.32}$$

and $\mathbb{E}[\|\mathbf{X}_i\|] \lesssim 1$ for all $i \in [\![l]\!]$ since the $\mathbf{X}_i$ are Wigner matrices. Therefore the proposition also holds for non-invertible A. □

The following result concerns the stability of the Dyson equation. The stability operator $\mathscr{L} : \mathbb{C}^{(l+1)N\times(l+1)N} \to \mathbb{C}^{(l+1)N\times(l+1)N}$ is given by

$$\mathscr{L}[\mathbf{R}] = \mathbf{R} - M_\delta\mathcal{S}[\mathbf{R}]M_\delta \tag{2.4.33}$$

and we will prove

Proposition 2.4.6 (Control of $\mathscr{L}$)**.** *Let q be a polynomial of the form (2.1.1) that is not a shifted square of a Wigner matrix and let the corresponding ρ have a regular edge at τ_0. There exists a $u \sim 1$ such that for all $z = E + \mathrm{i}\eta \in \mathbb{H}$ with $|z - \tau_0| < u$ and $\delta \in [0, 1]$ there exists an eigenvalue β with corresponding left and right eigenvectors $L, B \in \mathbb{C}^{(l+1)\times(l+1)}$ of $\mathscr{L}$ such that*

$$\begin{gathered}\|\mathscr{L}^{-1}\|_{\mathrm{sp}} \sim (\kappa+\eta)^{-\frac{1}{2}}, \quad \|(\mathbb{1} - \mathscr{P})\mathscr{L}^{-1}\|_{\mathrm{sp}} \lesssim 1, \quad |\beta| \sim (\kappa+\eta)^{\frac{1}{2}}, \\ |\langle L, B\rangle| \sim 1, \quad \|L\| + \|B\| \sim 1, \quad |\langle L, M_\delta\mathcal{S}[B]B\rangle| \sim 1,\end{gathered} \tag{2.4.34}$$

with $\mathbb{1}$ being the identity operator, $\mathscr{P}$ being the spectral projection onto B, i.e.

$$\mathscr{P} = (\langle L \otimes \mathbf{I}_N, B \otimes \mathbf{I}_N\rangle)^{-1}\langle L \otimes \mathbf{I}_N, \cdot\rangle(B \otimes \mathbf{I}_N) \tag{2.4.35}$$

and $\kappa = |E - \tau_0|$.

Furthermore, for any $C > 0$ and $E = \operatorname{Re} z$ with $C^{-1} < \operatorname{dist}(E, \operatorname{supp}(\rho)) < C$, there is an $\eta_0 > 0$ such that we have

$$\|\mathscr{L}^{-1}\|_{\mathrm{sp}} \sim_C 1 \tag{2.4.36}$$

uniformly for all $\eta \leq \eta_0$.

The proof will be given in Section 2.4.2.

Remark. *B and L being right and left eigenvectors of $\mathscr{L}$ with eigenvalue β is understood in the sense of*

$$\mathscr{L}[B] = \beta B \quad \textit{and} \quad \mathscr{L}^*[L] = \bar{\beta} L. \tag{2.4.37}$$

Here, we used the notation $R = R \otimes \mathbf{I}_N \in \mathbb{C}^{(l+1)N \times (l+1)N}$ introduced in (2.1.3). The adjoint is defined with respect to the scalar product $\langle \mathbf{R}, \mathbf{T} \rangle = \langle \mathbf{R}^* \mathbf{T} \rangle$.

Corollary 2.3.3, as well as Propositions 2.4.5 and 2.4.6 are the main ingredients to Proposition 2.4.3 and are in fact sufficient for $\delta = 1$. For $\delta = 0$ however, extra care is needed as the Ward identity for resolvents $\mathbf{G}$,

$$\mathbf{G}\mathbf{G}^* = \frac{\operatorname{Im} \mathbf{G}}{\eta}, \tag{2.4.38}$$

does not translate to generalized resolvents. Instead, we will estimate $\mathbf{G}_0 \mathbf{G}_0^*$ by $\mathbf{G}_1 \mathbf{G}_1^*$, which allows us to obtain the local law for $\delta = 0$ from the $\delta = 1$ case. The proof of Proposition 2.4.3 will be given in Section 2.5 and follows the general strategy from [3], modified to accommodate for the lack of a Ward identity.

2.4.2 Proof of Proposition 2.4.6

Proof of Proposition 2.4.6. We split the stability operator into $\mathscr{L} = \mathscr{L}^{(0)} + \mathscr{L}^{(1)}$ with

$$\mathscr{L}^{(0)}[\mathbf{R}] := \mathscr{L}[\underline{\mathbf{R}}] \quad \text{and} \quad \mathscr{L}^{(1)}[\mathbf{R}] := \mathscr{L}[\mathbf{R} - \underline{\mathbf{R}}]. \tag{2.4.39}$$

By (2.4.11) and (2.4.12), we have $\tilde{\Gamma}[\mathbf{R} - \underline{\mathbf{R}}] = \mathcal{S}_{\mathrm{o}}[\underline{\mathbf{R}}] = 0$. Thus we have

$$\mathscr{L}^{(0)}[\mathbf{R}] = \mathcal{L}[\underline{\mathbf{R}}] \otimes \mathbf{I}_N \quad \text{and} \quad \mathscr{L}^{(1)}[\mathbf{R}] = \mathbf{R} - \underline{\mathbf{R}} - M_\delta \mathcal{S}_{\mathrm{o}}[\mathbf{R}] M_\delta \tag{2.4.40}$$

with $\mathcal{L} : \mathbb{C}^{(l+1)\times(l+1)} \to \mathbb{C}^{(l+1)\times(l+1)}$ defined as

$$\mathcal{L}[R] := R - M_\delta \Gamma[R] M_\delta. \tag{2.4.41}$$

The image of $\mathscr{L}^{(0)}$ is given by

$$\{\mathbf{R} \in \mathbb{C}^{(l+1)N \times (l+1)N} : \underline{\mathbf{R}} = \mathbf{R}\} =: \mathcal{U} \tag{2.4.42}$$

and its kernel is given by $\mathcal{U}^\perp$, the orthogonal complement of $\mathcal{U}$. At the same time the image of $\mathscr{L}^{(1)}$ is contained in $\mathcal{U}^\perp$ and $\mathcal{U}$ is contained in the kernel of $\mathscr{L}^{(1)}$. That is, $\mathscr{L}$ decomposes into $\mathscr{L}^{(0)}$ acting on $\mathcal{U}$ and $\mathscr{L}^{(1)}$ acting on its orthogonal complement.

The behaviour of $\mathcal{L}$ is summarized in the following lemma.

Lemma 2.4.7. *Let q be a polynomial of the form (2.1.1) that is not a shifted square of a Wigner matrix and let the corresponding ρ have a regular edge at τ_0. There exists an $u \sim 1$ such that for all $z = E + \mathrm{i}\eta \in \mathbb{H}$ with $|z - \tau_0| < u$ and $\delta \in [0,1]$ there exists an eigenvalue β with corresponding normalized left and right eigenvectors L and B of $\mathcal{L}$ such that*

$$\begin{aligned} \|\mathcal{L}^{-1}\|_{\mathrm{sp}} \sim (\kappa+\eta)^{-\frac{1}{2}}, \quad \|(\mathbb{1}-\mathcal{P})\mathcal{L}^{-1}\|_{\mathrm{sp}} \lesssim 1, \quad |\beta| \sim (\kappa+\eta)^{\frac{1}{2}}, \\ |\langle L, B\rangle| \sim 1, \quad |\langle L, M_\delta \Gamma[B] B\rangle| \sim 1, \end{aligned} \tag{2.4.43}$$

with $\mathcal{P}$ being the spectral projection onto B, i.e. $\mathcal{P} = (\langle L, B\rangle)^{-1}\langle L, \cdot\rangle B$ and $\kappa = |E - \tau_+|$.

Furthermore, for any $C > 0$ and $E = \operatorname{Re} z$ with $C^{-1} \leq \operatorname{dist}(E, \operatorname{supp}(\rho)) \leq C$ there is an $\eta_0 > 0$ such that we have

$$\|\mathcal{L}^{-1}\|_{\mathrm{sp}} \sim_C 1 \tag{2.4.44}$$

uniformly for all $\eta \leq \eta_0$.

The proof of Lemma 2.4.7 is deferred to the end of the section. For $\mathcal{S}_{\mathrm{o}}$ we find

$$\|\mathcal{S}_{\mathrm{o}}[\mathbf{R}]\|_{\mathrm{hs}} \lesssim \frac{1}{N}\|\mathbf{R}\|_{\mathrm{hs}} \tag{2.4.45}$$

for all $\mathbf{R} \in \mathbb{C}^{(l+1)N \times (l+1)N}$. Thus $\mathcal{S}_{\mathrm{o}}$ is bounded by

$$\|\mathcal{S}_{\mathrm{o}}\|_{\mathrm{sp}} \lesssim \frac{1}{N}. \tag{2.4.46}$$

By Corollary 2.3.3, Lemma 2.4.1 and (2.4.18) we have $\|M_\delta\| \lesssim 1$ for all z such that $|z - \tau_0| \leq u$ and some $u > 0$. Combined with (2.4.46) it follows that there is a $C > 0$ such that

$$\|\mathscr{L}^{(1)}\|_{\mathrm{sp}} \leq 1 + CN^{-1} \quad \text{and} \quad \operatorname{Spec}\left(\mathscr{L}^{(1)}|_{\mathcal{U}^\perp}\right) \subset B_{CN^{-1}}(1), \tag{2.4.47}$$

where $B_\varepsilon(x)$ denotes the ε neighborhood of x. Thus for sufficiently large N the smallest eigenvalue of $\mathscr{L}$ equals that of $\mathcal{L}$ and the corresponding left and right eigenvectors of $\mathscr{L}$ are given by $L \otimes \mathbf{I}_N$ and $B \otimes \mathbf{I}_N$. The norm of the inverse of $\mathscr{L}$ is bounded by

$$\|\mathscr{L}^{-1}\|_{\mathrm{sp}} \leq \max\{1 + CN^{-1}, \|\mathcal{L}^{-1}\|_{\mathrm{sp}}\} \tag{2.4.48}$$

and

$$\|(\mathbb{1} - \mathscr{P})\mathscr{L}^{-1}\|_{\mathrm{sp}} \leq \max\{1 + CN^{-1}, \|(\mathbb{1} - \mathcal{P})\mathcal{L}^{-1}\|_{\mathrm{sp}}\}, \tag{2.4.49}$$

completing the proof of Proposition 2.4.6 □

Proof of Lemma 2.4.7. First, we prove that $\mathcal{L}$ has exactly one vanishing eigenvalue at τ_0, and from there on we conclude the proof with the help of a perturbative argument. We define

$$\mathcal{C}_J[R] = JRJ, \tag{2.4.50}$$

i.e. $\mathcal{C}_J$ is the projection onto the $(1,1)$ entry and in particular $\mathcal{C}_J[M_0] = mJ$. We also set $\mathcal{C}_J^\perp R := R - \mathcal{C}_J R$ and $\widetilde{\mathbb{C}}^{(l+1)\times(l+1)} := \mathrm{Image}\,\mathcal{C}_J^\perp$. For any $R \in \mathbb{C}^{(l+1)\times(l+1)}$ we denote $r := R_{11}$ and $\widetilde{R} := \mathcal{C}_J^\perp[R]$, i.e. $R = rJ + \widetilde{R}$. Let T_z be the matrix

$$T_z = \begin{pmatrix} 1 & 0 \\ 0 & zA \end{pmatrix} \in \mathbb{C}^{(l+1)\times(l+1)} \tag{2.4.51}$$

as well as

$$F : \begin{cases} \mathbb{C} \times \widetilde{\mathbb{C}}^{(l+1)\times(l+1)} \times \mathbb{H} & \to \mathbb{C}^{(l+1)\times(l+1)} \\ (r, \widetilde{R}, z) & \mapsto rJ + \widetilde{R} + T_z(z + \Gamma[rJ + \widetilde{R}]T_z)^{-1}. \end{cases} \tag{2.4.52}$$

Then $F(m, \widetilde{M}, z) = 0$ for $M = M_0[m]$ defined in (2.4.18). For $\delta = 0$, the stability operator $\mathcal{L}$ is the derivative of F in the sense that

$$\mathcal{L}[R] = D_R F(m, \widetilde{M}, z), \tag{2.4.53}$$

where $D_R F = \frac{\mathrm{d}}{\mathrm{d}\varepsilon} F(m + \varepsilon r, \widetilde{M} + \varepsilon\widetilde{R}, z)|_{\varepsilon=0}$ is the directional derivative of F in the direction R. We first consider the case when A is invertible. The case of non-invertible A will be treated afterwards. We define $\mathcal{B} := \mathcal{C}_{M^{-1}}\mathcal{L}$ on $z \in \mathbb{R} \setminus \mathrm{supp}(\rho)$, where $m(z)$ is defined as the unique analytical continuation to $\mathbb{C} \setminus \mathrm{supp}(\rho)$. Since $F(m, \widetilde{M}, z) = 0$ we have

$$\begin{aligned} \mathcal{B}[R] &= D_R[\mathcal{C}_{M^{-1}}F] \\ &= D_{rJ}\mathcal{C}_J[\mathcal{C}_{M^{-1}}F] + D_{\widetilde{R}}\mathcal{C}_J[\mathcal{C}_{M^{-1}}F] + D_{rJ}\mathcal{C}_J^\perp[\mathcal{C}_{M^{-1}}F] + D_{\widetilde{R}}\mathcal{C}_J^\perp[\mathcal{C}_{M^{-1}}F], \end{aligned} \tag{2.4.54}$$

where we used the linearity of the derivatives as well as the linearity of $\mathcal{B}$ in the second equality and we omitted the arguments of F. The above equation decomposes $\mathcal{B}$ into a two-by-two block operator with diagonal blocks $\mathcal{B}_{11}[rJ] := D_{rJ}\mathcal{C}_J[\mathcal{C}_{M^{-1}}F]$, $\mathcal{B}_{22}[\widetilde{R}] := D_{\widetilde{R}}\mathcal{C}_J^\perp[\mathcal{C}_{M^{-1}}F]$ and off-diagonal blocks $\mathcal{B}_{12}[\widetilde{R}] := D_{\widetilde{R}}\mathcal{C}_J[\mathcal{C}_{M^{-1}}F]$, $\mathcal{B}_{21}[rJ] := D_{rJ}\mathcal{C}_J^\perp[\mathcal{C}_{M^{-1}}F]$.

A tedious but straightforward calculation shows that $\mathcal{B}_{22}$ is invertible on the image of $\mathcal{C}_J^\perp$ and its inverse is given by

$$(\mathcal{B}_{22})^{-1}\left[\begin{pmatrix} 0 & r_{12}^t \\ r_{21} & \widehat{R} \end{pmatrix}\right] = \begin{pmatrix} 0 & h_{12}^t \\ h_{21} & \widehat{H} \end{pmatrix} \tag{2.4.55}$$

with

$$\begin{aligned} h_{12} =& -A^t V_0(mA(r_{21} - m^{-1}\widehat{R}w_0) + (1+mA)(r_{12} - m^{-1}\widehat{R}^t v_0)) \\ h_{21} =& -AV_0(mA^t(r_{12} - m^{-1}\widehat{R}^t v_0) + (1+mA^t)(r_{21} - m^{-1}\widehat{R}w_0)) \\ \widehat{H} =& \frac{A}{1+mA}\widehat{R}\frac{A}{1+mA} \\ & - AV_0(mA^t(r_{12} - m^{-1}\widehat{R}^t v_0) + (1+mA^t)(r_{21} - m^{-1}\widehat{R}w_0))m^{-1}v^t \\ & - wm^{-1}((r_{21}^t - m^{-1}w_0^t\widehat{R})mA^t + (r_{12}^t - m^{-1}v_0^t\widehat{R})(1+mA^t))V_0 A, \end{aligned} \tag{2.4.56}$$

where $V_0 = V_0(m)$ was introduced in (2.4.19) and v_0 and w_0 where defined in (2.4.15). Their explicit form in terms of $m = m_0$ is given in (2.4.18). From $F = 0$, we also have $\mathcal{C}_J^\perp \mathcal{C}_{M^{-1}} F = 0$ and both $\widetilde{M} = \widetilde{M}_0$ and z are uniquely defined by m (see (2.4.18) and (2.2.4)). Therefore the total derivative of $\mathcal{C}_J^\perp \mathcal{C}_{M^{-1}} F$ with respect to m is well defined and vanishes as well, i.e.

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}m} \mathcal{C}_J^\perp \mathcal{C}_{M^{-1}} F &= \mathcal{C}_J^\perp D_J \mathcal{C}_{M^{-1}} F + \mathcal{C}_J^\perp D_{\widetilde{M}'} \mathcal{C}_{M^{-1}} F + z'(m) \frac{\partial}{\partial z} \mathcal{C}_J^\perp \mathcal{C}_{M^{-1}} F \\ &= \mathcal{B}_{21}[J] + \mathcal{B}_{22}[\widetilde{M}'] \end{aligned} \tag{2.4.57}$$

where $\widetilde{M}' := \frac{\partial \widetilde{M}}{\partial m}$ and z' denote the derivative of $\mathcal{C}_J^\perp[M[m]]$ and z with respect to m. In the last step we also used $\frac{\partial}{\partial z} \mathcal{C}_J^\perp \mathcal{C}_{M^{-1}} F = 0$, which follows from (2.4.14). Since $\mathcal{B}_{22}$ is invertible on its image, (2.4.57) is equivalent to

$$(\mathcal{B}_{22})^{-1} \mathcal{B}_{21}[J] = -\widetilde{M}'. \tag{2.4.58}$$

Therefore the Schur complement of $\mathcal{B}_{22}$ is given by

$$(\mathcal{B}_{11} - \mathcal{B}_{12}(\mathcal{B}_{22})^{-1}\mathcal{B}_{21})[J] = \frac{\partial}{\partial m} C_J[\mathcal{C}_{M^{-1}} F] + D_{\widetilde{M}'} \mathcal{C}_J[\mathcal{C}_{M^{-1}} F]. \tag{2.4.59}$$

In other words, the Schur complement of $\mathcal{B}_{22}$ is the total derivative of $\mathcal{C}_J[\mathcal{C}_{M^{-1}} F]$ with respect to m for fixed z. Calculating it, we find

$$(\mathcal{B}_{11} - \mathcal{B}_{12}(\mathcal{B}_{22})^{-1}\mathcal{B}_{21})[J] = \left(\frac{1}{m^2} - \gamma'(m) \right) J = h(m) J, \tag{2.4.60}$$

where h was introduced in Definition 2.3.4. By Lemma 2.3.5, Lemma 2.3.6 and Lemma 2.6.3 we have $h(m) \neq 0$ for all $z \in \mathbb{R} \setminus \mathrm{supp}(\rho)$. Therefore, the Schur complement of $\mathcal{B}_{22}$ is invertible on the image of $\mathcal{C}_J$ for all $z \in \mathbb{R} \setminus \mathrm{supp}(\rho)$. Since both $\mathcal{B}_{22}$ and its Schur complement are invertible on their respective images for all $z \in \mathbb{R} \setminus \mathrm{supp}(\rho)$, the operator $\mathcal{B}$ for all $z \in \mathbb{R} \setminus \mathrm{supp}(\rho)$ is also invertible with its inverse given by the Schur complement formula

$$\begin{aligned} &\mathcal{B}^{-1}[R] \\ =& \left(\mathbb{1} - (\mathcal{B}_{22})^{-1}\mathcal{B}_{21}\right) \left(\mathcal{B}_{11} - \mathcal{B}_{12}(\mathcal{B}_{22})^{-1}\mathcal{B}_{21}\right)^{-1} \left[rJ - \mathcal{B}_{12}(\mathcal{B}_{22})^{-1}[\widetilde{R}] \right] + (\mathcal{B}_{22})^{-1}[\widetilde{R}] \\ =& h(m)^{-1} \left(r - \left\langle J, \mathcal{B}_{12}(\mathcal{B}_{22})^{-1}[\widetilde{R}] \right\rangle \right) M' + (\mathcal{B}_{22})^{-1}[\widetilde{R}], \end{aligned} \tag{2.4.61}$$

where $M' = J + \widetilde{M}'$ is the derivative of M with respect to m and the inverse of the Schur complement is acting on $\mathrm{span}(J)$. Therefore, $\mathcal{L}$ is also invertible with inverse $\mathcal{L}^{-1} = \mathcal{B}^{-1}\mathcal{C}_{M^{-1}}$ for $z \in \mathbb{R} \setminus \mathrm{supp}(\rho)$. Now let A be non-invertible. Then, $A + \varepsilon$ is invertible for all $\varepsilon \in (0, u]$ for some $u > 0$. We use an ε subscript to denote quantities with A being replaced by $A^\varepsilon := A + \varepsilon$, e.g. $\mathcal{L}^\varepsilon$, $(\mathcal{B}^{-1})^\varepsilon$ etc.

We have $\mathcal{L}^\varepsilon(\mathcal{B}^{-1}\mathcal{C}_{M^{-1}})^\varepsilon = \mathbb{1}$ for all $\varepsilon \in (0,u]$ and both the limits $\lim_{\varepsilon\searrow 0}\mathcal{L}^\varepsilon$ and $\lim_{\varepsilon\searrow 0}(\mathcal{B}^{-1}\mathcal{C}_{M^{-1}})^\varepsilon$ exist. Indeed, $\mathcal{L}^\varepsilon$ is the derivative of F^ε, which is smooth at $\varepsilon = 0$, proving the existence of $\lim_{\varepsilon\searrow 0}\mathcal{L}^\varepsilon$. Furthermore we explicitly calculate the expression for $(\mathcal{B}^{-1}\mathcal{C}_{M^{-1}})^\varepsilon$ in terms of A^ε and m^ε by using (2.4.17), (2.4.55), (2.4.56) and (2.4.61). In the resulting expression, $(A^\varepsilon)^{-1}$, appearing in (2.4.17), cancels and therefore the limit $\lim_{\varepsilon\searrow 0}(\mathcal{B}^{-1}\mathcal{C}_{M^{-1}})^\varepsilon$ exists. We leave the details to the reader. Therefore, $\mathcal{L}$ is also invertible for $z \in \mathbb{R}\setminus\operatorname{supp}(\rho)$ and we have $\mathcal{L}^{-1} = \lim_{\varepsilon\searrow 0}(\mathcal{B}^{-1}\mathcal{C}_{M^{-1}})^\varepsilon$.

For any $C > 0$ the norm of the operator $\mathcal{L}^{-1}$ is bounded on the compact subset $C^{-1} \le \operatorname{dist}(z,\operatorname{supp}(\rho)) \le C$ of $\mathbb{R}\setminus\operatorname{supp}(\rho)$, i.e. $\|\mathcal{L}^{-1}\|_{\mathrm{sp}} \sim_C 1$. By continuity, there is some $\eta_0 > 0$, depending on C, such that $\|\mathscr{L}^{-1}\|_{\mathrm{sp}} \sim_C 1$ still holds for $z = E + \mathrm{i}\eta \in \mathbb{H}$ with $C^{-1} \le \operatorname{dist}(E,\operatorname{supp}(\rho)) \le C$ and $\eta \le \eta_0$ and we have therefore shown (2.4.44).

Let $\mathcal{Q}_{M'}$ be the projection onto the orthogonal complement of M'. By (2.4.55), (2.4.56) and (2.4.61) we have

$$\begin{aligned}\|\mathcal{Q}_{M'}\mathscr{L}^{-1}\|_{\mathrm{sp}} &\lesssim \left\|\frac{1}{1+Am}\right\| + \|V_0\| \\ &\lesssim \operatorname{dist}\left(-1,\operatorname{Spec}(mA)\right) + \operatorname{dist}\left(-\frac{1}{2},\operatorname{Spec}(m\widehat{A})\right)\end{aligned} \tag{2.4.62}$$

uniformly in $z \in \mathbb{R}\setminus\operatorname{supp}(\rho)$. At $z = \tau_0$ we have $h(m) = 0$ and thus

$$\left\|\frac{(mA)^2}{(1+mA)^2}\right\| \le \operatorname{Tr}\left(\frac{(mA)^2}{(1+mA)^2}\right) + b^t\left(\frac{V_0}{m}\right)^3 b = 1. \tag{2.4.63}$$

In the above equation, we have equality if and only if $\operatorname{rank} A = 1$ and $b = 0$. Therefore, we have $\operatorname{Spec}(mA) \subset [-\frac{1}{2},\infty)$ and $-\frac{1}{2} \in \operatorname{Spec}(A)$ if and only if $\operatorname{rank} A = 1$ and $b = 0$. By Lemma 2.6.1, we have $\min\operatorname{Spec}(m\widehat{A}) = \min\operatorname{Spec}(mA)$ if and only if $mA \in \mathbb{R}^{l\times l}$. Therefore we have $-\frac{1}{2} \in \operatorname{Spec}(m\widehat{A})$ if and only if $\operatorname{rank} A = 1$, $A \in \mathbb{R}^{l\times l}$ and $b = 0$. That is, $-\frac{1}{2} \in \operatorname{Spec}(m\widehat{A})$ at τ_0 if and only if A is a shifted square of a Wigner matrix. If A is not a shifted square of a Wigner matrix, then (2.4.62) is uniformly bounded in some neighbourhood of τ_0. Thus $\mathcal{L}$ can at most have one vanishing eigenvalue at τ_0. Indeed let

$$H(m,z) := F(m,\widetilde{M}[m],z) = M[m] + T_z(z + \Gamma[M[m]]T_z)^{-1},$$

where $M[m]$ is given by the right hand side of (2.4.18). The spectral parameter z is also uniquely defined by m (see (2.2.4)), i.e. $H(m,z(m)) = 0$ with

$$z(m) := -\frac{1}{m} - \gamma(m).$$

Taking the derivative with respect to m we find

$$0 = \frac{\mathrm{d}}{\mathrm{d}m}H(m,z(m)) = \partial_m H + \frac{\mathrm{d}z}{\mathrm{d}m}\partial_z H. \tag{2.4.64}$$

By Corollary 2.3.3 the derivative $\frac{\mathrm{d}z}{\mathrm{d}m}$ vanishes at the edge. Therefore at $z = \tau_0$ the above equation simplifies to

$$\partial_m H = 0. \tag{2.4.65}$$

Calculating this we obtain

$$0 = M' - M\Gamma[M']M = \mathcal{L}[M'], \tag{2.4.66}$$

with $M' = M'[m]$ the derivative M with respect to m and we used $H = 0$. Therefore $B = M'[m]$ is the critical right eigenvector at τ_0. Next, we obtain the critical left eigenvector L. The adjoint of the stability operator $\mathcal{L}$ is

$$\mathcal{L}^*[R] = R - \Gamma[M^* R M^*]. \tag{2.4.67}$$

At any regular edge, we have $\operatorname{Im} M = 0$, i.e. $M = M^*$. Therefore

$$0 = \Gamma[\mathcal{L}[B]] = \Gamma[B] - \Gamma[\mathcal{C}_M[\Gamma[B]]] = \mathcal{L}^*[\Gamma[B]] \tag{2.4.68}$$

at $z = \tau_0$. Thus the critical left eigenvector is given by

$$L = \Gamma[B]. \tag{2.4.69}$$

Note that $L \neq 0$ since $L_{ii} = B_{11} = 1$ for all $1 < i \leq l + 1$ by the expression for Γ in (2.4.11). As L and B belong to the same non-degenerate eigenvalue they cannot be orthogonal and since $\mathcal{L}$ does not depend on N they satisfy

$$|\langle L, B\rangle| \sim 1 \tag{2.4.70}$$

at τ_0. Equation (2.4.70) is also satisfied in a u neighbourhood of τ_0, with $u \sim 1$, since L and B vary continuously in z.

To obtain $|\langle L, M\Gamma[B]B\rangle|$ we calculate the second total derivative

$$\begin{aligned} 0 &= \frac{\mathrm{d}^2}{\mathrm{d}m^2} H(m, z(m)) \\ &= \partial_m^2 H(m, z(m)) + \frac{\mathrm{d}^2 z}{\mathrm{d}m^2} \partial_z H(m, z(m)) \\ &\quad + \frac{\mathrm{d}z}{\mathrm{d}m} \left(\frac{\mathrm{d}z}{\mathrm{d}m} \partial_z^2 H(m, z(m)) + 2\partial_m \partial_z H(m, z(m)) \right). \end{aligned} \tag{2.4.71}$$

At the edge, $\frac{\mathrm{d}z}{\mathrm{d}m}$ once again vanishes, whereas $\frac{\mathrm{d}^2 z}{\mathrm{d}m^2} = c_0 \neq 0$ does not by Corollary 2.3.3 and Lemma 2.4.1. Thus calculating the derivatives, we arrive at

$$\begin{aligned} 0 &= -c_0 MJM + M'' - M\Gamma[M'']M - 2M\Gamma[M']M\Gamma[M']M \\ &= -c_0 MJM + \mathcal{L}[M''] - 2M\Gamma[M']M', \end{aligned} \tag{2.4.72}$$

where we used $H = 0$ multiple times and (2.4.66) in the last step. Next, we solve for the last term, use $M' = B$ and take an inner product with L:

$$\begin{aligned} \langle L, M\Gamma[B]B\rangle &= \frac{1}{2} \left(\langle L, \mathcal{L}[M'']\rangle - c_0 \langle L, MJM\rangle \right) \\ &= \frac{1}{2} \left(\langle \mathcal{L}^*[L], M''\rangle - c_0 \langle MLM, J\rangle \right) = -\frac{c_0}{2} B_{11} \neq 0. \end{aligned} \tag{2.4.73}$$

In the last equality, $MLM = M\Gamma[B]M = B$ was used. By continuity the relation $|\langle L, M\Gamma[B]B\rangle| \sim 1$ also holds in a u neighbourhood of τ_0 with $u \sim 1$. At τ_0, the operator $\mathcal{L}$ is independent of δ and therefore its vanishing eigenvectors L and B are as well. Consequently, (2.4.70) and (2.4.73) also hold for all $\delta \in [0,1]$ and as both expressions are also continuous in z for all δ, they also hold in some order one neighborhood of τ_0.

Next we study how $\mathcal{L}$ varies in z around τ_0 for arbitrary $\delta \in [0,1]$ and we make the dependence explicit by writing $\mathcal{L} = \mathcal{L}^z$, $M_\delta = M_\delta^z$, etc. and define $\mathcal{E}^z := \mathcal{L}^z - \mathcal{L}^{\tau_0}$. As the eigenvalues of $\mathcal{L}^z$ depend continuously on z, there is a $u > 0$ such that $\mathcal{L}^z$ has an isolated small eigenvalue β^z for all $|z - \tau_0| < u$. By perturbation theory β_z is given by

$$\beta^z = (\langle L^{\tau_0}, B^{\tau_0}\rangle)^{-1}\langle L^{\tau_0}, \mathscr{E}^z(B^{\tau_0})\rangle + \mathcal{O}(\|\mathscr{E}^z\|_{\mathrm{sp}}^2). \tag{2.4.74}$$

To estimate the right hand side, we first evaluate $\mathcal{E}^z[B^{\tau_0}]$ and find

$$\begin{aligned}\mathcal{E}^z[B^{\tau_0}] &= \mathcal{L}^z[B^{\tau_0}] - \mathcal{L}^{\tau_0}[B^{\tau_0}] = \mathcal{C}_{M_\delta^z}[L^{\tau_0}] - C_{M_\delta^{\tau_0}}[L^{\tau_0}] \\ &= M_\delta^{\tau_0} L^{\tau_0}(M_\delta^{\tau_0} - M_\delta^z) + (M_\delta^{\tau_0} - M_\delta^z)L^{\tau_0} M_\delta^z.\end{aligned} \tag{2.4.75}$$

We take the scalar product with L^{τ_0} and use the cyclic invariance of the trace to obtain

$$\begin{aligned}|\langle L^{\tau_0}, \mathcal{E}^z[B^{\tau_0}]\rangle| &= |\langle L^{\tau_0}(M_\delta^{\tau_0} + M_\delta^z)L^{\tau_0}(M_\delta^{\tau_0} - M_\delta^z)\rangle| \\ &= 2|\langle L^{\tau_0} M_\delta^{\tau_0} L^{\tau_0}(M_\delta^{\tau_0} - M_\delta^z)\rangle| + \mathcal{O}(|m_\delta^z - m_\delta^{\tau_0}|^2) \\ &= 2|\langle L^{\tau_0} M_\delta^{\tau_0} L^{\tau_0}(M_0^{\tau_0})'\rangle(m_\delta^{\tau_0} - m_\delta^z)| + \mathcal{O}(|m_\delta^z - m_\delta^{\tau_0}|^2 + \eta\delta) \\ &= 2|\langle L^{\tau_0}, M_\delta^{\tau_0}\Gamma[B^{\tau_0}]B^{\tau_0}\rangle||m_\delta^{\tau_0} - m_\delta^z| + \mathcal{O}(|m_\delta^z - m_\delta^{\tau_0}|^2 + \eta\delta) \\ &\sim |m_\delta^{\tau_0} - m_\delta^z| \sim \sqrt{\kappa + \eta}.\end{aligned} \tag{2.4.76}$$

Here, the third equality follows from the fact that $M_\delta[m_\delta] = M_0(m_\delta) + \mathcal{O}(\eta\delta)$ and M_0 is analytic in m_δ at $m_\delta^{\tau_0} = m^{\tau_0}$. In the fourth line we used (2.4.69), $M' = B$ and in the fifth line that $|\langle L^{\tau_0}, M_\delta^{\tau_0}\Gamma[B^{\tau_0}]B^{\tau_0}\rangle|$ is non-vanishing. In the last relation, we used Corollary 2.3.3 and Lemma 2.4.1. Taking the absolute value (2.4.74), the asymptotic behaviour of $|\beta^z|$ follows from (2.4.76) and (2.4.70), i.e.

$$|\beta^z| \sim \sqrt{\kappa + \eta} + \mathcal{O}(\|\mathscr{E}^z\|_{\mathrm{sp}}^2) \sim \sqrt{\kappa + \eta}. \tag{2.4.77}$$

In the last step, we used that

$$\|\mathscr{E}^z\|_{\mathrm{sp}} = \mathcal{O}(|m_\delta^z - m_\delta^{\tau_0}|). \tag{2.4.78}$$

□

2.5 Proof of the local law

In this section, we prove Theorem 2.2.9 and Proposition 2.2.10. For shifted squares of Wigner matrices (see Definition 2.4.2), we provide a direct proof below in

Lemma 2.5.1. The Stieltjes transform m and the resolvent $\mathbf{g}$ are submatrices of M_0 and $\mathbf{G}_0$, introduced in (2.4.18) and (2.4.4), respectively. For polynomials that are not shifted squares of Wigner matrices, the local laws are therefore a direct consequence of Proposition 2.4.3 and Proposition 2.4.4, the local laws for the linearization around regular edges and away from the spectrum, and we will spend most of the section proving them. The main steps of our proof for the edge local law, Proposition 2.4.3, follow the general strategy of [3, Proposition 3.3]. Here, we briefly describe the main ideas of the proof. Throughout this section we will use the notation $\mathbf{\Delta}_\delta := \mathbf{G}_\delta - M_\delta$. First, we establish a global law away from the spectrum, stated below in Proposition 2.5.7. Then, we use the global law as a starting point for a bootstrapping process. The bootstrapping proposition, Proposition 2.5.8, establishes a local law iteratively on scales ever closer to the optimal scale, $\eta \sim N^{-1+\varepsilon}$. Lemmata 2.5.4, 2.5.5 and 2.5.6 are auxiliary results used in the bootstrapping process. Lemma 2.5.5 establishes a bound for $\mathbf{\Delta}_\delta$ in terms of the error $\mathbf{D}$ and Θ_δ, the projection of $\mathbf{\Delta}_\delta$ onto its unstable direction, as well as an approximate quadratic equation for Θ_δ. Lemma 2.5.6 transforms the quadratic bound on Θ_δ into a linear bound. The most crucial difference between our proof and that of [3, Proposition 3.3] is addressed in Lemma 2.5.4. It provides a naive upper bound for $\mathbf{G}_0^*\mathbf{G}_0$ and allows us to estimate $\mathbf{G}_0^*\mathbf{G}_0$ in terms of $\mathbf{G}_1^*\mathbf{G}_1$. Unlike $\mathbf{G}_0$, the matrix $\mathbf{G}_1$ is a resolvent and thus satisfies $\mathbf{G}_1^*\mathbf{G}_1 = \eta^{-1} \operatorname{Im} \mathbf{G}_1$. As $\mathbf{D}$ is in turn bounded by $\mathbf{G}_\delta^*\mathbf{G}_\delta$ this step is necessary to obtain the correct upper bound for $\mathbf{G}_0 - M_0$. The proof of the local law away from the spectrum, Proposition 2.4.4, makes use of a similar but more simple strategy since $\mathscr{L}$ does not have an unstable direction away from $\operatorname{supp} \rho$.

Lemma 2.5.1. *Let q be a shifted square of a Wigner matrix as introduced in Definition 2.4.2 and let ρ have a regular edge at τ_0.*

1. *There is a $\kappa_0 > 0$, depending only on the coefficients of q, such that for all $\varepsilon, \gamma, D > 0$ and $z \in \mathbb{D}_\gamma^{\kappa_0}$ the isotropic local law (2.2.16) holds for all deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$ and the averaged local law (2.2.17) holds for all deterministic $\mathbf{B} \in \mathbb{C}^{N\times N}$. If additionally $E \notin \operatorname{supp}(\rho)$, we also have the improved local law (2.2.18).*
2. *For all $C > 0$ there is an $\eta_0 > 0$, depending only on the coefficients of q, such that the averaged local law (2.2.19) holds true for all deterministic $\mathbf{B} \in \mathbb{C}^{N\times N}$, $\gamma, \varepsilon, D > 0$ and $z \in \mathbb{G}_\gamma^{C,\eta_0}$.*

Proof. By assumption, q is a shifted square of a Wigner matrix. Thus it is of the form

$$q(\mathbf{X}) = q_{a,c}(\mathbf{W}) := a\mathbf{W}^2 + c, \tag{2.5.1}$$

with $a, c \in \mathbb{R}$, $a \neq 0$ and $\mathbf{W}$ being a Wigner matrix. By $\mathbf{g}_{a,c}$ we denote the resolvent of (2.5.1) and by $m_{a,c}$ the solution of (2.2.4) for (2.5.1) with explicitly stated dependence on a, c. We have

$$\mathbf{g}_{a,c}(z) = \frac{1}{a\mathbf{W}^2 + c - z} = \frac{1}{a}\mathbf{g}_{1,0}\left(\frac{z-c}{a}\right) \tag{2.5.2}$$

and

$$m_{a,c}(z) = \frac{1}{a} m_{1,0}\left(\frac{z-c}{a}\right). \tag{2.5.3}$$

Therefore $q_{a,c}$ has a regular edge at $a\tau_0 + c$ if and only if $q_{1,0}$ has a regular edge at τ_0 and Theorem 2.2.9 for general a, c follows from $a = 1$, $c = 0$. Thus let w.l.o.g. $a = 1$, $c = 0$ and we drop the subscripts again from m and $\mathbf{g}$.

For $\zeta \in \mathbb{H}$ let $\mathbf{g}_{\mathbf{W}} := \frac{1}{\mathbf{W}-\zeta}$ be the resolvent of $\mathbf{W}$ at spectral parameter ζ. We have

$$\mathbf{g}(z) = \frac{1}{2\sqrt{z}}\left(\mathbf{g}_{\mathbf{W}}(\sqrt{z}) + \mathbf{g}_{-\mathbf{W}}(\sqrt{z})\right), \tag{2.5.4}$$

where again the square root function is chosen such that the positive real axis is mapped to itself and with a branch cut along the negative real axis. Let $\rho_{\mathrm{sc}}(x) := \frac{1}{2\pi}\sqrt{(4-x^2)_+}$ with $(y)_+ := \max\{y, 0\}$ be the semi-circle density and m_{sc} its Stieltjes transform. By explicitly solving (2.2.4) for $q = \mathbf{W}^2$ and m_{sc} we find

$$m(z) = \frac{1}{\sqrt{z}}\left(-\sqrt{z} + \sqrt{z-4}\right) = \frac{1}{\sqrt{z}} m_{\mathrm{sc}}(\sqrt{z}). \tag{2.5.5}$$

for all $z \in \mathbb{H}$. The semi-circle density ρ_{sc} has regular edges at ± 2, therefore ρ has exactly one regular edge at $\tau_0 = 4$. Its left edge is a hard edge and has a singularity at zero. Combining (2.5.4) and (2.5.5) we find

$$\mathbf{g}(z) - m(z) = \frac{1}{2\sqrt{z}}\left(\left(\mathbf{g}_{\mathbf{W}}(\sqrt{z}) - m_{\mathrm{sc}}(\sqrt{z})\right) + \left(\mathbf{g}_{-\mathbf{W}}(\sqrt{z}) - m_{\mathrm{sc}}(\sqrt{z})\right)\right). \tag{2.5.6}$$

Note that $-\mathbf{W}$ is also a Wigner matrix. Therefore, Statement 1 of Lemma 2.5.1 around $\tau_0 = 4a + c$ follows from [3, Theorem 2.6] and Statement 2 of Lemma 2.5.1 Proposition 2.2.10 from [29, Theorem 2.1]. □

From now on we assume that q is not a shifted square of a Wigner matrix. To prove Proposition 2.4.3 and Proposition 2.4.4, we will require several intermediate results, which we state below. Throughout the remainder of the section, we assume w.l.o.g. that there is some $\alpha > 0$ such that

$$\|\mathbf{X}_i\| \lesssim N^{\alpha}. \tag{2.5.7}$$

The assumption can be removed by a standard argument using Chebyshev's inequality and our moment assumption (2.2.1).

We also introduce stochastic domination, a commonly used notation used to state high probability bounds in a way well adapted to our needs. It has first been introduced in a slightly different form in [26], for the form stated here see e.g. [27].

Definition 2.5.2 (Stochastic domination)**.** *Let $X = (X^{(N)})_{N\in\mathbb{N}}$ and $Y = (Y^{(N)})_{N\in\mathbb{N}}$ be families of non-negative random variables. We say X is stochastically dominated by Y if for all (small) $\varepsilon > 0$ and (large) $D > 0$*

$$\mathbb{P}\left(X^{(N)} \geq N^{\varepsilon} Y^{(N)}\right) \lesssim_{D,\varepsilon} N^{-D}. \tag{2.5.8}$$

We denote this relation by $X \prec Y$. If the constant in the definition depends on any other parameters α, we write $X \prec_\alpha Y$.

Furthermore, we introduce the norm $\|\cdot\|_* := \|\cdot\|_*^{K,\mathbf{x},\mathbf{y}}$ for deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{kN}$ and $k, K \in \mathbb{N}$. Our definition follows that of [3] and it goes back to [29] and we refer to those two works for details. First, we define the set

$$I_0 := \{\mathbf{x}, \mathbf{y}\} \cup \{\mathbf{e}_a, L^*_{a\cdot} : a \in [\![kN]\!]\}, \tag{2.5.9}$$

where $\mathbf{e}_a$ is the a^{th} standard base vector. Replacing a scalar index by a dot denotes the vector that runs over the entire range of the index, e.g. $\mathbf{R}_{a\cdot} := (\mathbf{R}_{ab})_{b\in[\![kN]\!]}$ is the a^{th} row vector of a matrix $\mathbf{R}$. For $j \in \mathbb{N}$ we define the set I_j recursively by

$$I_{j+1} := I_j \cup \{M_\delta \mathbf{u} : \mathbf{u} \in I_j\} \cup \{\kappa_c((M_\delta \mathbf{u})a, b\cdot), \kappa_d((M_\delta \mathbf{u})a, \cdot b) : \mathbf{u} \in I_j, a, b \in [\![kN]\!]\}, \tag{2.5.10}$$

where $\kappa(ab, cd)$ denotes the cumulant of the (a, b) entry and the (c, d) entry of $\sum_{j=1}^l K_j \otimes \mathbf{X}_j$ and κ_c and κ_d denote the decomposition of κ into its direct and its cross contribution according to the Hermitian symmetry. In (2.5.10) we also use the shorthand notation $\kappa(\mathbf{x}b, cd) := \sum_a x_a \kappa(ab, cd)$. Then $\|\cdot\|_*$ is defined as

$$\|\mathbf{R}\|_* := \sum_{0\le j\le K} N^{-\frac{j}{2K}} \|\mathbf{R}\|_{I_j} + N^{-\frac{1}{2}} \max_{\mathbf{u}\in I_K} \frac{\|\mathbf{R}\mathbf{u}\|}{\|\mathbf{u}\|}, \quad \|\mathbf{R}\|_{I_j} := \max_{\mathbf{u},\mathbf{v}\in I_j} \frac{|\langle \mathbf{v}, \mathbf{R}\mathbf{u}\rangle|}{\|\mathbf{u}\|\|\mathbf{v}\|} \tag{2.5.11}$$

The notion of stochastic domination is closely related to bounds in p-norms as can be seen from the following lemma.

Lemma 2.5.3 ([29, Lemma 5.4])**.** *Let $\mathbf{R} \in C^{kN\times kN}$, $k \in \mathbb{N}$, be a random matrix and Φ be a stochastic control parameter. Then the following holds true*

1. *If $\Phi \gtrsim N^{-C}$ and $\|\mathbf{R}\| \lesssim N^C$ for some $C > 0$ and $|\langle \mathbf{x}, \mathbf{R}\mathbf{y}\rangle| \prec \Phi\|\mathbf{x}\|\|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y}$, then $\|\mathbf{R}\|_p \lesssim_{\varepsilon,p} N^\varepsilon \Phi$ for all $\varepsilon > 0$, $p \in \mathbb{N}$.*
2. *If $\|\mathbf{R}\|_p \lesssim_{\varepsilon,p} N^\varepsilon \Phi$ for all $\varepsilon > 0$, $p \in \mathbb{N}$ then $\|\mathbf{R}\|_*^{K,\mathbf{x},\mathbf{y}} \prec \Phi$ for any fixed $K \in \mathbb{N}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$.*

Lemma 2.5.4 (A priori bound)**.** $\mathbf{G}_0$ *satisfies the a priori bound*

$$\mathbf{G}_0^*\mathbf{G}_0 \lesssim \frac{1}{\eta^2}(1 + \|\mathbf{X}\|^4) \tag{2.5.12}$$

uniformly in $z \in \mathbb{H}$ with $\eta \lesssim 1$.

Proof. We recall that $\mathbf{G}_0$ is given by

$$\mathbf{G}_0 = \begin{pmatrix} \mathbf{g} & \mathbf{g}\mathbf{X}^t A \\ A\mathbf{X}\mathbf{g} & -A + A\mathbf{X}\mathbf{g}\mathbf{X}^t A \end{pmatrix}. \tag{2.5.13}$$

We estimate $\mathbf{G}_0$ blockwise to obtain

$$\mathbf{G}_0\mathbf{G}_0^* \leq \|\mathbf{G}_0\|^2 \lesssim \|\mathbf{g}\|^2(1+\|\mathbf{X}\|^4) \leq \frac{1}{\eta^2}(1+\|\mathbf{X}\|^4). \tag{2.5.14}$$

The last inequality holds because $\mathbf{g}$ is a resolvent. □

The lemma is used in the last step of the following estimate on $\mathbf{G}_0^*\mathbf{G}_0$,

$$\begin{aligned}\mathbf{G}_0^*\mathbf{G}_0 &\leq 2(\mathbf{G}_1^*\mathbf{G}_1 + (\mathbf{G}_0-\mathbf{G}_1)^*(\mathbf{G}_0-\mathbf{G}_1)) \\ &= 2(\mathbf{G}_1^*\mathbf{G}_1 + \eta^2\mathbf{G}_1^*(I_{l+1}-J)\mathbf{G}_0^*\mathbf{G}_0(I_{l+1}-J)\mathbf{G}_1) \lesssim (1+\|\mathbf{X}\|^4)\mathbf{G}_1^*\mathbf{G}_1.\end{aligned} \tag{2.5.15}$$

In the first inequality, we applied Lemma 2.6.2 from the appendix with $R = \mathbf{G}_1$ and $T = \mathbf{G}_0 - \mathbf{G}_1$. In the second step, we used

$$\mathbf{G}_0 - \mathbf{G}_1 = \mathbf{G}_0(\mathbf{G}_1^{-1} - \mathbf{G}_0^{-1})\mathbf{G}_1 = -\mathrm{i}\eta\mathbf{G}_0(I_{l+1}-J)\mathbf{G}_1. \tag{2.5.16}$$

Taking the p-norm on both sides of (2.5.15) and applying Hölder's inequality we obtain

$$\|\mathbf{G}_0^*\mathbf{G}_0\|_p \lesssim_p (1+\|\|\mathbf{X}\|^4\|_{2p})\|\mathbf{G}_1^*\mathbf{G}_1\|_{2p} \lesssim_p \|\mathbf{G}_1^*\mathbf{G}_1\|_{2p} = \frac{\|\operatorname{Im}\mathbf{G}_1\|_{2p}}{\eta} \tag{2.5.17}$$

for all $p \in \mathbb{N}$. Here we use $\|Y\|_p := (\mathbb{E}[|Y|^p])^{1/p}$ for scalar random variables Y. In the second to last step, we used $\|\mathbf{X}\|_p \lesssim_p 1$, which follows from Lemma 2.5.3 since $\|\mathbf{X}\| \prec 1$ and $\|\mathbf{X}\| \leq N^C$ by assumption (2.5.7).

We define the sets

$$\begin{aligned}\mathbb{D}^{\kappa_0} &:= \{z \in \mathbb{H} : |E-\tau_0| \leq \kappa_0\}, \\ \mathbb{G}^{C,\eta_0} &:= \{z \in \mathbb{H} : C^{-1} < \operatorname{dist}(E, \operatorname{supp}(\rho)) < C, \eta \leq \eta_0\}\end{aligned} \tag{2.5.18}$$

and the random variable

$$\Theta_\delta := \frac{\langle L \otimes \mathbf{I}_N, \mathbf{\Delta}_\delta\rangle}{\langle L \otimes \mathbf{I}_N, B \otimes \mathbf{I}_N\rangle}. \tag{2.5.19}$$

With Θ_δ we can write the projection of $\mathbf{\Delta}_\delta$ onto the critical direction (see (2.4.35)) as $\mathscr{P}[\mathbf{\Delta}_\delta] = \Theta_\delta B$. Furthermore, let $\chi(A)$ denote the indicator function on event A. We import following lemmata from [3, Proposition 3.3] and [3, Lemma 3.9].

Lemma 2.5.5 ([3, Proposition 3.3]). *Let $\delta \in \{0,1\}$ and $\|\cdot\|_* = \|\cdot\|_*^{K,\mathbf{x},\mathbf{y}}$. There is a $\kappa_0 \sim 1$ and deterministic matrices $\mathbf{R}_i$ with $\|\mathbf{R}_i\| \lesssim 1$ for $i = 1,2$ such that*

$$\mathbf{\Delta}_\delta\chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}) = (\Theta_\delta B - \mathscr{L}^{-1}[(\mathbb{1}-\mathscr{P})[M_\delta\mathbf{D}]] + \mathcal{E})\chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}), \tag{2.5.20}$$

with an error function $\mathcal{E}$ of size

$$\|\mathcal{E}\|_* = \mathcal{O}\left(N^{\frac{2}{K}}(|\Theta_\delta|^2 + \|\mathbf{D}\|_*^2)\right) \tag{2.5.21}$$

and Θ_δ satisfying the approximate quadratic equation

$$\left(\xi_1\Theta_\delta + \xi_2\Theta_\delta^2\right)\chi(\|\mathbf{\Delta}_\delta\|_* \le N^{-\frac{3}{K}}) = \mathcal{O}\left(N^{\frac{2}{K}}\|\mathbf{D}\|_*^2 + |\langle \mathbf{R}_1, \mathbf{D}\rangle| + |\langle \mathbf{R}_2, \mathbf{D}\rangle|\right) \tag{2.5.22}$$

with $\xi_1 \sim \sqrt{\kappa+\eta}$, $\xi_2 \sim 1$ uniformly in $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{(l+1)N}$ and $z \in \mathbb{D}^{\kappa_0}$.

Proof. The proof of [3, Proposition 3.3] uses [3, Proposition 3.1] as an input. After replacing [3, Proposition 3.1] by our analogous Proposition 2.4.6 the proof follows theirs line by line. □

Lemma 2.5.6 ([3, Lemma 3.9])**.** *Let $d = d(\eta)$ be a monotonically decreasing function in $\eta \ge N^{-1}$ and assume $0 \le d \lesssim N^{-\varepsilon}$ for some $\varepsilon > 0$. Suppose there are $\kappa_0, \gamma > 0$ such that*

$$\begin{aligned} |\xi_1\Theta_\delta + \xi_2\Theta_\delta^2| &\lesssim d \text{ for all } z \in \mathbb{D}_\gamma^{\kappa_0}, \\ |\Theta_\delta| &\lesssim \min\left\{\frac{d}{\sqrt{\kappa+\eta}}, \sqrt{d}\right\} \text{ for some } z_0 \in \mathbb{D}_\gamma^{\kappa_0}. \end{aligned} \tag{2.5.23}$$

Then also $|\Theta_\delta| \lesssim \min\{d/\sqrt{\kappa+\eta}, \sqrt{d}\}$ for all $z' \in \mathbb{D}_\gamma^{\kappa_0}$ with $\operatorname{Re} z' = \operatorname{Re} z_0$ and $\operatorname{Im} z' \le \operatorname{Im} z_0$.

Proposition 2.5.7 (Global Law)**.** *For all $C > 0$ and some $\kappa_0 > 0$ with $\kappa_0 \sim 1$ there is an $\eta_0 > 0$ such that for all $\delta \in [0,1]$, $z \in \mathbb{D}^{\kappa_0} \cup \mathbb{G}^{C,\eta_0}$ and $\varepsilon, D > 0$ the isotropic global law,*

$$\mathbb{P}\left(|\langle \mathbf{x}, \mathbf{\Delta}_\delta \mathbf{y}\rangle| > \|\mathbf{x}\|\|\mathbf{y}\|\frac{N^\varepsilon}{\sqrt{N}}\right) \lesssim_{\varepsilon,D,\eta,C} N^{-D} \tag{2.5.24}$$

holds for all deterministic $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{(l+1)N}$. Additionally, we have a global averaged law,

$$\mathbb{P}\left(|\langle \mathbf{B}\mathbf{\Delta}_\delta\rangle| > \|\mathbf{B}\|\frac{N^\varepsilon}{N}\right) \lesssim_{\varepsilon,D,\eta,C} N^{-D}, \tag{2.5.25}$$

for all deterministic $\mathbf{B} \in \mathbb{C}^{(l+1)N\times(l+1)N}$.

Proof. First, consider $\delta = 1$ and A being invertible. Then for any $z \in \mathbb{H}$ the matrix $\mathbf{G}_1$ is a resolvent of the Hermitian random matrix $\mathbf{L} - EJ$ with spectral parameter $\mathrm{i}\eta$. In this regime the global law is covered by [29, Theorem 2.1].

Now consider A to be non-invertible. Then there is a $u > 0$ s.t. $A + \varepsilon$ is invertible for all $0 < \varepsilon < u$ and a global law for $A + \varepsilon$ also follows from [29, Theorem 2.1]. For $C > 0$ fix a $z \in \mathbb{D}^{\kappa_0} \cup \mathbb{G}^{C,\eta_0}$ with κ_0, η_0 sufficiently small for Proposition 2.4.6 to be applicable. $\mathscr{L}$ is continuous in $\varepsilon \in [0,u]$ for sufficiently small $u \sim 1$ and Proposition 2.4.6 holds true uniformly for sufficiently small ε. As M_1 and $\mathbf{G}_1$ are also continuous in $\varepsilon \in [0,u]$ for some $u \sim 1$, the proposition follows from a stochastic continuity argument.

For $\delta \in [0,1)$ and arbitrary A the global law follows from another stochastic continuity argument as M_δ, $\mathbf{G}_\delta$ and $\mathscr{L}$ are also continuous in δ. □

Before starting our bootstrapping argument, we introduce the following notions for isotropic and averaged stochastic dominance for random matrices $\mathbf{R}$ and deterministic control parameters Λ,

$$\begin{aligned} |\mathbf{R}| \prec \Lambda \text{ in } \mathbb{D} &\Leftrightarrow \|\mathbf{R}\|_*^{K,\mathbf{x},\mathbf{y}} \prec \Lambda \text{ uniform in } \mathbf{x}, \mathbf{y} \text{ and } z \in \mathbb{D} \\ |\mathbf{R}|_{\mathrm{av}} \prec \Lambda \text{ in } \mathbb{D} &\Leftrightarrow \frac{|\langle \mathbf{BR} \rangle|}{\|\mathbf{B}\|} \prec \Lambda \text{ uniform in } \mathbf{B} \neq 0 \text{ and } z \in \mathbb{D}. \end{aligned} \tag{2.5.26}$$

Proposition 2.5.8 (Bootstrapping)**.** *Assume the following:*

1. The isotropic and averaged local laws,

$$\begin{aligned} |\mathbf{\Delta}_\delta| &\prec N^{\frac{2}{K}} \left(\sqrt{\frac{\operatorname{Im} m}{N\eta}} + \frac{N^{\frac{2}{K}}}{N\eta} \right), \\ |\mathbf{\Delta}_\delta|_{\mathrm{av}} &\prec \begin{cases} \dfrac{N^{\frac{2}{K}}}{N\eta}, & E \in \operatorname{supp}\rho, \\ \dfrac{N^{\frac{2}{K}}}{N(\eta+\kappa)} + \dfrac{N^{\frac{4}{K}}}{N^2\eta^2\sqrt{\eta+\kappa}}, & E \notin \operatorname{supp}\rho, \end{cases} \end{aligned} \tag{2.5.27}$$

hold on $z = E + \mathrm{i}\eta \in \mathbb{D}_{\gamma_0}^{\kappa_0}$ for some γ_0, κ_0, K and $\delta \in \{0, 1\}$.

2. The isotropic and averaged local laws,

$$|\mathbf{\Delta}_\delta| \prec_C N^{\frac{2}{K}} \left(\frac{1}{\sqrt{N}} + \frac{N^{\frac{2}{K}}}{N\eta} \right), \quad |\mathbf{\Delta}_\delta|_{\mathrm{av}} \prec_C \left(\frac{1}{N} + \frac{N^{\frac{4}{K}}}{(N\eta)^2} \right) \tag{2.5.28}$$

hold on $z = E + \mathrm{i}\eta \in \mathbb{G}_{\gamma_0}^{C,\eta_0}$ for some γ_0, C, η_0, K and $\delta \in \{0, 1\}$.

Then, for all $\gamma > 0$ with $\gamma \geq \frac{100}{K}$ there is a $\gamma_s > 0$ independent of γ_0 such that the local laws (2.5.27) also hold on $\mathbb{D}_{\gamma_1}^{\kappa_0}$ with $\gamma_1 = \max\{\gamma, \gamma_0 - \gamma_s\}$ and the local laws (2.5.28) also hold on $\mathbb{G}_{\gamma_1}^{C,\eta_0}$ with $\gamma_1 = \max\{\gamma, \gamma_0 - \gamma_s\}$.

Proof. We first prove the local law (2.5.27) for $z \in \mathbb{D}_{\gamma_1}^{\kappa_0}$ and then comment on the modifications necessary to prove (2.5.28) for $z \in \mathbb{G}_{\gamma_1}^{C,\eta_0}$. We begin by proving that $\eta \mapsto \eta\|\mathbf{G}_\delta\|_p$ is monotonically non-decreasing in η. For $\delta = 1$, the proof is given in the proof of [29, Proposition 5.5]. For $\delta = 0$, the proof modifies as follows. Let $\varepsilon > 0$. For fixed E, we define $f(\eta) := \eta\|\mathbf{G}_0(E + \mathrm{i}\eta)\|_p$. It satisfies

$$\begin{aligned} &\liminf_{\varepsilon \to 0} \frac{f(\eta+\varepsilon) - f(\eta)}{\varepsilon} \\ =&\liminf_{\varepsilon \to 0} \|\mathbf{G}_0(E + \mathrm{i}(\eta+\varepsilon))\|_p + \frac{\eta(\|\mathbf{G}_0(E + \mathrm{i}(\eta+\varepsilon))\|_p - \|\mathbf{G}_0(E + \mathrm{i}\eta)\|_p)}{\varepsilon} \\ \geq&\|\mathbf{G}_0(E + \mathrm{i}\eta)\|_p - \lim_{\varepsilon \to 0} \eta \left\| \frac{\mathbf{G}_0(E + \mathrm{i}(\eta+\varepsilon)) - \mathbf{G}_0(E + \mathrm{i}\eta)}{\varepsilon} \right\|_p \\ =&\|\mathbf{G}_0(E + \mathrm{i}\eta)\|_p - \eta\|\mathbf{G}_0(E + \mathrm{i}\eta) J \mathbf{G}_0(E + \mathrm{i}\eta)\|_p \end{aligned} \tag{2.5.29}$$

To obtain a bound for the last term, we estimate

$$\begin{aligned}\eta|\langle \mathbf{x}, \mathbf{G}_0 J \mathbf{G}_0 \mathbf{y}\rangle| &\leq \frac{\eta}{2}(\langle \mathbf{x}, \mathbf{G}_0 J \mathbf{G}_0^* \mathbf{x}\rangle + \langle \mathbf{y}, \mathbf{G}_0^* J \mathbf{G}_0 \mathbf{y}\rangle) \\ &= \frac{1}{2}(\langle \mathbf{x}, \operatorname{Im} \mathbf{G}_0 \mathbf{x}\rangle + \langle \mathbf{y}, \operatorname{Im} \mathbf{G}_0 \mathbf{y}\rangle)\end{aligned} \tag{2.5.30}$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{(l+1)N}$. In the last step, we used the Ward identity for generalized resolvents,

$$\mathbf{G}_0 J \mathbf{G}_0 = \frac{\operatorname{Im} \mathbf{G}_0}{\eta}. \tag{2.5.31}$$

Thus we find

$$\eta\|\mathbf{G}_0 J \mathbf{G}_0\|_p \leq \sup_{\|\mathbf{x}\|,\|\mathbf{y}\|=1} \left(\mathbb{E}\left(\frac{1}{2}(\langle \mathbf{x}, \operatorname{Im} \mathbf{G}_0 \mathbf{x}\rangle + \langle \mathbf{y}, \operatorname{Im} \mathbf{G}_0 \mathbf{y}\rangle)\right)^p\right)^{\frac{1}{p}} \leq \|\mathbf{G}_0\|_p, \tag{2.5.32}$$

where we used

$$|\langle \mathbf{x}, \operatorname{Im} \mathbf{R}\mathbf{x}\rangle| \leq |\langle \mathbf{x}, \mathbf{R}\mathbf{x}\rangle| \tag{2.5.33}$$

in the last step. Therefore

$$\liminf_{\varepsilon \to 0} \frac{f(\eta+\varepsilon) - f(\eta)}{\varepsilon} \geq 0 \tag{2.5.34}$$

and the claim follows.

For $\delta = 0, 1$, assume the local law (2.5.27) on $\mathbb{D}_{\gamma_0}^{\kappa_0}$. Then $\|\mathbf{G}_\delta\|_p \sim_p 1$ on $\mathbb{D}_{\gamma_0}^{\kappa_0}$ and by monotonicity of $\eta\|\mathbf{G}_\delta\|_p$ we find

$$\|\mathbf{G}_\delta\|_p \lesssim_p N^{\gamma_s}. \tag{2.5.35}$$

on $\mathbb{D}_{\gamma_1}^{\kappa_0}$. We choose $\gamma_s < \frac{1}{4}$. Then Proposition 2.4.5 yields

$$\|\mathbf{D}\|_p \lesssim_{p,\varepsilon} N^{\varepsilon + c\gamma_s} \sqrt{\frac{\|\mathbf{G}_\delta \mathbf{G}_\delta^*\|_q}{N}} \quad \text{and} \quad \|\mathbf{D}\|_p^{\mathrm{av}} \lesssim_{p,\varepsilon} N^{\varepsilon + c\gamma_s} \frac{\|\mathbf{G}_\delta \mathbf{G}_\delta^*\|_q}{N}. \tag{2.5.36}$$

For $\delta = 0$ estimate the quadratic term by making use of (2.5.17) and then use the Ward identity on $\mathbf{G}_1 \mathbf{G}_1^*$ for both $\delta = 0, 1$ to obtain

$$\begin{aligned}\|\mathbf{D}\|_p &\lesssim_{p,\varepsilon} N^{\varepsilon + c\gamma_s} \sqrt{\frac{\|\operatorname{Im} \mathbf{G}_1\|_{2q}}{\eta N}} \lesssim_p \frac{N^{\frac{c'}{2}\gamma_s}}{\sqrt{\eta N}}, \\ \|\mathbf{D}\|_p^{\mathrm{av}} &\lesssim_{p,\varepsilon} N^{\varepsilon + c\gamma_s} \frac{\|\operatorname{Im} \mathbf{G}_1\|_{2q}}{\eta N} \lesssim_p \frac{N^{c'\gamma_s}}{\eta N}.\end{aligned} \tag{2.5.37}$$

on $\mathbb{D}_{\gamma_1}^{\kappa_0}$ for sufficiently small $\varepsilon > 0$ and some modified $c' > 0$. Note that we are allowed to exchange the q norm in the bound for $\delta = 1$ by a $2q$ norm as the norm is increasing in q. After using Lemma 2.5.3 to turn the p-norm bound in the first equation of (2.5.37) into a bound on the *-norm bound, we get from Lemma 2.5.5

$$|\xi_1 \Theta_\delta + \xi_2 \Theta_\delta^2| \chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}) \prec \frac{N^{\frac{2}{K} + c'\gamma_s}}{\eta N} \tag{2.5.38}$$

on $\mathbb{D}_{\gamma_1}^{\kappa_0}$. The left hand side is also Lipschitz continuous with Lipschitz constant $\prec \eta^{-2} \leq N^2$. Thus the inequality

$$|\xi_1\Theta_\delta + \xi_2\Theta_\delta^2|\chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}) \leq N^{-\frac{10}{K}} \tag{2.5.39}$$

holds with very high probability on all of $\mathbb{D}_{\gamma_1}^{\kappa_0}$ for sufficiently large K and sufficiently small γ_s. By our assumption the local law, (2.5.27), holds on $\mathbb{D}_{\gamma_0}^{\kappa_0}$. In conjunction with another stochastic continuity argument, we also get the bound

$$|\Theta_\delta| \leq \min\left\{\frac{N^{-\frac{10}{K}}}{\sqrt{\kappa+\eta}}, N^{-\frac{5}{K}}\right\} \tag{2.5.40}$$

on all of $\mathbb{D}_{\gamma_0}^{\kappa_0}$ with very high probability. Therefore Lemma 2.5.6 can be applied with $d = N^{-10/K}$ and we obtain

$$|\Theta_\delta|\chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}) \prec N^{-\frac{5}{K}}. \tag{2.5.41}$$

Next, we turn this into a bound on $\mathbf{\Delta}_\delta$. To do so we again use the *-norm bound obtained from applying Lemma 2.5.3 to (2.5.37) to find $\|\mathbf{D}\|_* \prec N^{-\frac{7}{K}}$ for sufficiently small γ_s and sufficiently large K. Using both bounds on (2.5.20) we get

$$\|\mathbf{\Delta}_\delta\|_*\chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}) \lesssim \left(|\Theta_\delta| + N^{\frac{2}{K}}\|\mathbf{D}\|_*\right)\chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}) \prec N^{-\frac{5}{K}}. \tag{2.5.42}$$

We have thus found a "forbidden" area for $\|\mathbf{\Delta}_\delta\|_*$. With the aid of a standard stochastic continuity argument, we remove the indicator function from the bounds (2.5.41) and (2.5.42) to get to the rough bounds

$$\|\mathbf{\Delta}_\delta\|_* \prec N^{-\frac{5}{K}} \quad \text{and} \quad |\Theta_\delta| \prec N^{-\frac{5}{K}}. \tag{2.5.43}$$

Since $\mathbf{x}$ and $\mathbf{y}$ were arbitrary we the first bound becomes $|\mathbf{\Delta}_\delta| \prec N^{-5/K}$. Now, assume that $\max\{|\mathbf{\Delta}_0|, |\mathbf{\Delta}_1|\} \prec \Lambda$ and $\max\{|\Theta_0|, |\Theta_1|\} \prec \theta$ for some deterministic $\theta \leq \Lambda \leq N^{-3/K}$. Then we know from Lemma 2.5.5 and Proposition 2.4.5 as well as (2.5.17) that

$$\max_i\{|\mathbf{\Delta}_i|\} \prec \theta + N^{\frac{2}{K}}\sqrt{\frac{\operatorname{Im} m + \Lambda}{N\eta}} \quad \text{and} \quad \max_i\{|\xi_1\Theta_i + \xi_2\Theta_i^2|\} \prec N^{\frac{2}{K}}\frac{\operatorname{Im} m + \Lambda}{N\eta} \tag{2.5.44}$$

are also deterministic bounds. Here, we also used $\operatorname{Im} m \sim \langle \operatorname{Im} M_1\rangle$ by Corollary 2.3.3, Lemma 2.4.1 as well as (2.4.18). The first bound in (2.5.44) self-improving and applying it iteratively yields

$$\max_i\{|\mathbf{\Delta}_i|\} \prec \theta + N^{\frac{2}{K}}\left(\frac{N^{\frac{2}{K}}}{N\eta} + \sqrt{\frac{\operatorname{Im} m + \theta}{N\eta}}\right) \tag{2.5.45}$$

and therefore the second bound in (2.5.44) improves to

$$\max_i\{|\xi_1\Theta_i + \xi_2\Theta_i^2|\} \prec N^{\frac{2}{K}}\frac{\operatorname{Im} m + \theta}{N\eta} + N^{\frac{4}{K}}\frac{1}{(N\eta)^2}. \tag{2.5.46}$$

Now we separately treat $\operatorname{Re} z \in \operatorname{supp}(\rho)$ and $\operatorname{Re} z \notin \operatorname{supp}(\rho)$ and we start with the former. Then by Corollary 2.3.3 we know that $\operatorname{Im} m \sim \sqrt{\kappa+\eta}$. For fixed θ we apply Lemma 2.5.6 with

$$d = N^{\frac{2}{K}} \frac{\sqrt{\kappa+\eta}+\theta}{N\eta} + N^{\frac{4}{K}} \frac{1}{(N\eta)^2} \tag{2.5.47}$$

to obtain

$$\max_i\{|\Theta_i|\} \prec \min\left\{\frac{d}{\sqrt{\kappa+\eta}}, \sqrt{d}\right\}. \tag{2.5.48}$$

This is also a self-improving bound and iterating it gives

$$\max_i\{|\Theta_i|\} \prec N^{\frac{2}{K}} \frac{1}{N\eta}, \quad \text{hence} \quad \max_i\{|\Delta_i|\} \prec N^{\frac{2}{K}} \left\{\sqrt{\frac{\operatorname{Im} m}{N\eta}} + \frac{N^{\frac{2}{K}}}{N\eta}\right\}. \tag{2.5.49}$$

For $\operatorname{Re} z \notin \operatorname{supp}(\rho)$ we have $\operatorname{Im} m \sim \frac{\eta}{\sqrt{\kappa+\eta}}$, again by Corollary 2.3.3. Analogously to the $\operatorname{Re} z \in \operatorname{supp}(\rho)$ we obtain

$$\max_i\{|\Theta_i|\} \prec N^{\frac{2}{K}} \frac{1}{N(\eta+\kappa)} + N^{\frac{4}{K}} \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}. \tag{2.5.50}$$

Finally we use (2.5.20), (2.4.31) and the bounds on $\max_i\{|\Theta_i|\}$ to arrive at the averaged bounds

$$\max_i\{|\mathbf{\Delta}_i|_{\mathrm{av}}\} \prec N^{\frac{2}{K}} \begin{cases} \frac{1}{N\eta} \text{ if } \operatorname{Re} z \in \operatorname{supp}\rho_1 \\ \frac{1}{N(\eta+\kappa)} + \frac{N^{\frac{2}{K}}}{N^2\eta^2(\eta+\kappa)^{\frac{1}{2}}} \text{ if } \operatorname{Re} z \notin \operatorname{supp}\rho_1. \end{cases} \tag{2.5.51}$$

This concludes the proof of (2.5.27) on $\mathbb{D}_{\gamma_1}^{\kappa_0}$.

Now, assume (2.5.28) on $\mathbb{G}_{\gamma_0}^{C,\eta_0}$. From (2.4.33) we have

$$\mathscr{L}[\mathbf{\Delta}_\delta] = -M_\delta \mathbf{D} + M_\delta \mathcal{S}[\mathbf{\Delta}_\delta]\mathbf{\Delta}_\delta. \tag{2.5.52}$$

We apply $\mathscr{L}^{-1}$ to the equation and take its *-norm for some deterministic $\mathbf{x}, \mathbf{y}$ to find

$$\begin{aligned} \|\mathbf{\Delta}_\delta\|_* &\le \|\mathscr{L}^{-1}[M_\delta \mathbf{D}]\|_* + \|\mathscr{L}^{-1}[M_\delta \mathcal{S}[\mathbf{\Delta}_\delta]\mathbf{\Delta}_\delta]\|_* \\ &\le \|\mathscr{L}^{-1}\|_{*\to*} \left(\|M_\delta \mathbf{D}\|_* + \|M_\delta \mathcal{S}[\mathbf{\Delta}_\delta]\mathbf{\Delta}_\delta\|_*\right) \\ &\lesssim_C N^{\frac{2}{K}} \|\mathbf{D}\|_* + N^{\frac{2}{K}} \|\mathbf{\Delta}_\delta\|_*^2. \end{aligned} \tag{2.5.53}$$

Here, $\|\mathscr{L}^{-1}\|_{*\to*}$ denotes the operator norm of $\mathscr{L}^{-1}$ with respect to the *-norm. In the last estimate, we have used $\|\mathscr{L}^{-1}\|_{*\to*} \lesssim \|\mathscr{L}^{-1}\|_{\mathrm{sp}} \lesssim_C 1$ by [29, Equation (70c)] and Proposition 2.4.6 as well as

$$\|M_\delta \mathbf{R}\|_* \lesssim N^{2/K} \|\mathbf{R}\|_* \quad \text{and} \quad \|M_\delta \mathcal{S}[\mathbf{R}]\mathbf{R}\|_* \lesssim N^{2/K} \|\mathbf{R}\|_*^2$$

by [29, Equation (70a) and (70b)]. Equation (2.5.42) also holds on $\mathbb{G}_{\gamma_1}^{C,\eta_0}$ by the same argument as in the case $z \in \mathbb{D}_{\gamma_1}^{\kappa_0}$ and for sufficiently large K and sufficiently small

γ_s we have in particular $\|\mathbf{D}\|_* \prec N^{-\frac{7}{K}}$ for $\delta \in \{0,1\}$. We use this bound on with (2.5.53) to estimate

$$\|\mathbf{\Delta}_\delta\|_* \chi(\|\mathbf{\Delta}_\delta\|_* \leq N^{-\frac{3}{K}}) \prec_C N^{-\frac{5}{K}}. \tag{2.5.54}$$

By a stochastic continuity argument, we establish the rough bound

$$\|\mathbf{\Delta}_\delta\|_* \prec_C N^{-\frac{5}{K}} \quad \text{and therefore} \quad |\mathbf{\Delta}_\delta| \prec_C N^{-\frac{5}{K}} \tag{2.5.55}$$

since $\mathbf{x}$ and $\mathbf{y}$ were arbitrary. Now, assume $\max_i\{|\mathbf{\Delta}_i|\} \prec_C \Lambda$ for some deterministic $\Lambda \leq N^{-3/K}$. Then we have another deterministic bound in the form of

$$\max_i\{|\mathbf{\Delta}_i|\} \prec_C N^{\frac{2}{K}} \sqrt{\frac{\operatorname{Im} m + \Lambda}{N\eta}}. \tag{2.5.56}$$

Iterating this self-improving bound and using $\operatorname{Im} m \sim_C \eta$ from Lemma 2.6.3 we find

$$\max_i\{|\mathbf{\Delta}_i|\} \prec_C \frac{N^{\frac{4}{K}}}{N\eta} + N^{\frac{2}{K}} \frac{1}{\sqrt{N}}. \tag{2.5.57}$$

Finally from (2.5.57), (2.5.52) and (2.4.31) we obtain the averaged bound

$$\max_i\{|\mathbf{\Delta}_i|_{\mathrm{av}}\} \prec_C \frac{N^{\frac{4}{K}}}{(N\eta)^2} + \frac{1}{N}. \tag{2.5.58}$$

Therefore we have shown that (2.5.28) holds on $\mathbb{G}_{\gamma_1}^{C,\eta_0}$. □

Proof of Propositions 2.4.3 and 2.4.4. For all $K > 0$ Equation (2.5.27) holds on $\mathbb{D}_1^{\kappa_0}$ for some κ_0 and (2.5.28) on $\mathbb{G}_1^{C,\eta_0}$ for all $C > 0$ and some η_0 depending on C due to the global law, Proposition 2.5.7. The local laws, Proposition 2.4.3 and Proposition 2.4.4, follow immediately from applying Proposition 2.5.8 finitely many times in the respective domains. □

This concludes the proof of Theorem 2.2.9 and Proposition 2.2.10 and we are left with proving their Corollaries, 2.2.11 and 2.2.12.

Proof of Corollary 2.2.11. Let λ_i, $i \in [\![N]\!]$, denote the eigenvalues of $q(\mathbf{X})$ with eigenvectors $\mathbf{v}_i$. let $\mathbf{x} \in \mathbb{C}^N$ be a deterministic and normalized vector and let $\mathbf{v}$ be an eigenvector of $q(\mathbf{X})$ with eigenvalue λ such that $|\lambda - \tau_0| < \kappa_0$ for κ_0 from Theorem 2.2.9. Evaluating $\mathbf{g}$ at spectral parameter $\lambda + \mathrm{i}\eta$ have with very high probability that

$$1 \gtrsim \operatorname{Im}\langle \mathbf{x}, \mathbf{g}(\lambda + \mathrm{i}\eta)\mathbf{x}\rangle = \sum_{i=1}^{N} \eta \frac{\eta}{\eta^2 + (\lambda - \lambda_i)^2} |\langle \mathbf{x}, \mathbf{v}_i\rangle|^2 \geq \frac{|\langle \mathbf{x}, \mathbf{v}_i\rangle|^2}{\eta} \tag{2.5.59}$$

for all $\eta \geq N^{-1+\varepsilon}$ and all $\varepsilon > 0$. As the deterministic vector $\mathbf{x}$ was arbitrary, the eigenvalue delocalization, (2.2.20), follows. □

Proof of Corollary 2.2.12. Let q have regular edge at τ_0. W.l.o.g. we can assume that $\tau_0 = \tau_+$ is a right edge. Otherwise, consider the right edge of $-q$. From (2.2.18) and Proposition 2.2.10 it follows that

$$\mathbb{P}\left(|\langle \mathbf{B}(\mathbf{g} - m)\rangle| > \|\mathbf{B}\| N^{\varepsilon}\left(\frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}\right)\right) \lesssim_{\varepsilon,\gamma,D,C} N^{-D} \tag{2.5.60}$$

for all deterministic $\mathbf{B} \in \mathbb{C}^{N\times N}$ and $z \in \mathbb{D}_{\gamma}^{\kappa_0} \cup \mathbb{G}_{\gamma}^{C,\eta_0}$ with $E \notin \operatorname{supp}(\rho)$. In particular, this holds true for $C \geq \kappa_0^{-1}$. Following [31, Chapter 11.1], this implies for all $\varepsilon > 0$ that with very high probability there are no eigenvalues λ such that $\lambda \notin \operatorname{supp}(\rho)$ and $N^{-2/3+\varepsilon} \leq \lambda - \tau_+ \leq C$, i.e.

$$\mathbb{P}\left(\exists \lambda \in \operatorname{Spec}(q(\mathbf{X})) \,:\, N^{-\frac{2}{3}+\varepsilon} \leq \lambda - \tau_+ < C\right) \lesssim_{\varepsilon,D,C} N^{-D}. \tag{2.5.61}$$

Additionally, it follows from the trivial bound (2.4.32) that there is a $K > 0$ such that

$$\mathbb{P}\left(\|q(\mathbf{X})\| \geq K\right) \lesssim_{D} N^{-D}. \tag{2.5.62}$$

Combining (2.5.61) for some $C > 0$ such that $C + \tau_0 \geq K$ and (2.5.62), we have

$$\mathbb{P}\left(\exists \lambda \in \operatorname{Spec}(q(\mathbf{X})) \,:\, \lambda - \tau_+ \geq N^{-\frac{2}{3}+\varepsilon}\right) \lesssim_{\varepsilon,D} N^{-D}. \tag{2.5.63}$$

By a standard argument, (2.5.63) in conjunction with the averaged local law, (2.2.17), implies eigenvalue rigidity around the edge as in Theorem 2.2.12 (see e.g. [31, Chapter 11.2-11.4]). □

2.6 Appendix

2.6.1 The entrywise real part of a Hermitian matrix.

Let $n \in \mathbb{N}$ and $H \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and $\widehat{H} = \frac{1}{2}(H + H^t)$ be its entrywise real part. The following lemma summarizes how the two matrices are related.

Lemma 2.6.1. *Let h_i and $\widehat{h}_i$ be the eigenvalues of H and $\widehat{H}$ respectively, both arranged in non-increasing order. The following holds true*

1. *$h_1 \geq \widehat{h}_1$ and $h_n \leq \widehat{h}_n$.*
2. *$\|H\| \geq \|\widehat{H}\|$.*
3. *If $H \geq 0$, then $\widehat{H} \geq 0$ and conversely $H \leq 0$ implies $\widehat{H} \leq 0$.*
4. *Let additionally $\operatorname{rank} H = 1$. Then we have $\operatorname{rank} \widehat{H} = 1$ if and only if $H \in \mathbb{R}^{l\times l}$. Otherwise we have $\operatorname{rank} \widehat{H} = 2$.*

Proof. Let $\widetilde{H} = \frac{1}{2\mathrm{i}}(H - H^t)$ be the entrywise imaginary part of H, i.e. $H = \widehat{H} + \mathrm{i}\widetilde{H}$. First, we will prove $h_1 \geq \widehat{h}_1$. Note that for all $a \in \mathbb{R}$ the norms of $H+a$ and $\widehat{H}+a$ are given by $\|H+a\| = \max\{h_1+a, -h_n-a\}$ and $\|\widehat{H}+a\| = \max\{\widehat{h}_1+a, -\widehat{h}_n-a\}$. Thus for sufficiently large $a \in \mathbb{R}$ (depending on H) we have $\|H+a\| = h_1 + a$ and $\|\widehat{H}+a\| = \widehat{h}_1 + a$. For such an a let $v_1 \in \mathbb{R}^n$ be a normalized eigenvector of $\widehat{H}+a$ corresponding to the eigenvalue $\widehat{h}_1 + a$ (it can be chosen purely real since $\widehat{H}$ is a real symmetric matrix). Then

$$\begin{aligned} h_1 + a = \|H+a\| \geq \|(H+a)v_1\| &= \sqrt{\|(\widehat{H}+a)v_1\|^2 + \|\widetilde{H}v_1\|^2} \\ &\geq \|(\widehat{H}+a)v_1\| = \widehat{h}_1 + a. \end{aligned} \tag{2.6.1}$$

Here the second equality holds since $\widetilde{H}v_1$ is a purely imaginary vector. The claim $h_1 \geq \widehat{h}_1$ follows. Choosing a sufficiently small, we find $h_n \leq \widehat{h}_n$ by a similar argument. Since $\|H\| = \max\{h_1, -h_n\}$ and $\|\widehat{H}\| = \max\{\widehat{h}_1, -\widehat{h}_n\}$, the inequality $\|H\| \geq \|\widehat{H}\|$ follows. Next, let $H \geq 0$. Then $h_n \geq 0$ and thus $\widehat{h}_n \geq h_n \geq 0$. The inequality $\widehat{H} \geq 0$ follows. Similarly, $H \leq 0$ implies $\widehat{H} \leq 0$.

Now let $\operatorname{rank} H = 1$. Then $\operatorname{rank} \widehat{H} = 1$ if $H \in \mathbb{R}^{n\times n}$ is clear. Let on the other hand $H \in \mathbb{C}^{n\times n} \setminus \mathbb{R}^{n\times n}$. Then $H = \alpha v v^*$ for some $\alpha \in \mathbb{R}\setminus\{0\}$ and normalized $v \in \mathbb{C}^n \setminus e^{\mathrm{i}\varphi}\mathbb{R}^n$ for all $\varphi \in \mathbb{R}$. Then $\widehat{H} = \frac{\alpha}{2}(vv^* + \bar{v}\bar{v}^*)$ and since v and $\bar{v}$ are only linearly dependent if $v \in e^{\mathrm{i}\varphi}\mathbb{R}^n$ the claim $\operatorname{rank}\widehat{H} = 2$ follows. □

2.6.2 Matrix inequality

Here we provide a proof for the matrix inequality used in (2.5.15).

Lemma 2.6.2. *Let $R, T \in \mathbb{C}^{n\times n}$ be arbitrary matrices. Then the following inequality holds:*

$$(R+T)^*(R+T) \leq 2(R^*R + T^*T). \tag{2.6.2}$$

Proof. We first note that

$$(R+T)^*(R+T) = R^*R + T^*T + R^*T + T^*R \tag{2.6.3}$$

and it is thus sufficient to bound the last two terms. Let $v \in \mathbb{C}^n$ be arbitrary. Since $R^*T + T^*R$ is Hermitian, $\langle v, (R^*T + T^*R)v\rangle \in \mathbb{R}$ and we estimate

$$\begin{aligned} \langle v, (R^*T + T^*R)v\rangle &= 2\operatorname{Re}(\langle Rv, Tv\rangle) \leq 2\|Rv\|\|Tv\| \\ &\leq \|Rv\|^2 + \|Tv\|^2 = \langle v, (R^*R + T^*T)v\rangle, \end{aligned} \tag{2.6.4}$$

where we used the Cauchy-Schwarz inequality in the second step. As v was arbitrary, $R^*T + R^*T \leq R^*R + T^*T$ follows and we conclude the lemma. □

2.6.3 Some properties of the solution of the Dyson Equation

Proof of Proposition 2.2.3. There can at most be one analytic function in the upper half-plane that satisfies (2.2.4) and $\lim_{z\to\infty} zm(z) = -1$ since (2.2.4) is stable at infinity. Thus we are left with proving existence. First, consider the case of A being invertible. Let $\{s_1, \ldots, s_l\}$ be a family of free semi-circular variables in a C^* probability space $(\mathscr{S}, \tau)$ (see [28, Appendix B]) and let $s := (s_i)_{i\in[\![l]\!]}$. Define

$$\mathbf{L}_{\mathrm{sc}} = K_0 \otimes \mathbb{1} + \sum_{j=1}^{l} K_j \otimes s_j. \tag{2.6.5}$$

Following the proof of [28, Lemma 2.6], the matrix-valued function $M_0 : \mathbb{H} \to \mathbb{C}^{(l+1)\times(l+1)}$, given by

$$M_0(z) = (\mathrm{id} \otimes \tau)(\mathbf{L}_{\mathrm{sc}} - zJ \otimes \mathbb{1})^{-1}, \tag{2.6.6}$$

is a solution to (2.4.14) for $\delta = 0$. Thus its (1,1) entry satisfies (2.4.20) for $\delta = 0$ and therefore (2.2.4). Using the Schur complement formula, we find

$$(M_0(z))_{11} = \tau(q(s) - z \otimes \mathbb{1})^{-1}. \tag{2.6.7}$$

The polynomial $q(s)$, defined in (2.1.1) is self-adjoint, thus $(M_0)_{11}$ is analytic on $\mathbb{H}$, has positive imaginary part and $\lim_{z\to\infty} z(M_0(z))_{11} = -1$. Therefore $m := (M_0)_{11}$ is the unique function that satisfies all conditions of Proposition 2.2.3.

Now, let A be non-invertible. Then there is a $u \sim 1$ such that $A^\varepsilon := A + \varepsilon$ is invertible for all $0 < \varepsilon < u$. Let q^ε and γ^ε be the objects given by replacing A with $A + \varepsilon$ in the definitions of q in (2.1.1) and γ in (2.2.5), respectively. By the above argument, the function

$$m^\varepsilon(z) := \tau(q^\varepsilon(s) - z \otimes \mathbb{1})^{-1} \tag{2.6.8}$$

is a solution to the equation

$$-\frac{1}{m^\varepsilon} = z + \gamma^\varepsilon(m^\varepsilon). \tag{2.6.9}$$

For any fixed $z \in \mathbb{H}$, both m^ε and γ^ε are continuous in ε at $\varepsilon = 0$. Thus $m := m^0$ solves (2.2.4) and m is analytic on $\mathbb{H}$, has positive imaginary part and satisfies $\lim_{z\to\infty} zm(z) = -1$. Therefore it is the unique function that satisfies all conditions of Proposition 2.2.3. □

The function m is a Stieltjes transform of a real-valued probability measure ρ with compact support. As such it can be analytically extended to $\mathbb{C} \setminus \mathrm{supp}(\rho)$. The following lemma summarizes the properties of the extension, also called m, on and above the real axis outside of the spectrum

Lemma 2.6.3. *For the analytical extension of m to $\mathbb{C}\setminus\mathrm{supp}(\rho)$ we have the following.*

1. *The function m is real-valued on $\mathbb{R} \setminus \operatorname{supp}(\rho)$ and $m'(E) > 0$ for all $E \in \mathbb{R} \setminus \operatorname{supp}(\rho)$.*

2. *For all $C > 0$ there is an $\eta_0 > 0$ such that for all $z = E + \mathrm{i}\eta$ with $C^{-1} \leq \operatorname{dist}(E, \operatorname{supp}(\rho)) \leq C$ and $0 < \eta \leq \eta_0$ we have $\operatorname{Im} m \sim_C \eta$.*

Proof. The fact that $m(E) \in \mathbb{R}$ for all $\mathbb{R} \setminus \operatorname{supp}(\rho)$ follows immediately for from (2.2.6) and taking the derivative of (2.2.6) we find

$$m'(E) = \int_{\mathbb{R}} \frac{\rho(\mathrm{d}x)}{(x-E)^2} > 0. \tag{2.6.10}$$

Using $m(E) \in \mathbb{R}$ we have for $z \in \mathbb{H}$

$$\operatorname{Im} m(z) = \operatorname{Im}(m(z) - m(E)) = \eta m'(E) + \mathcal{O}(\eta^2). \tag{2.6.11}$$

By (2.6.10) and continuity of the derivative, we have that $m'(E)$ is bounded from above and bounded away from 0 on all compact subsets of $\mathbb{R} \setminus \operatorname{supp}(\rho)$. Therefore we have $m'(E) \sim_C 1$ for all $C > 0$ and E such that $C^{-1} \leq \operatorname{dist}(E, \operatorname{supp}(\rho)) \leq C$. □

Proof of Lemma 2.4.1. The function m is a Stieltjes transform and as such analytically extendable to $\mathbb{C} \setminus \operatorname{supp}(\rho)$. Since, by assumption, ρ has a regular edge at $\tau_\pm$ it can also be continuously extended to $\tau_\pm$. For $x \in (\mathbb{R} \setminus \operatorname{supp}(\rho)) \cup \{\tau_\pm\}$ we denote this extension by $m(x)$. By continuity it satisfies

$$-\frac{1}{m(x)} = x + \gamma(m(x)). \tag{2.6.12}$$

Consider (2.4.20) at $z \in \mathbb{H}$ as a perturbation to (2.6.12). More precisely, consider the following equation in the unknown function $\widetilde{m}_\delta(z)$,

$$\frac{1}{m(x)} - \frac{1}{\widetilde{m}_\delta(z)} = \gamma_\delta(\widetilde{m}_\delta(z)) - \gamma_0(m(x)) + (z - x), \tag{2.6.13}$$

where we have used $\gamma_0 = \gamma$. Define by $\dot{\gamma}_\delta(y) := \partial_t \gamma_{\frac{t}{\mathrm{i}\eta}}(y)|_{t=\mathrm{i}\delta\eta}$ the partial derivative with respect to $\mathrm{i}\delta\eta$ and by $\gamma'_\delta(y) := \partial_y \gamma_\delta(y)$ the derivative with respect to its variable.

Expanding (2.6.13) in $m_\delta(z) - m(x)$ and $\mathrm{i}\delta\eta$ we find

$$\begin{aligned}\left(\frac{1}{m(x)^2} - \gamma_0'(m(x))\right)(\widetilde{m}_\delta(z) - m(x)) - \left(\frac{1}{m(x)^3} + \frac{1}{2}\gamma_0''(m(x))\right)(\widetilde{m}_\delta(z) - m(x))^2 \\ = \mathrm{i}\dot{\gamma}_0(m(x))\delta\eta + (z - x) + \mathcal{O}(|\widetilde{m}_\delta(z) - m(x)|^3 + |\widetilde{m}_\delta(z) - m(x)|\eta + \eta^2).\end{aligned} \tag{2.6.14}$$

We identify $\frac{1}{m^2} - \gamma_0'(m) = h(m)$ and $-\frac{2}{m^3} - \gamma_0''(m) = h'(m)$ with h defined in (2.3.11). Now let $x = \tau_\pm$ and note that $m(\tau_\pm) = m_\pm$. By Lemma 2.3.5 as well as Lemma 2.3.6 we have

$$h(m_\pm) = 0, \quad |h'(m_\pm)| \sim 1 \quad \text{and} \quad \pm h'(m_\pm) > 0. \tag{2.6.15}$$

Furthermore one finds $\dot{\gamma}_0(y) > 0$ for all $y \in \mathbb{R} \setminus \mathscr{S}(\gamma)$. Thus (2.6.14) constitutes an approximate quadratic equation in $\widetilde{m}_\delta(z) - m_\pm$ and there are $u, u_1 \sim 1$ such that for all $z \in \mathbb{H}$ with $0 < |z - \tau_\pm| \leq u$, there are exactly two solutions $\widetilde{m}_\delta(z)$ with $|\widetilde{m}_\delta(z) - m_\pm| \leq u_1$. One of them has positive imaginary part and one of them has negative imaginary part. Combining (2.6.14) and (2.6.15), we find that the unique solution with positive imaginary part satisfies both parts of (2.4.22) and we denote it by m_δ.

Now let $C > 0$, $x = E = \operatorname{Re} z$ and $C^{-1} < \operatorname{dist}(E, \operatorname{supp}(\rho)) < C$. Then we have

$$h(m(E)) = (m'(E))^{-1} \sim_C 1 \tag{2.6.16}$$

by Lemma 2.6.3 and the fact that $C^{-1} < \operatorname{dist}(E, \operatorname{supp}(\rho)) < C$ constitutes a compact subset of $\mathbb{R} \setminus \operatorname{supp}(\rho))$. Therefore (2.6.14) is an approximate linear equation in $\widetilde{m}_\delta(z) - m(E)$. Hence there are $u, u_1 \sim 1$ such that for all $0 < |z - E| = \eta \leq u$ there is a unique solution $\widetilde{m}_\delta(z)$ with $|\widetilde{m}_\delta(z) - m(E)| \leq u_1$ to (2.6.14). We denote this solution by m_δ and combining (2.6.15) with (2.6.16) we find that it satisfies both parts of (2.4.23).

From now on, let A be invertible and

$$M_\delta(z) = (\mathrm{id} \otimes \tau)(\mathbf{L}_{\mathrm{sc}} - (zJ + \delta\eta(I - J)) \otimes \mathbb{1})^{-1}, \tag{2.6.17}$$

where $\mathbf{L}_{\mathrm{sc}}$ was defined in (2.6.5). For all $\eta > 0$, the imaginary part of $zJ + \delta\eta(I - J)$ is positive. Therefore, by Lemma [43, Lemma 5.4], the function M_δ is the unique solution to (2.4.14) with $\operatorname{Im} M_\delta(z) > 0$. In particular, this implies that its (1,1) entry satisfies (2.4.20). By the Schur complement formula, it is given by

$$(M_\delta(z))_{11} = \tau(q_\delta(s) - z \otimes \mathbb{1})^{-1}, \tag{2.6.18}$$

where q_δ denotes the polynomial q, defined in (2.1.1), with A replaced by A_δ. For $E \notin \operatorname{supp}(\rho)$ consider

$$\begin{aligned}(M_\delta)_{11} - m &= \tau((q_\delta(s) - z \otimes \mathbb{1})^{-1} - (q_0(s) - z \otimes \mathbb{1})^{-1}) \\ &= \tau((q_\delta(s) - z \otimes \mathbb{1})^{-1}(q_0(s) - q_\delta(s))(q_0(s) - z \otimes \mathbb{1})^{-1}).\end{aligned} \tag{2.6.19}$$

The middle factor satisfies

$$\|q_0(s) - q_\delta(s)\| = \left\| \sum_{i,j=1}^{l} (A - A_\delta) s_i s_j \right\| \lesssim \eta\delta. \tag{2.6.20}$$

Since $E \notin \operatorname{supp}(\rho)$, the term $q_0(s) - E$ is invertible and by continuity in η we have

$$\|(q_0(s) - z \otimes \mathbb{1})^{-1}\| \lesssim_E 1 \tag{2.6.21}$$

uniformly in η for $0 < \eta \leq 1$. In particular, for η sufficiently small (depending on E), we have

$$\|(q_0(s) - z \otimes \mathbb{1})^{-1}(q_0(s) - q_\delta(s))\| \leq \frac{1}{2}. \tag{2.6.22}$$

Therefore the first factor in (2.6.19) factorizes further into

$$(q_\delta(s) - z \otimes \mathbb{1})^{-1} = (1 \otimes \mathbb{1} + (q_0(s) - z \otimes \mathbb{1})^{-1}(q_0(s) - q_\delta(s)))^{-1}(q_0(s) - z \otimes \mathbb{1})^{-1} \tag{2.6.23}$$

and

$$\|(1 \otimes \mathbb{1} + (q_0(s) - z \otimes \mathbb{1})^{-1}(q_0(s) - q_\delta(s)))^{-1}\| \leq 2. \tag{2.6.24}$$

Combining (2.6.19) with (2.6.20), (2.6.21), (2.6.23) and (2.6.24), we find

$$|(M_\delta)_{11} - m| \lesssim_E \delta\eta. \tag{2.6.25}$$

As shown in the first part of the proof, there are $u, u_1 > 0$ such that (2.6.13) has a unique solution m_δ with positive imaginary part and $|m_\delta(z) - m_\pm| \leq u_1$ for all $z \in \mathbb{H}$ such that $|z - \tau_\pm| \leq u$. For such u, u_1 we choose a u_2 with $0 < u_2 < u$ such that $|m_z - m_\pm| \leq \frac{u_1}{2}$ for all $z \in \mathbb{H}$ with $|z - \tau_\pm| \leq u_2$ (such a u_2 exists by Corollary 2.3.3). Fix $E \notin \operatorname{supp}(\rho)$ with $|E - \tau_\pm| < u_2$. We choose $\eta = \operatorname{Im} z$ sufficiently small such that $|z - \tau_\pm| < u_2$ and $|(M_\delta)_{11} - m| \lesssim \frac{u_1}{2}$ (which is possible by (2.6.25)). Thus there is a $z \in \mathbb{H}$ with $|z - \tau_\pm| \leq u$ such that

$$|(M_\delta)_{11}(z) - m_\pm| \leq |(M_\delta)_{11}(z) - m(z)| + |m(z) - m_\pm| \leq u_1. \tag{2.6.26}$$

Since this condition is unique among solutions to (2.4.20) with positive imaginary part, we must have $(M_\delta)_{11}(z) = m_\delta(z)$ for some $z \in \mathbb{H}$ with $|z - \tau_\pm| \leq u$. By continuity of both m_δ and $(M_\delta)_{11}$ we must have $(M_\delta)_{11}(z) = m_\delta(z)$ for all $z \in \mathbb{H}$ with $|z - \tau_\pm| \leq u$.

For $C > 0$ and E with $C^{-1} < \operatorname{dist}(E, \operatorname{supp}(\rho)) < C$ we find similarly from Lemma 2.6.3 and (2.6.25) that for all $u_1 > 0$ there is an $\eta > 0$ such that

$$|(M_\delta)_{11}(z) - m(E)| \leq |(M_\delta)_{11}(z) - m(z)| + |m(z) - m(E)| \leq u_1. \tag{2.6.27}$$

Thus $(M_\delta)_{11}(z)$ must be the unique solution to (2.4.20) that satisfies (2.4.23), which concludes the proof of the lemma. □

Chapter 3

Examples

In this chapter, we present a number of examples of quadratic polynomials to illustrate some interesting features in their limiting density. Recall that a self-adjoint quadratic polynomial q in l Wigner matrices $\mathbf{X} = (\mathbf{X}_i)_{i\in[\![l]\!]}$ is of the form

$$q(\mathbf{X}_1, \dots, \mathbf{X}_l) = \sum_{i,j=1}^{l} \mathbf{X}_i A_{ij} \mathbf{X}_j + \sum_{i=1}^{l} b_i \mathbf{X}_i + c. \tag{3.0.1}$$

The Stieltjes transform m of its limiting density is uniquely defined by Proposition 2.2.3. In particular, m satisfies the self-consistent equation

$$-m^{-1} = z - \operatorname{Tr} A(I_l + Am)^{-1} + mb^t \left((I_l + 2m\widehat{A})^{-2} (I_l + m\widehat{A}) \right) b - c. \tag{3.0.2}$$

One notable feature of the equation is that for homogeneous polynomials, that is polynomials for which all first and zeroth order coefficients are vanishing, m only depends on the eigenvalues of A, not however on its eigenvectors. This can be seen very well when comparing the polynomials

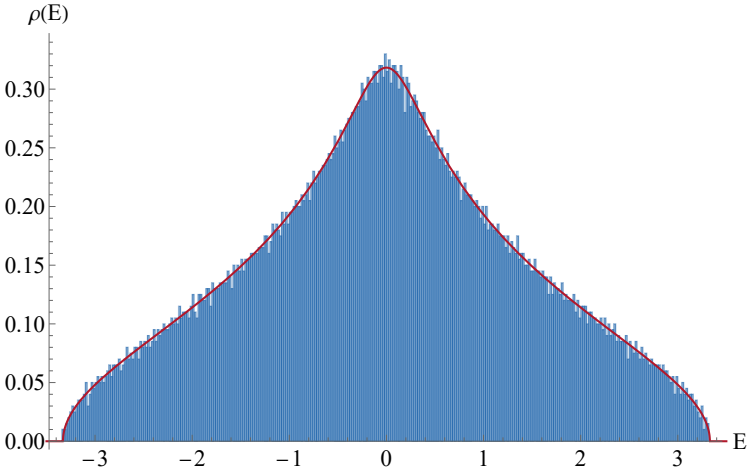
$$q_1(\mathbf{X}_1, \mathbf{X}_2) := \mathbf{X}_1\mathbf{X}_2 + \mathbf{X}_2\mathbf{X}_1 + r\mathbf{X}_1 \quad \text{and} \quad q_2(\mathbf{X}_1, \mathbf{X}_2) := \mathrm{i}(\mathbf{X}_1\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1) + r\mathbf{X}_1. \tag{3.0.3}$$

with $r \in \mathbb{R}$. The respective matrices of the second-order coefficients are given by

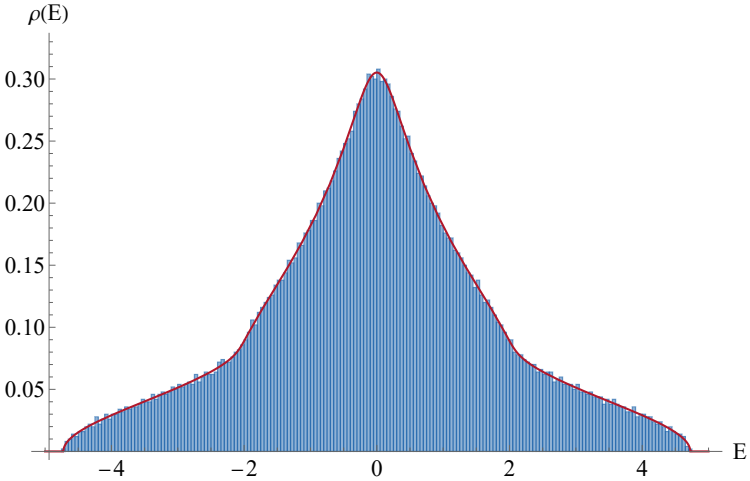
$$A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}. \tag{3.0.4}$$

They both have eigenvalues ± 1 and therefore have the same limiting density for $r = 0$. For $r \neq 0$, they are visibly different, however. We display the eigenvalue density of p_1 for $r = 0$ and $r = 1$ in Figure 3.1 and of p_2, also for $r = 0$ and $r = 1$, in Figure 3.2.

Remark. *All plots in this chapter show the limiting density, obtained by solving (3.0.2) for $m(z)$ and then taking the limit $\rho(E) = \lim_{\eta \searrow 0} \frac{1}{\pi} \operatorname{Im} m(E + \mathrm{i}\eta)$. The limiting density is shown as a red line. The histograms additionally displayed in the plots have been obtained by a single sampling of the respective polynomials of GOE Wigner matrices of dimension $N = 10^4$.*



(a) $r = 0$



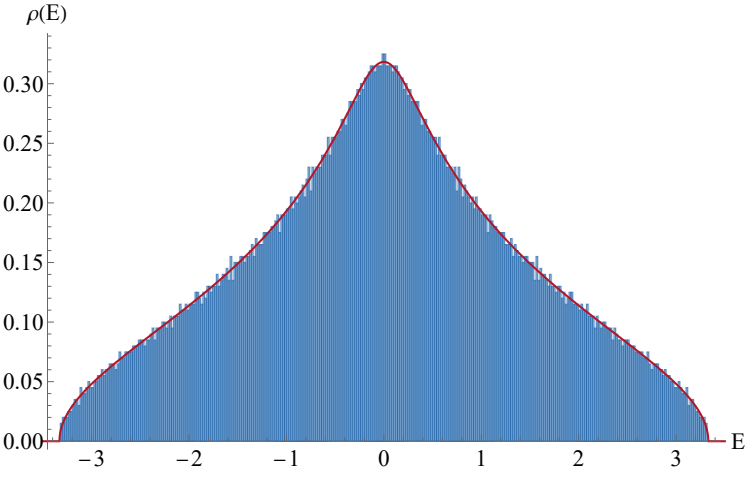
(b) $r = 1$

Figure 3.1: Eigenvalue distribution for $q_1(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{X}_1\mathbf{X}_2 + \mathbf{X}_2\mathbf{X}_1 + r\mathbf{X}_1$

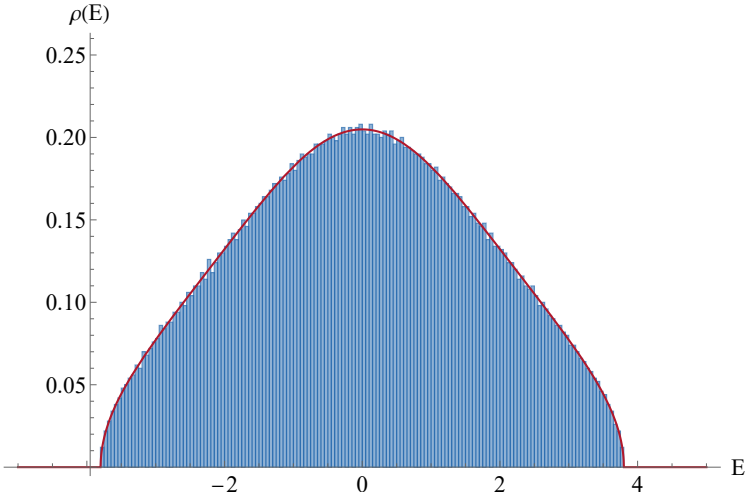
Next, we consider reducible polynomials. Recall that reducible polynomials are, up to a shift and scaling, of the form

$$q_r = (\mathbf{Y} - \xi)(\mathbf{Y} - \xi)^* \tag{3.0.5}$$

where $\xi \geq 0$ is a positive real number and $\mathbf{Y} = v^*\mathbf{X} \in \mathbb{C}^{N\times N}$, for some normalized $v \in \mathbb{C}^l$, see also Definition 2.2.1. In Proposition 2.2.8, we have proved that reducible polynomials have hard edges if $\xi \leq \xi_c$ for $\xi_c = 2s^{-1}$ and s was defined in (2.2.11). For $\xi < \xi_c$, the density ρ has an inverse square root singularity at its hard edge and for $\xi > \xi_c$, the density has two regular edges. At $\xi = \xi_c$, a different behaviour can be seen depending on if $\mathbf{Y}$ is Hermitian or non-Hermitian. If $\mathbf{Y}$ is Hermitian, then we have $\rho(E) \sim E^{1/4}$ for $E \searrow 0$ and if $\mathbf{Y}$ is non-Hermitian, then we have $\rho(E) \sim E^{1/3}$



(a) $r = 0$



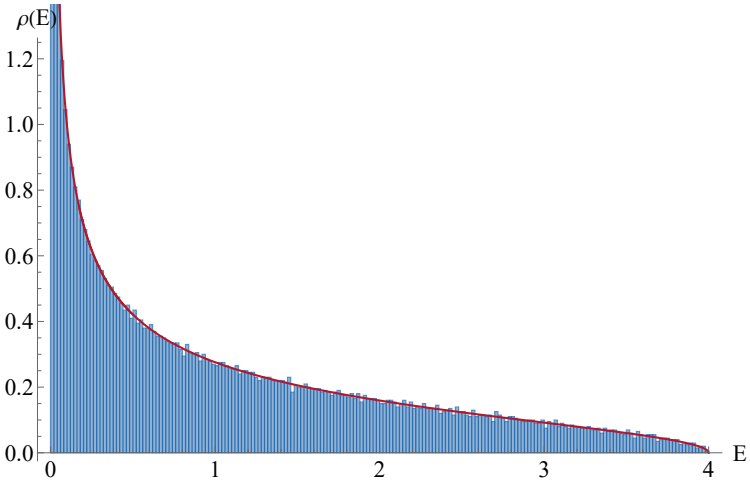
(b) $r = 1$

Figure 3.2: Eigenvalue distribution for $q_2(\mathbf{X}_1, \mathbf{X}_2) = \mathrm{i}(\mathbf{X}_1\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1) + r\mathbf{X}_1$

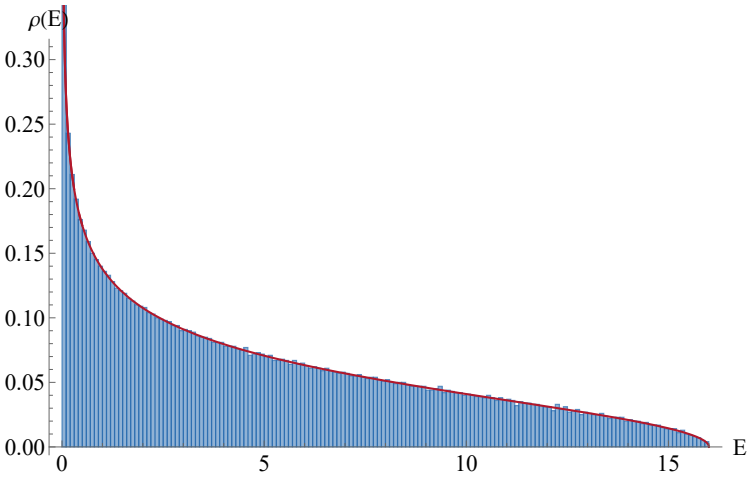
for $E \searrow 0$. This difference can be seen for the polynomials

$$\begin{aligned} q_3(\mathbf{X}) &:= (\mathbf{X} - \xi)^2 \\ q_4(\mathbf{X}_1, \mathbf{X}_2) &:= \left(\sqrt{\frac{4}{5}}\mathbf{X}_1 + \mathrm{i}\sqrt{\frac{1}{5}}\mathbf{X}_2 - \xi\right)\left(\sqrt{\frac{4}{5}}\mathbf{X}_1 + \mathrm{i}\sqrt{\frac{1}{5}}\mathbf{X}_2 - \xi\right)^*. \end{aligned} \tag{3.0.6}$$

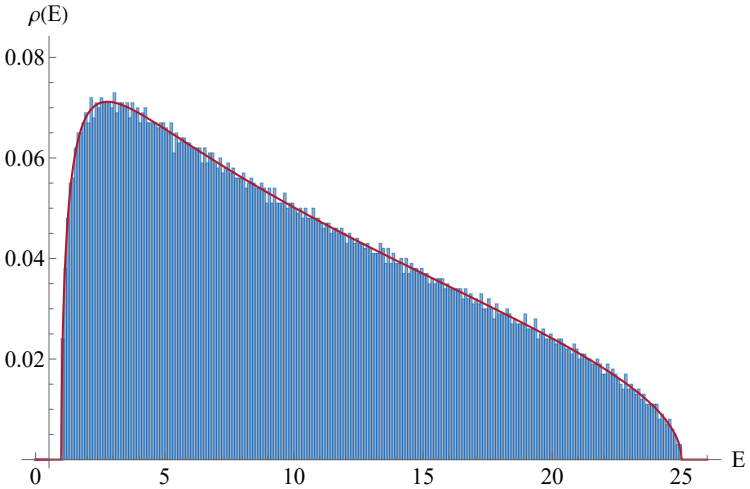
The matrix $\mathbf{Y}$ for the polynomial q_3 is clearly Hermitian, whereas in the case of q_4, $\mathbf{Y}$ is a (non-Hermitian) elliptic matrix (see e.g. [39]). For q_3, the critical value is given by $\xi_c = 2$ and for q_4, we have $\xi_c = \frac{8}{5}$. In Figure 3.3 and Figure 3.4, we display the eigenvalue density of q_3 and q_4 respectively for different values of ξ. It is clearly visible, that the eigenvalue density of q_4 has a stronger singularity at 0 for $\xi = \xi_c$ than q_3, while for $\xi \neq \xi_c$, both eigenvalue densities show the same asymptotic behaviour.



(a) $\xi = 0$

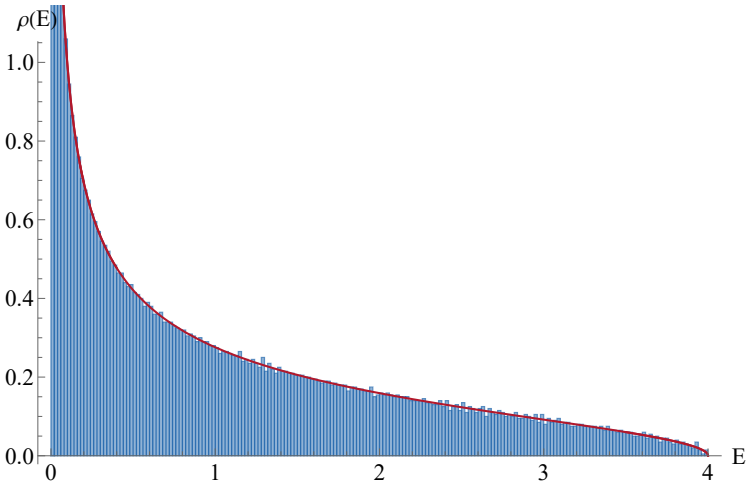


(b) $\xi = \xi_c$

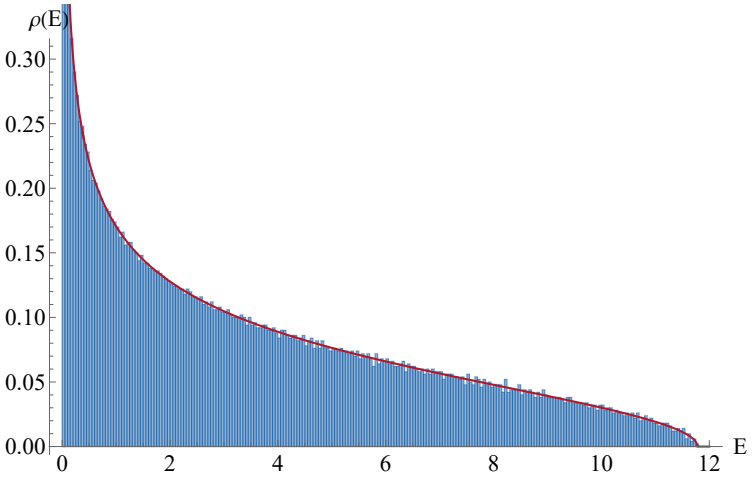


(c) $\xi > \xi_c$

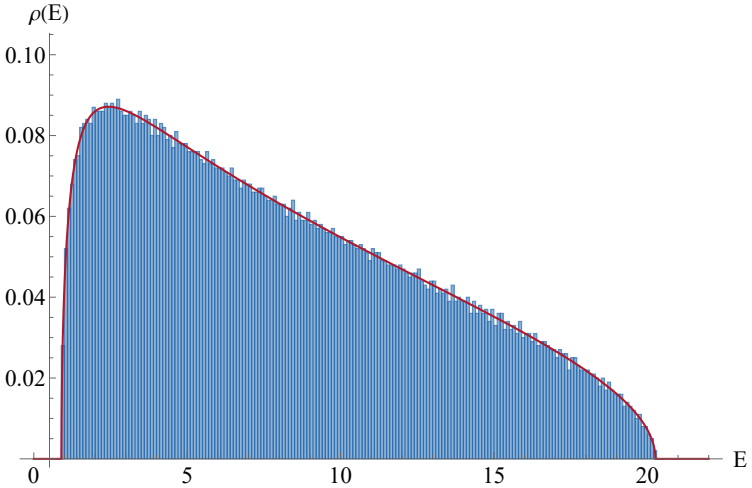
Figure 3.3: Eigenvalue distribution for $q_3(\mathbf{X}) = (\mathbf{X} - \xi)^2$



(a) $\xi = 0$



(b) $\xi = \xi_c$



(c) $\xi > \xi_c$

Figure 3.4: Eigenvalue distribution for
$q_4(\mathbf{X}_1, \mathbf{X}_2) = \left(\sqrt{\frac{4}{5}}\mathbf{X}_1 + \mathrm{i}\sqrt{\frac{1}{5}}\mathbf{X}_2 - \xi\right)\left(\sqrt{\frac{4}{5}}\mathbf{X}_1 + \mathrm{i}\sqrt{\frac{1}{5}}\mathbf{X}_2 - \xi\right)^*$

Chapter 4

Introduction to Non-Hermitian Random Matrix Theory

So far we have only discussed ensembles of Hermitian random matrices. An important feature of them is that their eigenvalues are confined to the real line. If we drop the Hermitian symmetry constraint, this is no longer the case and in general, we expect to see eigenvalues on the whole of the complex plane. As a consequence, the resolvent method as described in Section 1.1 is no longer directly applicable as the method relied on being able to control the distance between the spectral parameter and the spectrum, which is always bounded from below by the imaginary part of the spectral parameter. This makes understanding the spectral statistics of non-Hermitian random matrices a much harder endeavour compared to their Hermitian counterpart. Despite this, a lot of progress has been made over the last few years. Here, we introduce a method that allows us to extract information about the spectrum of a non-Hermitian matrix by instead studying a larger dimensional Hermitian block matrix. The method is originally due to Girko and it has been used by Tao and Vu to prove a global law for matrices with centered i.i.d. entries. We refer to such matrices as **Girko matrices** and more precisely, we define them as follows.

Definition 4.0.1 (Girko matrix)**.** *Let $\mathbf{X} \in \mathbb{C}^{N\times N}$ be a random matrix whose entries are i.i.d. copies of $\frac{1}{\sqrt{N}}\zeta$, where ζ is a random variable satisfying*

$$\mathbb{E}[\zeta] = 0, \quad \textit{and} \quad \mathbb{E}[|\zeta|^2] = 1. \tag{4.0.1}$$

Furthermore, ζ satisfies the exponential decay property

$$\mathbb{P}\left(|\zeta| > \lambda\right) \leq \frac{1}{\vartheta}\mathrm{e}^{-\vartheta\lambda} \tag{4.0.2}$$

for all $\lambda > 0$ and some $\vartheta > 0$. Then $\mathbf{X}$ is called a Girko matrix.

Remark. *Some of the statements covered here also hold under weaker assumptions than (4.0.2). However, for the purpose of this introduction, we will not worry about the best possible moment assumptions.*

If the entries of a Girko matrix follow a complex Gaussian distribution, the joint probability density function of the eigenvalues can be explicitly calculated and it is given by

$$\varphi_N(z_1,\dots,z_N) = \frac{1}{Z}\left(\prod_{1\le i<j\le N}|z_i-z_j|^2\right)\exp\left(-\sum_{k=1}^{N}|z_k|^2\right), \tag{4.0.3}$$

where Z is a normalizing constant. This was first done by Ginibre [37] and Girko matrices with (real or complex) Gaussian entries are therefore also called **(real or complex) Ginibre matrices**. Ginibre also showed that the limiting density σ_{c} for complex Ginibre matrices is given by the uniform distribution on the unit disc, i.e.

$$\sigma_{\mathrm{c}}(z) := \frac{1}{\pi}\begin{cases}1 & \text{if } |z|\le 1\\ 0 & \text{if } |z|>1.\end{cases} \tag{4.0.4}$$

For more general distributions there is no explicit formula for the joint probability density function for finite N. Nonetheless, we still expect the eigenvalues to converge to the same limiting distribution that only depends on the model data in the form of entry variance and expectation value but not on the specific distribution of the entries. This result was first proven under optimal conditions on the global scale by Tao and Vu in [61] and on an optimal local scale in a series of papers by Bourgade, Yau and Yin [20, 21, 70]. For an overview of the circular law, we refer to [18]. The following form of the local circular law is given in [70].

Theorem 4.0.2 (The local circular law)**.** *Let $\mathbf{X}\in\mathbb{C}^{N\times N}$ be a Girko matrix with eigenvalues μ_j, $j\in[\![N]\!]$. Furthermore, let $z_0\in\mathbb{C}$, $\alpha\in(0,\frac{1}{2})$, $f\in C_0^2(\mathbb{C})$ and $f_{N,z_0}(z)=N^{2\alpha}f(N^{\alpha}(z-z_0))$. Then we have the local law*

$$\left|\frac{1}{N}\sum_{j=1}^{N}f_{N,z_0}(\mu_j)-\int_{\mathbb{C}}f_{N,z_0}(z)\sigma_{\mathrm{c}}(z)\mathrm{d}^2z\right| \prec \frac{\|\Delta f\|_{L_1}}{N^{1-2\alpha}}$$

uniformly in z_0 and f.

4.1 Proof strategy of Theorem 4.0.2

The main idea to prove Theorem 4.0.2 is to replace the matrix $\mathbf{X}-z$ with a related Hermitian matrix. This method is known as **Girko's Hermitization trick** and was first introduced in [38]. There are two standard procedures to Hermitizise $\mathbf{X}-z$. One consists of studying

$$\widetilde{\mathbf{H}}^z := (\mathbf{X}-z)(\mathbf{X}^*-\bar{z})\in\mathbb{C}^{N\times N} \tag{4.1.1}$$

and the other of studying

$$\mathbf{H}^z := \begin{pmatrix}0 & \mathbf{X}-z\\ \mathbf{X}^*-\bar{z} & 0\end{pmatrix}\in\mathbb{C}^{2N\times 2N}. \tag{4.1.2}$$

Here, $\mathbf{X}^*$ denotes the adjoint matrix of $\mathbf{X}$. Bourgade, Yau and Yin used (4.1.1) in their original proof of the local circular law. We, however, will utilize (4.1.2). To our knowledge this Hermitization method was first used in [33] and our explanations follow the overview given in [1]. We will only discuss the bulk case, i.e. throughout this section we assume that

$$z_0 \in \mathbb{D}_< := \{z \in \mathbb{C} : |z| \leq 1 - \tau_*\} \tag{4.1.3}$$

for some fixed $\tau_* > 0$. The constants in our bounds will in general depend on τ_*. Furthermore, we will assume that the entries of $\mathbf{X}$ are absolutely continuous and their density satisfies the following assumption.

Assumption 4.1.1. *Let $\mathbf{X} \in \mathbb{C}^{N\times N}$ be a Girko matrix whose entries are i.i.d. copies of $\frac{1}{\sqrt{N}}\zeta$. We assume that ζ has a density, i.e. there is a probability density function $p : \mathbb{C} \to [0, \infty)$ such that*

$$\mathbb{P}(\zeta \in B) = \int_B p(z)\mathrm{d}^2 z \tag{4.1.4}$$

for all Borel sets $B \subset \mathbb{C}$. Also, we assume there are $\alpha, \beta > 0$ such that $p \in L^{1+\alpha}(\mathbb{C})$ and

$$\|p\|_{L^{1+\alpha}} \lesssim N^{\beta}. \tag{4.1.5}$$

The central observation by Girko was that the first term on the left-hand side of (4.0.2) can be expressed in terms of the log-determinant of $\mathbf{X} - z\mathbf{I}_N$ for $z \in \mathbb{C}$. Namely, we have

$$\frac{1}{N}\sum_{j=1}^{N} f_{N,z_0}(\mu_j) = \frac{1}{2\pi N}\int_{\mathbb{C}} \Delta f_{N,z_0}(z) \log|\det(\mathbf{X} - z\mathbf{I}_N)|\mathrm{d}^2 z. \tag{4.1.6}$$

The above equation holds true since

$$\frac{1}{2\pi}\log|\det(\mathbf{X} - z\mathbf{I}_N)| = \frac{1}{2\pi}\sum_{i=1}^{N}\log(|\mu_i - z|) \tag{4.1.7}$$

and $(2\pi)^{-1}\log|\cdot|$ is the Green's function of the Laplace operator in two dimensions. The log determinant of $\mathbf{X} - z$ and the one of its Hermitization, $\mathbf{H}^z$, which was defined in (4.1.2), are connected by

$$\log|\det(\mathbf{X} - z)| = \frac{1}{2}\log|\det(\mathbf{H}^z)|. \tag{4.1.8}$$

From here on for the rest of this thesis, we no longer use η for the imaginary part of z. Instead, $\eta > 0$ denotes a positive real number independent of z. Let $\mathbf{G} \in \mathbb{C}^{2N\times 2N}$ denote the resolvent of $\mathbf{H}^z$ at spectral parameter $\mathrm{i}\eta$ for $\eta > 0$, i.e.

$$\mathbf{G} = \mathbf{G}^z(\mathrm{i}\eta) := (\mathbf{H}^z - \mathrm{i}\eta)^{-1}. \tag{4.1.9}$$

For all $T > 0$, the log-determinant of $\mathbf{H}^z$ and its resolvent are connected by

$$\log|\det(\mathbf{H}^z)| = \log|\det(\mathbf{H}^z - \mathrm{i}T)| - 2N\int_0^T \mathrm{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle \mathrm{d}\eta. \tag{4.1.10}$$

This equality can be seen by evaluating the integral on the right hand side. Combining (4.1.6), (4.1.8) and (4.1.10), we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N f_{N,z_0}(\mu_i) = \frac{1}{4\pi N}\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\log|\det(\mathbf{H}^z - \mathrm{i}T)|\mathrm{d}^2 z \\ - \frac{1}{2\pi}\int_{\mathbb{C}}\Delta f_{N,z_0}(z)\int_0^T \mathrm{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle \mathrm{d}\eta \mathrm{d}^2 z.\end{aligned} \tag{4.1.11}$$

Now, since $\mathbf{G}$ is the resolvent of a Hermitian matrix with independent entries, it is known that it satisfies the following averaged local law uniformly in $z \in \mathbb{D}_<$ and $N^c > \eta > N^{-1+\varepsilon}$ for all $c, \varepsilon > 0$:

$$|\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle - m^z(\mathrm{i}\eta)| \prec \begin{cases} \frac{1}{N\eta} & \text{if } N^{-1+\varepsilon} \le \eta \le 1 \\ \frac{1}{N\eta^2} & \text{if } 1 \le \eta \le N^c. \end{cases} \tag{4.1.12}$$

Here, $m^z(\mathrm{i}\eta)$ is the unique solution with positive imaginary part to the scalar Dyson equation

$$-\frac{1}{m^z(\mathrm{i}\eta)} = \mathrm{i}\eta + m^z(\mathrm{i}\eta) - \frac{|z|^2}{\mathrm{i}\eta + m^z(\mathrm{i}\eta)}. \tag{4.1.13}$$

For a proof of this local law see e.g. [1, Theorem 5.2] or [2, Theorem 5.2]. The solution m^z is connected to the limiting spectral measure (4.0.4) by

$$\sigma_{\mathrm{c}}(z) = -\frac{1}{2\pi}\int_0^\infty \Delta_z(\mathrm{Im}\, m^z(\mathrm{i}\eta) - h(\eta))\mathrm{d}\eta \tag{4.1.14}$$

for any function h that is independent of z (see e.g. Equation (4.8) combined with Equation (3.5) in [1]). For $T > 0$ we choose

$$h(\eta) = \chi_{[T,\infty)}(\eta)\frac{1}{\eta}, \tag{4.1.15}$$

where $\chi_{[T,\infty)}$ denotes the indicator function on the interval $[T, \infty)$. Since $\mathrm{Im}\, m^z(\mathrm{i}\eta) = \eta^{-1} + \mathcal{O}(\eta^{-2})$ for $\eta \to \infty$, this choice of h ensures that $\mathrm{Im}\, m^z(\mathrm{i}\eta) - h(\eta) = \mathcal{O}(\eta^{-2})$ for $\eta \to \infty$. Therefore, we can use Green's identity to obtain

$$\int_{\mathbb{C}} f_{N,z_0}(z)\sigma_{\mathrm{c}}(z)\mathrm{d}^2 z = -\frac{1}{2\pi}\int_{\mathbb{C}}\Delta f_{N,z_0}(z)\int_0^\infty (\mathrm{Im}\, m^z(\mathrm{i}\eta) - h(\eta))\,\mathrm{d}\eta \mathrm{d}^2 z. \tag{4.1.16}$$

Combining (4.1.11) and (4.1.16) we find

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N} f_{N,z_0}(\mu_j) - \int_{\mathbb{C}} f_{N,z_0}(z)\sigma_{\mathrm{c}}(z)\mathrm{d}^2 z \\
=&\frac{1}{4\pi N}\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\log\left|\det(\mathbf{H}^z - \mathrm{i}T)\right|\mathrm{d}^2 z \\
&-\frac{1}{2\pi}\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\int_0^T \left(\mathrm{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle - \mathrm{Im}\, m^z(\mathrm{i}\eta)\right)\mathrm{d}\eta\mathrm{d}^2 z \\
&+\frac{1}{2\pi}\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\int_T^\infty \left(\mathrm{Im}\, m^z(\mathrm{i}\eta) - \frac{1}{\eta}\right)\mathrm{d}\eta\mathrm{d}^2 z.
\end{aligned} \tag{4.1.17}$$

It is straightforward to see that the first and third term on the right-hand side of (4.1.17) are stochastically dominated by $N^{-(1-2\alpha)}\|\Delta f\|_{L_1}$ for the choice $T = N^C$ for some large but fixed $C > 0$. We omit the details. The second term can be bounded with the aid of the following lemma.

Lemma 4.1.2 ([1, Lemma 5.8])**.** *Define $I(z)$ by*

$$I(z) := \int_0^T \left|\mathrm{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle - \mathrm{Im}\, m^z(\mathrm{i}\eta)\right|\mathrm{d}\eta. \tag{4.1.18}$$

For every $\delta > 0$ and $p \in \mathbb{N}$, we have the moment bound

$$\sup_{z\in\mathbb{D}_<} \mathbb{E}[I(z)^p] \lesssim_{\delta,p} N^{-p(1-\delta)}. \tag{4.1.19}$$

The lemma leads to the desired bound in (4.0.2) for the second term in (4.1.17) since

$$\begin{aligned}
&\mathbb{E}\left|\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\int_0^T \left(\mathrm{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle - \mathrm{Im}\, m^z(\mathrm{i}\eta)\right)\mathrm{d}\eta\mathrm{d}^2 z\right|^p \\
\le&\int_{\mathbb{C}^p}\prod_{j=1}^{p} |\Delta f_{N,z_0}(z_j)|\mathbb{E}[I(z)^p]^{\frac{1}{p}}\mathrm{d}^2 z_j \lesssim_{p,\delta} \|\Delta f\|_1^p N^{p(1-\delta-2\alpha)}.
\end{aligned} \tag{4.1.20}$$

Here, we have used Hölder's inequality in the first estimate and Lemma 4.1.2 in the second estimate. Using Chebyshev's inequality, we transform the high moment bound into the stochastic domination bound

$$\left|\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\int_0^T \left(\mathrm{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle - \mathrm{Im}\, m^z(\mathrm{i}\eta)\right)\mathrm{d}\eta\mathrm{d}^2 z\right| \prec \|\Delta f\|_1 N^{-1+2\alpha}. \tag{4.1.21}$$

To prove Lemma 4.1.2, we first split $I(z)$ into several terms. We denote the eigenvalues of $\mathbf{H}^z$ by λ_i, $i \in [\![2N]\!]$ and sort them in a non-increasing order, i.e. $\lambda_i \ge \lambda_{i+1}$. Due to the block structure of $\mathbf{H}^z$, its eigenvalues come in symmetric pairs around the origin, i.e. $\lambda_i = -\lambda_{2N+1-i}$ for all $i \in [\![N]\!]$. Furthermore, the positive eigenvalues

of $\mathbf{H}^z$ correspond exactly to the singular values of $\mathbf{X} - z$ and in particular we have $\lambda_N = \|(\mathbf{X} - z)^{-1}\|^{-1}$.

As a consequence of the eigenvalue symmetry we have

$$\int_0^x \operatorname{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle \mathrm{d}\eta = \frac{1}{2N}\sum_{i=1}^{n} \log\left(1 + \frac{x^2}{\lambda_i^2}\right) \tag{4.1.22}$$

for all $x \geq 0$. Splitting the η integral in $I(z)$ into $\eta \leq N^{-1+\varepsilon}$ and $\eta > N^{-1+\varepsilon}$ and using (4.1.22) we find for any $k \in \mathbb{N}$ that

$$\begin{aligned} I(z) \leq \frac{1}{2N}\sum_{\lambda_i \in [0, N^{-k}]} \log\left(1 + \frac{N^{-2+2\varepsilon}}{\lambda_i^2}\right) + \frac{1}{2N}\sum_{\lambda_i > N^{-k}} \log\left(1 + \frac{N^{-2+2\varepsilon}}{\lambda_i^2}\right) \\ + \int_0^{N^{-1+\varepsilon}} \operatorname{Im} m^z(\mathrm{i}\eta)\mathrm{d}\eta + \int_{N^{-1+\varepsilon}}^{T} \left|\operatorname{Im}\langle \mathbf{G}^z(\mathrm{i}\eta)\rangle - \operatorname{Im} m^z(\mathrm{i}\eta)\right| \mathrm{d}\eta. \end{aligned} \tag{4.1.23}$$

The exponent k will be fixed later. The moment bound (4.1.19) is proved separately for each term in (4.1.23). For the bounds of the second, third and fourth term in (4.1.23) we refer to the proof of Lemma 5.8 in [1], but we present the estimate of the first term, as it shows how the smallest singular value estimate of $\mathbf{X} - z$, presented in Proposition 4.1.3 below, enters the proof of the circular law. We estimate the p^{th} moment of the first term by

$$\begin{aligned} \mathbb{E}\left[\frac{1}{2N}\sum_{\lambda_i \in [0, N^{-k}]} \log\left(1 + \frac{N^{-2+2\varepsilon}}{\lambda_i^2}\right)\right]^p &\leq \mathbb{E}\left[\log^p\left(1 + \frac{N^{-2+2\varepsilon}}{\lambda_N^2}\right)\chi(\lambda_N \leq N^{-k})\right] \\ &\lesssim \mathbb{E}\left[|\log \lambda_N|^p \chi(\lambda_N \leq N^{-k})\right]. \end{aligned} \tag{4.1.24}$$

The expectation value of the last term is given by

$$\mathbb{E}\left[|\log \lambda_N|^p \chi(\lambda_N \leq N^{-k})\right] = p\int_{k\log N}^{\infty} \mathbb{P}(\lambda_N \leq \mathrm{e}^{-t}) t^{p-1}\mathrm{d}t. \tag{4.1.25}$$

This expression is estimated with the help of the following proposition.

Proposition 4.1.3 ([1, Proposition 5.7])**.** *Under Assumption 4.1.1 we have*

$$\mathbb{P}\left(\|(\mathbf{X} - z)^{-1}\|^{-1} \leq \frac{u}{N}\right) \lesssim_\alpha u^{\frac{2\alpha}{1+\alpha}} N^{\beta+1} \tag{4.1.26}$$

uniformly in $u > 0$ and $z \in \mathbb{C}$.

Using Proposition 4.1.3 for $u = \mathrm{e}^{-t}N$ to estimate (4.1.25) and consequently (4.1.24), we find

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{2N}\sum_{\lambda_i \in [0, N^{-k}]} \log\left(1 + \frac{N^{-2+2\varepsilon}}{\lambda_i^2}\right)\right]^p \\ &\lesssim_\alpha N^{\beta+1+\frac{2\alpha}{1+\alpha}} \int_{k\log N}^{\infty} e^{-\frac{2\alpha t}{1+\alpha}} t^{p-1}\mathrm{d}t \lesssim_{\alpha,\beta,p} N^{-p}, \end{aligned} \tag{4.1.27}$$

where the last step holds for sufficiently large k depending on p.

Remark. *In the general case, without Assumption 4.1.1, Proposition 4.1.3 is no longer valid. Instead, the following theorem by Tao and Vu is used.*

Theorem 4.1.4 ([61, Theorem 2.1])**.** *Let A, C_1 be positive constants, and let w be a complex-valued random variable with non-zero finite variance (in particular, the second moment is finite). Then there are positive constants B and C_2 such that the following holds: if $\mathbf{W} \in \mathbb{C}^{N\times N}$ is the random matrix of order N whose entries are i.i.d. copies of w, and $\mathbf{M} \in \mathbb{C}^{N\times N}$ is a deterministic matrix of order N with spectral norm at most N^{C_1} , then,*

$$\mathbb{P}\left(\|(\mathbf{M}+\mathbf{W})^{-1}\| \geq N^B\right) \leq C_2 N^{-A}. \tag{4.1.28}$$

It requires a different technique to estimate $I(z)$ though, since Theorem 4.1.4 is not sufficient to estimate (4.1.25).

Chapter 5

The Spherical Ensemble

5.1 Introduction

In the last chapter, we have been studying the eigenvalue distribution of Girko matrices $\mathbf{X}$. This is equivalent to asking about the distribution of solutions to the eigenvalue equation $\det(\mathbf{X} - z\mathbf{I}_N) = 0$ for $z \in \mathbb{C}$. A natural continuation of that question is to replace the identity matrix with another Girko matrix $\mathbf{Y}$, independent of $\mathbf{X}$, and ask for the distribution of solutions to the equation

$$\det(\mathbf{X} - z\mathbf{Y}) = 0. \tag{5.1.1}$$

Equations of this kind are known as **generalized eigenvalue equations**. If $\mathbf{Y}$ is invertible, which is the case with high probability if $\mathbf{Y}$ is a Girko matrix (see Theorem 4.1.4), then $z \in \mathbb{C}$ being a solution to (5.1.1) is equivalent to z being an eigenvalue $\mathbf{Y}^{-1}\mathbf{X}$. This is a simple example of a multivariate rational function of non-Hermitian random matrices and we explore its eigenvalue density throughout this chapter.

We denote the eigenvalues of $\mathbf{Y}^{-1}\mathbf{X}$ by μ_i, $i \in [\![N]\!]$ and its empirical eigenvalue distribution by σ_N, i.e.

$$\sigma_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\mu_i}. \tag{5.1.2}$$

First, we aim to find a candidate for the limiting spectral measure of $\mathbf{Y}^{-1}\mathbf{X}$. To do so, consider $\mathbf{X}^{\mathrm{G}}, \mathbf{Y}^{\mathrm{G}}$ to be complex Ginibre matrices and denote the eigenvalues of $(\mathbf{Y}^{\mathrm{G}})^{-1}\mathbf{X}^{\mathrm{G}}$ by μ_i^{G} and its empirical eigenvalue distribution by σ_N^{G}. Krishnapur made the observation in [48] that we have

$$\mathbb{E}[\sigma_N^{\mathrm{G}}] = \sigma_{\mathrm{s}} \tag{5.1.3}$$

for all $N \in \mathbb{N}$. Here, σ_{s} denotes the absolutely continuous distribution with density

$$\sigma_{\mathrm{s}}(z) := \frac{1}{\pi}\frac{1}{(1+|z|^2)^2} \tag{5.1.4}$$

and we briefly present Krishnapur's argument. Note that σ_{s} corresponds exactly to the uniform distribution under projection onto the Riemann sphere. Rotations on the sphere are given by the Möbius transformations of the type

$$z \mapsto \frac{z\alpha - \bar{\beta}}{z\beta - \bar{\alpha}} \tag{5.1.5}$$

with $\alpha, \beta \in \mathbb{C}$ such that $|\alpha|^2 + |\beta|^2 = 1$ and we show that the distribution of eigenvalues of $(\mathbf{Y}^{\mathrm{G}})^{-1}\mathbf{X}^{\mathrm{G}}$ is invariant under (5.1.5). To do so define $\widetilde{\mathbf{X}}^{\mathrm{G}}, \widetilde{\mathbf{Y}}^{\mathrm{G}}$ by

$$\widetilde{\mathbf{X}}^{\mathrm{G}} := \alpha\mathbf{X}^{\mathrm{G}} - \beta\mathbf{Y}^{\mathrm{G}} \quad \text{and} \quad \widetilde{\mathbf{Y}}^{\mathrm{G}} := \bar{\beta}\mathbf{X}^{\mathrm{G}} - \bar{\alpha}\mathbf{Y}^{\mathrm{G}}. \tag{5.1.6}$$

Then, $(\widetilde{\mathbf{X}}^{\mathrm{G}}, \widetilde{\mathbf{Y}}^{\mathrm{G}})$ has the same distribution as $(\mathbf{X}^{\mathrm{G}}, \mathbf{Y}^{\mathrm{G}})$ since $\mathbf{X}^{\mathrm{G}}, \mathbf{Y}^{\mathrm{G}}$ are Ginibre matrices. Therefore, the solutions to $\det\left(\mathbf{X}^{\mathrm{G}} - z\mathbf{Y}^{\mathrm{G}}\right) = 0$ have the same distribution as the solutions to $\det\left(\widetilde{\mathbf{X}}^{\mathrm{G}} - z\widetilde{\mathbf{Y}}^{\mathrm{G}}\right) = 0$. At the same time we have

$$\begin{aligned} \det\left(\widetilde{\mathbf{X}}^{\mathrm{G}} - z\widetilde{\mathbf{Y}}^{\mathrm{G}}\right) &= \det\left((z\alpha - \bar{\beta})\mathbf{X}^{\mathrm{G}} - (z\beta - \bar{\alpha})\mathbf{Y}^{\mathrm{G}}\right) \\ &= (z\beta - \bar{\alpha})^N \det\left(\frac{z\alpha - \bar{\beta}}{z\beta - \bar{\alpha}}\mathbf{X}^{\mathrm{G}} - \mathbf{Y}^{\mathrm{G}}\right). \end{aligned} \tag{5.1.7}$$

Therefore, the zeros of $\det\left(\widetilde{\mathbf{X}}^{\mathrm{G}} - z\widetilde{\mathbf{Y}}^{\mathrm{G}}\right)$ are precisely given by the set

$$\left\{\frac{\mu_i^{\mathrm{G}}\alpha - \bar{\beta}}{\mu_i^{\mathrm{G}}\beta - \bar{\alpha}}\right\}_{i \in [\![N]\!]} \tag{5.1.8}$$

and it has the same distribution as $\{\mu_i^{\mathrm{G}}\}_{i \in [\![N]\!]}$. Since the rotation was chosen arbitrarily, (5.1.3) follows. Beyond the expectation value of the empirical measure, Krishnapur also calculated the exact joint probability density function φ_N of the eigenvalues in the Ginibre case. He showed that

$$\varphi_N(z_1, \ldots, z_N) = \frac{1}{Z}\left(\prod_{1 \le i < j \le N} \|\varphi(z_i) - \varphi(z_j)\|_{\mathbb{R}^3}^2\right), \tag{5.1.9}$$

where Z denotes a normalization constant, φ the projection onto the Riemann sphere and $\|\cdot\|_{\mathbb{R}^3}$ denotes the euclidean distance on the sphere as a subset of $\mathbb{R}^3$.

As in the other ensembles studied before, we expect that the equality in expectation, (5.1.3) also holds in a sense of weak convergence in probability for $N \to \infty$ and that this limit does not depend on the precise entry distribution. Indeed both of these claims were proved by Bordenave, who established a global law under optimal assumptions in [17]. As the empirical eigenvalue distribution of $\mathbf{Y}^{-1}\mathbf{X}$ tends to the uniform distribution after projection on the sphere, the ensemble is known as the **spherical ensemble**.

5.2 The local law for the spherical ensemble

It turns out that the convergence of σ_N to σ_s does not only hold on a global scale but also on all local scales above the typical eigenvalue spacing. This **local spherical law** is stated and proven under Assumption 4.1.1 in this section.

Theorem 5.2.1 (The local spherical law)**.** *Let $C > 0$ be a large but fixed number, $\alpha \in (0, \frac{1}{2})$ and let $\mathbf{X}, \mathbf{Y} \in \mathbb{C}^{N\times N}$ be two independent Girko matrices satisfying assumption 4.1.1 and let μ_j, $j \in \mathbb{N}$ denote the eigenvalues of $\mathbf{Y}^{-1}\mathbf{X}$. For $z_0 \in \mathbb{C}$, $|z_0| \le N^C$, $f \in C_0^2(\mathbb{C})$, we define f_{N,z_0} by $f_{N,z_0}(z) = N^{2\alpha} f(N^\alpha(z - z_0))$. Then we have the local law*

$$\left| \frac{1}{N}\sum_{j=1}^{N} f_{N,z_0}(\mu_j) - \int_{\mathbb{C}} f_{N,z_0}(z)\sigma_\mathrm{s}(z)\mathrm{d}^2 z \right| \prec \frac{\|\Delta f\|_{L_1}}{N^{1-2\alpha}} \tag{5.2.1}$$

uniformly in z_0 and f.

Proof. As in (4.1.6), we express the eigenvalues of $\mathbf{Y}^{-1}\mathbf{X}$ in terms of an integral over its log-determinant by

$$\frac{1}{N}\sum_{j=1}^{N} f_{N,z_0}(\mu_j) = \frac{1}{2\pi N}\int_{\mathbb{C}} \Delta f_{N,z_0}(z) \log\left|\det\left(\mathbf{Y}^{-1}\mathbf{X} - z\right)\right| \mathrm{d}^2 z. \tag{5.2.2}$$

From the multiplicativity of the determinant we have

$$\begin{aligned} &\log|\det\left(\mathbf{Y}^{-1}\mathbf{X} - z\right)| \\ &= \log\left(|\det(\mathbf{Y})^{-1}||\det(\mathbf{X} - z\mathbf{Y})|\right) = \log(|\det(\mathbf{X} - z\mathbf{Y})|) - \log(|\mathbf{Y}|). \end{aligned} \tag{5.2.3}$$

The crucial observation is that the second term does not depend on z. Integrating it against $\Delta f_{N,z_0}$ for $f \in C_0^2$ in (5.2.2) yields therefore no contribution. Next, we define $\mathbf{H}_\mathrm{s}^z \in \mathbb{C}^{2N\times 2N}$, the Hermitization of $\mathbf{X} - z\mathbf{Y}$, by

$$\mathbf{H}_\mathrm{s}^z = \begin{pmatrix} 0 & \mathbf{X} - z\mathbf{Y} \\ \mathbf{X}^* - \bar{z}\mathbf{Y}^* & 0 \end{pmatrix}. \tag{5.2.4}$$

Then

$$\log(|\det(\mathbf{X} - z\mathbf{Y})|) = \frac{1}{2}\log\left(|\mathbf{X} - \bar{z}\mathbf{Y}|^2\right) = \frac{1}{2}\log|\det \mathbf{H}_\mathrm{s}^z|. \tag{5.2.5}$$

We define the resolvent of $\mathbf{H}_\mathrm{s}^z$ at spectral parameter $\mathrm{i}\eta$, $\eta > 0$ by $\mathbf{G}_\mathrm{s}^z(\mathrm{i}\eta) := (\mathbf{H}_\mathrm{s}^z - \mathrm{i}\eta)^{-1}$. Analogous to (4.1.10), we have

$$\log|\det \mathbf{H}_\mathrm{s}^z| = \log|\det(\mathbf{H}_\mathrm{s}^z - \mathrm{i}T)| - 2N\int_0^T \langle \operatorname{Im} \mathbf{G}_\mathrm{s}^z(\mathrm{i}\eta)\rangle \mathrm{d}\eta \tag{5.2.6}$$

for all $T > 0$.

By assumption, $\mathbf{X}$ and $\mathbf{Y}$ are independent Girko matrices. Therefore

$$\mathbf{Z} := v(z)^{-1/2}(\mathbf{X} - z\mathbf{Y}), \quad \text{with} \quad v(z) := (1 + |z|^2) \tag{5.2.7}$$

is also a Girko matrix. Consequently, up to an overall scaling factor of $\sqrt{v(z)}$, the Hermitization of the spherical ensemble $\mathbf{H}_{\mathrm{s}}^z$ has the same distribution as the Hermitization for the circular ensemble, (4.1.2), evaluated at $z = 0$. From [1, Theorem 5.2], we therefore have the local law

$$|\langle \mathbf{G}_{\mathrm{s}}^z(\mathrm{i}\eta)\rangle - m_{\mathrm{s}}^z(\mathrm{i}\eta)| \prec \frac{1}{N\eta} \tag{5.2.8}$$

uniformly for $z \in \mathbb{C}$ and $N^{-1+\varepsilon} \leq \eta \leq N^{2C}$ for $\varepsilon > 0$ and $C > 0$ from the statement of the theorem. Here, m_{s}^z is given by

$$m_{\mathrm{s}}^z(\mathrm{i}\eta) := \mathrm{i}\frac{1}{2v(z)}\left(\sqrt{4v(z) + \eta^2} - \eta\right). \tag{5.2.9}$$

It is connected to σ_{s} by

$$\sigma_{\mathrm{s}}(z) = -\frac{1}{2\pi}\int_0^\infty \Delta_z \operatorname{Im} m_{\mathrm{s}}^z(\mathrm{i}\eta)\mathrm{d}\eta = \frac{1}{\pi}\frac{1}{(1+|z|^2)^2}. \tag{5.2.10}$$

This can be seen by evaluating the integral on the right-hand side, using the explicit form of m_{s}^z given in (5.2.9). We combine (5.2.10), (5.2.2), (5.2.3), (5.2.5) and (5.2.6) to obtain the following expression for the left-hand side of (5.2.1)

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^N f_{N,z_0}(\mu_j) - \int_{\mathbb{C}} f_{N,z_0}(z)\sigma(z)\mathrm{d}^2 z \\
=&\frac{1}{4\pi N}\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\log|\det(\mathbf{H}_{\mathrm{s}}^z - \mathrm{i}T)|\,\mathrm{d}^2 z \\
&\quad - \frac{1}{2\pi}\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\int_0^T \langle \operatorname{Im}\mathbf{G}_{\mathrm{s}}^z(\mathrm{i}\eta)\rangle - \operatorname{Im} m_{\mathrm{s}}^z(\mathrm{i}\eta)\mathrm{d}\eta\mathrm{d}^2 z \\
&\quad + \frac{1}{2\pi}\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\int_T^\infty \left(\operatorname{Im} m_{\mathrm{s}}^z(\mathrm{i}\eta) - \frac{1}{\eta}\right)\mathrm{d}\eta\mathrm{d}^2 z.
\end{aligned} \tag{5.2.11}$$

As in the case of (4.1.17), we have freely added the η^{-1} term in the last integral to ensure convergence of the integral at infinity. We estimate all terms in (5.2.11) separately along the lines of [1] and start with the first term. We have

$$\log|\det(\mathbf{H}_{\mathrm{s}}^z - \mathrm{i}T)| = 2N\log T + \sum_{j=1}^N \log\left(1 + \frac{\lambda_j^2}{T^2}\right). \tag{5.2.12}$$

Since the first term on the right-hand side does not depend on z, its integral against $\Delta f_{N,z_0}(z)$ vanishes and we have

$$\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\log|\det(\mathbf{H}_{\mathrm{s}}^z - \mathrm{i}T)|\,\mathrm{d}^2 z \leq \int_{\mathbb{C}} \Delta|f_{N,z_0}(z)|\frac{\operatorname{Tr}((\mathbf{H}_{\mathrm{s}}^z)^2)}{T^2}\mathrm{d}^2 z, \tag{5.2.13}$$

where we have used $\log(x+1) \leq x$. For our choice $T = N^{2C}$ we estimate the trace term by

$$\begin{aligned}\sup_{|z|\leq N^c} \frac{\mathrm{Tr}((\mathbf{H}_{\mathrm{s}}^z)^2)}{T^2} &= \sup_{|z|\leq N^c} \frac{1}{T^2}\sum_{i,j=1}^{N}(x_{ij}-zy_{ij})(\bar{x}_{ij}-\bar{z}\bar{y}_{ij}) \\ &\leq \sup_{|z|\leq N^c} \frac{1}{T^2}\sum_{i,j=1}^{N}|x_{ij}|^2+|z|^2|y_{ij}|^2 \prec \sup_{|z|\leq N^c} N\frac{|z|^2}{T^2} \leq N^{-2C+1}.\end{aligned} \tag{5.2.14}$$

For the upper bound of the second term in (5.2.11), recall that $v(z)^{-1/2}\mathbf{H}_{\mathrm{s}}^z$ has the same distribution as $\mathbf{H}^z$ from (4.1.2) at $z=0$. The resolvent of $\mathbf{H}_{\mathrm{s}}^z$ satisfies

$$\mathbf{G}_{\mathrm{s}}^z(\mathrm{i}\eta) = \frac{1}{\sqrt{v(z)}}\left(\frac{1}{\sqrt{v(z)}}\mathbf{H}_{\mathrm{s}}^z - \mathrm{i}\frac{\eta}{\sqrt{v(z)}}\right)^{-1} \tag{5.2.15}$$

and therefore it has the same distribution as $v(z)^{-1/2}\mathbf{G}^0(\mathrm{i}v(z)^{-1/2}\eta)$, where $\mathbf{G}^z$ is the resolvent of $\mathbf{H}^z$. Similarly, we have $m_{\mathrm{s}}^z(\mathrm{i}\eta) = v(z)^{-1/2}m^0(\mathrm{i}v(z)^{-1/2}\eta)$, where m^z was defined as the unique solution with positive imaginary part to (4.1.13). Then we have for all $p \in \mathbb{N}$ that

$$\begin{aligned}&\mathbb{E}\left[\left(\int_0^T |\langle \mathrm{Im}\,\mathbf{G}_{\mathrm{s}}^z(\mathrm{i}\eta)\rangle - \mathrm{Im}\,m_{\mathrm{s}}^z(\mathrm{i}\eta)|\mathrm{d}\eta\right)^p\right] \\ =&\mathbb{E}\left[\left(\frac{1}{\sqrt{v(z)}}\int_0^T \left|\left\langle \mathrm{Im}\,\mathbf{G}^0\left(\frac{\mathrm{i}\eta}{\sqrt{v(z)}}\right)\right\rangle - \mathrm{Im}\,m^0\left(\frac{\mathrm{i}\eta}{\sqrt{v(z)}}\right)\right|\mathrm{d}\eta\right)^p\right] \\ =&\mathbb{E}\left[\left(\int_0^{\frac{T}{\sqrt{v(z)}}} |\langle \mathrm{Im}\,\mathbf{G}^0(\mathrm{i}\eta)\rangle - \mathrm{Im}\,m^0(\mathrm{i}\eta)|\mathrm{d}\eta\right)^p\right] \leq \mathbb{E}[I(0)^p]\end{aligned} \tag{5.2.16}$$

where $I(z)$ was defined in (4.1.18). The last inequality holds since the integrand is positive and $v(z) \geq 1$. Therefore, Lemma 4.1.2 can be applied to obtain the estimate

$$\mathbb{E}\left[\left(\int_0^T |\langle \mathrm{Im}\,\mathbf{G}_{\mathrm{s}}^z(\mathrm{i}\eta)\rangle - \mathrm{Im}\,m_{\mathrm{s}}^z(\mathrm{i}\eta)|\mathrm{d}\eta\right)^p\right] \lesssim_{\delta,p} N^{-p(1-\delta)}. \tag{5.2.17}$$

By an argument analogous to (4.1.20) and (4.1.21), this moment bound is converted into the stochastic domination bound

$$\left|\int_{\mathbb{C}} \Delta f_{N,z_0}(z)\int_0^T \langle \mathrm{Im}\,\mathbf{G}_{\mathrm{s}}^z(\mathrm{i}\eta)\rangle - \mathrm{Im}\,m_{\mathrm{s}}^z(\mathrm{i}\eta)\mathrm{d}\eta\mathrm{d}^2z\right| \prec \|\Delta f\|_1 N^{-1+2\alpha}. \tag{5.2.18}$$

Lastly, the term integrand in the last term in (5.2.11) satisfies

$$|\,\mathrm{Im}\,m_{\mathrm{s}}^z(\mathrm{i}\eta) - \frac{1}{\eta}| = \mathcal{O}\left(\frac{v(z)}{\eta^3}\right) = \mathcal{O}(\eta^{-2}), \tag{5.2.19}$$

for $\eta \geq T$. The last estimate follows from $v(z) = 1 + |z|^2 \lesssim N^{2C} = T$. Therefore we have

$$\left| \int_{\mathbb{C}} \Delta f_{N,z_0}(z) \int_T^\infty \left(\operatorname{Im} m_{\mathrm{s}}^z(\mathrm{i}\eta) - \frac{1}{\eta} \right) \mathrm{d}\eta \mathrm{d}^2 z \right| \lesssim \frac{\|\Delta f\|_1}{T} = N^{-2C} \|\Delta f\|_1. \tag{5.2.20}$$

Using (5.2.14), (5.2.18) and (5.2.20) to estimate (5.2.11), we conclude the proof of Theorem 5.2.1.

□

Remark. *The idea behind the proof of Theorem 5.2.1 can be easily generalized to other rational functions of random matrices. For any rational function of the type*

$$\mathbf{R} = \mathbf{Q}^{-1}\mathbf{P} \in \mathbb{C}^{N \times N}, \tag{5.2.21}$$

with $\mathbf{Q}, \mathbf{P} \in \mathbb{C}^{N \times N}$, a proof of a local law for $\mathbf{R}$ reduces to a proof of a local law for the Hermitization

$$\mathbf{H}_{\mathbf{R}}^z = \begin{pmatrix} 0 & \mathbf{P} - z\mathbf{Q} \\ \mathbf{P}^* - \bar{z}\mathbf{Q}^* & 0 \end{pmatrix} \in \mathbb{C}^{N \times N} \tag{5.2.22}$$

as well as a smallest singular value estimate for $\mathbf{P} - z\mathbf{Q}$ in the form of Theorem 4.1.4 for all $z \in \mathbb{C}$. Two examples of rational functions for which both of these results are readily available are

$$\mathbf{R}_1 := (\varepsilon\mathbf{Y} + \sqrt{1-\varepsilon^2}\mathbf{Z})^{-1}(\varepsilon\mathbf{X} + \sqrt{1-\varepsilon^2}\mathbf{Z}) \quad \textit{and} \quad \mathbf{R}_2 := (\varepsilon\mathbf{Y} + \sqrt{1-\varepsilon^2}\mathbf{I}_N)^{-1}\mathbf{X}, \tag{5.2.23}$$

where $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are independent Girko matrices and $\varepsilon \in (0, 1]$ for $\mathbf{R}_1$ and $\varepsilon \in [0, 1]$ for $\mathbf{R}_2$. The ensemble $\mathbf{R}_1$ describes a generalization of the spherical ensemble, where we allow for correlation between the respective entries of $\mathbf{Q}_1$ and $\mathbf{P}_1$. For $\varepsilon = 1$, we recover the spherical ensemble and for $\varepsilon \to 0$, the ensemble approaches the identity matrix. The ensemble $\mathbf{R}_2$ describes an interpolation between the spherical and the Girko ensemble. For $\varepsilon = 1$, we again recover the spherical ensemble and for $\varepsilon = 0$, we recover the Girko ensemble.

Bibliography

[1] J. Alt, L. Erdős, and T. Krüger. *"Local inhomogeneous circular law"*. Ann. Appl. Probab. **28**(1): 148–203 (2018).

[2] J. Alt, L. Erdős, and T. Krüger. *"Spectral radius of random matrices with independent entries"*. Probab. Math. Phys. **2**(2): 221–280 (2021).

[3] J. Alt, L. Erdős, T. Krüger, and D. Schröder. *"Correlated random matrices: band rigidity and edge universality"*. Ann. Probab. **48**(2): 963–1001 (2020).

[4] G. W. Anderson. *"Convergence of the largest singular value of a polynomial in independent Wigner matrices"*. Ann. Probab. **41**(3B): 2103–2181 (2013).

[5] G. W. Anderson. *"A local limit law for the empirical spectral distribution of the anticommutator of independent Wigner matrices"*. Annales de l'I.H.P. Probabilités et statistiques **51**(3): 809–841 (2015).

[6] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. volume 118 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge (2010).

[7] F. Augeri. *"Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails"*. Electron. J. Probab. **21**: No. 32 (2016).

[8] F. Augeri, A. Guionnet, and J. Husson. *"Large deviations for the largest eigenvalue of sub-Gaussian matrices"*. Comm. Math. Phys. **383**(2): 997–1050 (2021).

[9] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, New York (2010).

[10] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, New York (2010).

[11] Z. D. Bai and Y. Q. Yin. *"Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix"*. Ann. Probab. **16**(4): 1729–1741, (1988).

[12] M. Banna and T. Mai. *"Hölder continuity of cumulative distribution functions for noncommutative polynomials under finite free Fisher information"*. J. Funct. Anal. **279**(8): 108710 (2020).

[13] Z. Bao, L. Erdős, and K. Schnelli. *"Local law of addition of random matrices on optimal scale"*. Comm. Math. Phys. **349**(3): 947–990 (2017).

[14] Z. Bao, L. Erdős, and K. Schnelli. *"Spectral rigidity for addition of random matrices at the regular edge"*. J. Funct. Anal. **279**(7): 108639 (2020).

[15] S. T. Belinschi and M. Capitaine. *"Spectral properties of polynomials in independent Wigner and deterministic matrices"*. J. Funct. Anal. **273**(12): 3901–3963 (2017).

[16] S. T. Belinschi, T. Mai, and R. Speicher. *"Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem"*. J. Reine Angew. Math. **732**: 21–53 (2017).

[17] C. Bordenave. *"On the spectrum of sum and product of non-Hermitian random matrices"*. Electron. Commun. Probab. **16**: 104–113 (2011).

[18] C. Bordenave and D. Chafaï. *"Around the circular law"*. Probab. Surv. **9**: 1–89 (2012).

[19] P. Bourgade, L. Erdős, and H.-T. Yau. *"Edge universality of beta ensembles"*. Comm. Math. Phys. **332**(1): 261–353 (2014).

[20] P. Bourgade, H.-T. Yau, and J. Yin. *"Local circular law for random matrices"*. Probab. Theory Related Fields **159**(3-4): 545–595 (2014).

[21] P. Bourgade, H.-T. Yau, and J. Yin. *"The local circular law II: the edge case"*. Probab. Theory Related Fields **159**(3-4): 619–660 (2014).

[22] I. Charlesworth and D. Shlyakhtenko. *"Free entropy dimension and regularity of non-commutative polynomials"*. J. Funct. Anal. **271**(8): 2274–2292 (2016).

[23] B. Collins, A. Guionnet, and F. Parraud. *"On the operator norm of non-commutative polynomials in deterministic matrices and iid GUE matrices"*. Camb. J. Math. **10**(1): 195–260, (2022).

[24] N. A. Cook, A. Guionnet, and J. Husson. *"Spectrum and pseudospectrum for quadratic polynomials in Ginibre matrices"*. Ann. Inst. Henri Poincaré Probab. Stat. **58**(4): 2284–2320 (2022).

[25] K. Dykema. *"On certain free product factors via an extended matrix model"*. J. Funct. Anal. **112**(1): 31–60 (1993).

[26] L. Erdős, A. Knowles, and H.-T. Yau. *"Averaging fluctuations in resolvents of random band matrices"*. Ann. Henri Poincaré **14**(8): 1837–1926 (2013).

[27] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. *"The local semicircle law for a general class of random matrices"*. Electron. J. Probab. **18**: No. 59 (2013).

[28] L. Erdős, T. Krüger, and Yu. Nemish. *"Local laws for polynomials of Wigner matrices"*. J. Funct. Anal. **278**(12): 108507 (2020).

[29] L. Erdős, T. Krüger, and D. Schröder. *"Random matrices with slow correlation decay"*. Forum Math. Sigma **7**: No. e8 (2019).

[30] L. Erdős and Y. Xu. *"Small deviation estimates for the largest eigenvalue of Wigner matrices"*. Bernoulli **29**(2): 1063–1079 (2023).

[31] L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*. volume 28 of *Courant Lecture Notes in Mathematics*, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI (2017).

[32] L. Erdős, H.-T. Yau, and J. Yin. *"Rigidity of eigenvalues of generalized Wigner matrices"*. Adv. Math. **229**(3): 1435–1515 (2012).

[33] J. Feinberg and A. Zee. *"Non-Hermitian random matrix theory: method of Hermitian reduction"*. Nuclear Phys. B **504**(3): 579–608 (1997).

[34] O. N. Feldheim and S. Sodin. *"A universality result for the smallest eigenvalues of certain sample covariance matrices"*. Geom. Funct. Anal. **20**(1): 88–123 (2010).

[35] J. Fronk, T. Krüger, and Y. Nemish. *"Norm Convergence Rate for Multivariate Quadratic Polynomials of Wigner Matrices"*, (2023). DOI: 10.48550/arXiv.2308.16778.

[36] F. Gesztesy and E. Tsekanovskii. *"On matrix-valued Herglotz functions"*. Math. Nachr. **218**: 61–138 (2000).

[37] J. Ginibre. *"Statistical ensembles of complex, quaternion, and real matrices"*. J. Mathematical Phys. **6**: 440–449 (1965).

[38] V. L. Girko. *"The circular law"*. Teor. Veroyatnost. i Primenen. **29**(4): 669–679, (1984).

[39] V. L. Girko. *"Strong elliptic law"*. Random Oper. Stochastic Equations **5**(3): 269–306 (1997).

[40] F. Götze, A. Naumov, and A. Tikhomirov. *"Local laws for non-Hermitian random matrices and their products"*. Random Matrices Theory Appl. **9**(4): 2150004 (2020).

[41] F. Götze and A. Tikhomirov. *"On the Asymptotic Spectrum of Products of Independent Random Matrices"*, (2011). ArXiv:1012.2710.

[42] U. Haagerup, H. Schultz, and S. Thorbjørnsen. *"A random matrix approach to the lack of projections in $C^*_{\mathrm{red}}(\mathbb{F}_2)$"*. Adv. Math. **204**(1): 1–83 (2006).

[43] U. Haagerup and S. Thorbjørnsen. *"A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group"*. Ann. of Math. (2) **162**(2): 711–775 (2005).

[44] J. W. Helton, T. Mai, and R. Speicher. *"Applications of realizations (aka linearizations) to free probability"*. J. Funct. Anal. **274**(1): 1–79 (2018).

[45] J. W. Helton, S. A. McCullough, and V. Vinnikov. *"Noncommutative convexity arises from linear matrix inequalities"*. J. Funct. Anal. **240**(1): 105–191 (2006).

[46] J. W. Helton, R. Rashidi Far, and R. Speicher. *"Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints"*. Int. Math. Res. Not. IMRN **2007**(22): rnm086 (2007).

[47] S. C. Kleene. *Representation of Events in Nerve Nets and Finite Automata*. RAND Corporation (1951).

[48] M. Krishnapur. *"From random matrices to random analytic functions"*. Ann. Probab. **37**(1): 314–346 (2009).

[49] B. Landon and H.-T. Yau. *"Edge statistics of Dyson Brownian motion"*. arXiv: 1712.03881 , (2017).

[50] J. O. Lee and K. Schnelli. *"Edge universality for deformed Wigner matrices"*. Rev. Math. Phys. **27**(8): 1550018 (2015).

[51] J. O. Lee and J. Yin. *"A necessary and sufficient condition for edge universality of Wigner matrices"*. Duke Math. J. **163**(1): 117–173 (2014).

[52] T. Mai, R. Speicher, and M. Weber. *"Absence of algebraic relations and of zero divisors under the assumption of full non-microstates free entropy dimension"*. Adv. Math. **304**: 1080–1107 (2017).

[53] C. Male. *"The norm of polynomials in large random and deterministic matrices"*. Probab. Theory Related Fields **154**(3-4): 477–532 (2012).

[54] Yu. Nemish. *"Local law for the product of independent non-Hermitian random matrices with independent entries"*. Electron. J. Probab. **22**: No. 22 (2017).

[55] S. O'Rourke and A. Soshnikov. *"Products of independent non-Hermitian random matrices"*. Electron. J. Probab. **16**: No. 81, 2219–2245, (2011).

[56] N. S. Pillai and J. Yin. *"Edge universality of correlation matrices"*. Ann. Statist. **40**(3): 1737–1763 (2012).

[57] N. S. Pillai and J. Yin. *"Universality of covariance matrices"*. Ann. Appl. Probab. **24**(3): 935–1001 (2014).

[58] M. P. Schützenberger. *"On the definition of a family of automata"*. Information and Control **4**: 245–270, (1961).

[59] D. Shlyakhtenko and P. Skoufranis. *"Freely independent random variables with non-atomic distributions"*. Trans. Amer. Math. Soc. **367**(9): 6267–6291 (2015).

[60] A. Soshnikov. *"Universality at the edge of the spectrum in Wigner random matrices"*. Comm. Math. Phys. **207**(3): 697–733 (1999).

[61] T. Tao and V. Vu. *"Random matrices: the circular law"*. Commun. Contemp. Math. **10**(2): 261–307 (2008).

[62] T. Tao and V. Vu. *"Random matrices: universality of local eigenvalue statistics up to the edge"*. Comm. Math. Phys. **298**(2): 549–572 (2010).

[63] T. Tao and V. Vu. *"Random matrices: The Universality phenomenon for Wigner ensembles"*, (2012).

[64] T. Tao and V. Vu. *"Random matrices: sharp concentration of eigenvalues"*. Random Matrices Theory Appl. **2**(3): 1350007 (2013).

[65] C. A. Tracy and H. Widom. *"Level-spacing distributions and the Airy kernel"*. Phys. Lett. B **305**(1-2): 115–118 (1993).

[66] C. A. Tracy and H. Widom. *"On orthogonal and symplectic matrix ensembles"*. Comm. Math. Phys. **177**(3): 727–754, (1996).

[67] D. Voiculescu. *"Limit laws for random matrices and free products"*. Invent. Math. **104**(1): 201–220 (1991).

[68] E. P. Wigner. *"Characteristic vectors of bordered matrices with infinite dimensions"*. Ann. of Math. (2) **62**: 548–564 (1955).

[69] J. Wishart. *"The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population"*. Biometrika **20A**(1/2): 32–52, (1928).

[70] J. Yin. *"The local circular law III: general case"*. Probab. Theory Related Fields **160**(3-4): 679–732 (2014).

Acknowledgements

First and foremost, I would like to thank my supervisor Torben Krüger. Over countless Zoom meetings and unfortunately much fewer in-person meetings, he introduced me to the beautiful world of random matrix theory and I am deeply grateful for your scientific guidance, support and patience over the last three years. I would also like to thank my earlier scientific mentors Manfred Salmhofer and Andreas Mielke, as well as my co-author Yuriy Nemish.

I would like to thank Frederik for his help with the Danish translation of my abstract and my proofreaders Paula and Sabiha. Your feedback was invaluable!

Thank you to all current and former members of QMATH. Between coffee breaks, running, bouldering, pub quizzes and so many more things, it has always been a great working environment and I have enjoyed my time with all of you a lot. In particular, I would like to thank Frederik, Freek, Itai, Johannes, Paula, Sabiha and Tommaso for your support - both scientific and emotional.

Thank you also to my friends in Heidelberg, Julia, Jannis, Max, Jens and Iris. All of you made a difference on my journey here. I would also like to thank my parents, Jutta and John, for all your support along the way. Thank you to Eva. Without you, I would not have made it through the PhD. And lastly, I do not have words to express my gratitude towards my sister, Rebecca, for everything, but especially for the time we had together. I know you would have been proud of me.