Dyson equation and eigenvalue statistics of random matrices

by

Johannes Alt

July, 2018

A thesis presented to the
Graduate School
of the
Institute of Science and Technology Austria, Klosterneuburg, Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

IST AUSTRIA
Institute of Science and Technology

The thesis of Johannes Alt, titled *Dyson equation and eigenvalue statistics of random matrices*, is approved by:

Supervisor: László Erdős, IST Austria, Klosterneuburg, Austria

Signature: ____________________

Committee Member: Jan Maas, IST Austria, Klosterneuburg, Austria

Signature: ____________________

Committee Member: Jiří Černý, University of Basel, Basel, Switzerland

Signature: ____________________

Defense Chair: Maximilian Jösch, IST Austria, Klosterneuburg, Austria

Signature: ____________________

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: ____________________________

Johannes Alt

July, 2018

Biographical Sketch

Johannes Alt obtained a BSc in Mathematics and a BSc in Physics from Saarland University in 2012. He completed a MSc in Mathematics at Ludwig-Maximilians University Munich in 2014 and was awarded the Carathéodory prize for an outstanding master thesis in Mathematics. In September 2014, he then joined IST Austria as a graduate student. His research interests lie in probability theory, analysis and mathematical physics. These areas symbiotically interplay in random matrix theory on which Johannes' current research focuses. He is particularly interested in rigorously analyzing universal characteristics of eigenvalue statistics of large random matrices.

List of Publications

The present thesis contains the following publications and preprints.

[1] *The local semicircle law for random matrices with a fourfold symmetry*, J. Math. Phys. **56** (2015), no. 10 (arXiv:1506.04683).

[2] *Local law for random Gram matrices*, Electron. J. Probab. **22** (2017), no. 25, 41 pp., *with László Erdős and Torben Krüger* (arXiv:1606.07353).

[3] *Local inhomogeneous circular law*, Ann. Appl. Probab. **28** (2018), no. 1, 148–203, *with László Erdős and Torben Krüger* (arXiv:1612.07776).

[4] *Location of the spectrum of Kronecker random matrices*, arXiv:1706.08343, 2017, *with László Erdős, Torben Krüger and Yuriy Nemish.* Accepted to Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques.

[5] *Singularities of the density of states of random Gram matrices*, Electron. Commun. Probab. **22** (2017), no. 63, 13 pp. (arXiv:1708.08442).

[6] *Correlated random matrices: Band rigidity and Edge universality*, arXiv:1804.07744, 2018, *with László Erdős, Torben Krüger and Dominik Schröder.*

[7] *The Dyson equation with linear self-energy: spectral bands, edges and cusps*, arXiv:1804.07752, 2018, *with László Erdős and Torben Krüger.*

Acknowledgments

I would like to use this opportunity to thank all those who have supported me during my PhD studies. There are quite a few people whom I want to mention by name.

I am grateful to all current and former members of the research group of my supervisor, László Erdős, at IST Austria during my PhD studies. I have benefited a lot from many inspiring discussions with them and the motivating atmosphere.

Moreover, I very much appreciate the support of my thesis committee with the internal members, László Erdős and Jan Maas, from IST Austria and Jiří Černý as external member.

Some people have directly contributed to some research projects of this thesis. They are László Erdős, Torben Krüger, Yuriy Nemish and Dominik Schröder (in alphabetical order). I am very grateful to them for our fruitful collaborations which have been a constant source of inspiration and motivation to me.

Overall, my PhD studies have been a very good and rewarding experience for me. To a large extend, this is due to my supervisor, László Erdős. He has always made a lot of time to listen to my concerns and questions, to answer them, to introduce me to new mathematical ideas, to discuss about my research projects and to teach me how to efficiently communicate mathematics. László has been the best supervisor I could have imagined and, for that, I want to express my deep gratitude to him.

Finally, I would like to say "thank you" to my friends, my family and in particular my parents, Claudia and Andreas. They have been a constant support for me during my whole undergraduate and graduate studies. I appreciate their previous and future help very much.

The author acknowledges support of the European Research Council under ERC Advanced Grant, RANMAT, No. 338804.

Abstract

The eigenvalue density of many large random matrices is well approximated by a deterministic measure, the *self-consistent density of states*. In the present work, we show this behaviour for several classes of random matrices. In fact, we establish that, in each of these classes, the self-consistent density of states approximates the eigenvalue density of the random matrix on all scales slightly above the typical eigenvalue spacing.

For large classes of random matrices, the self-consistent density of states exhibits several universal features. We prove that, under suitable assumptions, random Gram matrices and Hermitian random matrices with decaying correlations have a 1/3-Hölder continuous self-consistent density of states ρ on $\mathbb{R}$, which is analytic, where it is positive, and has either a square root edge or a cubic root cusp, where it vanishes. We, thus, extend the validity of the corresponding result for Wigner-type matrices from [4, 5, 7].

We show that ρ is determined as the inverse Stieltjes transform of the normalized trace of the unique solution $m(z)$ to the *Dyson equation*

$$-m(z)^{-1} = z - a + S[m(z)]$$

on $\mathbb{C}^{N\times N}$ with the constraint $\operatorname{Im} m(z) \geq 0$. Here, z lies in the complex upper half-plane, a is a self-adjoint element of $\mathbb{C}^{N\times N}$ and S is a positivity-preserving operator on $\mathbb{C}^{N\times N}$ encoding the first two moments of the random matrix. In order to analyze a possible limit of ρ for $N \to \infty$ and address some applications in free probability theory, we also consider the Dyson equation on infinite dimensional von Neumann algebras.

We present two applications to random matrices. We first establish that, under certain assumptions, large random matrices with independent entries have a rotationally symmetric self-consistent density of states which is supported on a centered disk in $\mathbb{C}$. Moreover, it is infinitely often differentiable apart from a jump on the boundary of this disk. Second, we show edge universality at all regular (not necessarily extreme) spectral edges for Hermitian random matrices with decaying correlations.

Contents

List of Tables

List of Figures

List of Symbols

We list a number of general symbols used through the whole text. In the individual chapters, we introduce further notation which will be used only within the respective chapter.

$\mathbb{Z}$	Integers
$\mathbb{N}$	Positive integers
$[n] \coloneqq \{1, \ldots, n\} \subset \mathbb{N}$	Set of first n positive integer, $n \in \mathbb{N}$
$\mathbb{R}$	Real numbers
$\mathbb{R}_+ = (0, \infty)$	Strictly positive real numbers
$\mathbb{R}_0^+ = [0, \infty)$	Nonnegative real numbers
$\mathbb{C}$	Complex numbers
i	Imaginary unit
$\operatorname{Re} z$, $\operatorname{Im} z$	Real and imaginary part of $z \in \mathbb{C}$
$\mathbb{H} \coloneqq \{z \in \mathbb{C} \colon \operatorname{Im} z > 0\}$	Complex upper half-plane
x_+	Positive part for $x \in \mathbb{R}$
x_-	Negative part for $x \in \mathbb{R}$
$D_r(z)$, $D(z, r)$	Disk in $\mathbb{C}$ of radius $r > 0$ centered at $z \in \mathbb{C}$
$\operatorname{supp} \rho$	Support of the measure ρ or the function ρ
$\mathbb{P}$	Probability measure on probability space
$\mathbb{E}$	Expectation with respect to $\mathbb{P}$
$\mathbf{1}(A)$, $\mathbf{1}_A$	Characteristic function of an event A
$\mathbb{1}$	Identity matrix or identity operator
$\operatorname{Spec}(X)$	Spectrum of the matrix X or the operator X

$\mathrm{Tr}(X)$	Trace of a matrix X, sum of its diagonal elements
$\mathrm{tr}(X) = \frac{1}{n}\,\mathrm{Tr}(X)$	Normalized trace of an $n \times n$ matrix X
$\det(X)$	Determinant of a matrix X
$\mathrm{dist}(a, B) = \inf\{d(a, b) \colon b \in B\}$	Distance of $a \in X$ from a set $B \subset X$ in a metric space (X, d)

List of Abbreviations

RMT: Random matrix theory

MDE: Matrix Dyson equation

WDM: Wigner-Dyson-Mehta universality conjecture

QVE: Quadratic vector equation

CHAPTER 1

Introduction

The study of eigenvalue densities of large random matrices has a long history. In a seminal work, it was initiated by Wigner in the 1950's [157]. He proved that the eigenvalue density of an $N \times N$ Hermitian matrix with independent (up to the symmetry constraint) and centered entries of variance $1/N$ converges to a semicircular distribution when N tends to infinity [158]. Such matrices are now called *Wigner matrices* and the convergence result is referred to as *Wigner's semicircle law*. Figure 1.1 shows Wigner's semicircle law, $\rho_{\mathrm{sc}}(x) := \frac{1}{2\pi}\sqrt{(4-x^2)_+}$, and the eigenvalue density of a sampled Wigner matrix.

Wigner's semicircle law is the first instance of the *universality phenomenon* in random matrix theory (RMT) since he showed that the limit of the eigenvalue density is independent of the precise distribution of the matrix entries. Moreover, Wigner conjectured that the distribution of the gaps of consecutive eigenvalues of Wigner matrices follows a universal law which only depends on the basic symmetry type of the random matrix, i.e., whether it is a real symmetric or a complex Hermitian matrix. Nowadays, it is a common belief in RMT that many features of the eigenvalue statistics of large random matrices are universal in the sense that they do not depend on fine details of the random matrix ensemble[†] but hold true for large classes of random matrices with the same "symmetry" type.

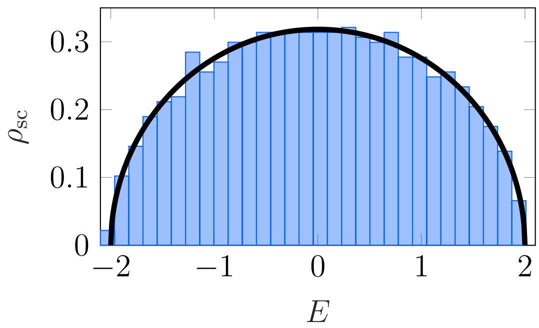


FIGURE 1.1. Wigner's semicircle law ρ_{sc} and eigenvalue density of a 1000×1000 Gaussian Wigner matrix

[†] By a slight abuse of terminology, we use the terms "random matrix" and "random matrix ensemble" interchangeably. Strictly speaking, the latter usually denotes the induced probability measure on the space of Hermitian matrices but we do not make this distinction here.

Since Wigner's ground-breaking ideas, verifying this belief is one of the main objectives in RMT and many works have been devoted to this goal.

The present work contributes to this goal for five classes of random matrices:

- Hermitian random matrices with a special fourfold symmetry,
- Random Gram matrices,
- Random matrices with independent entries,
- Kronecker random matrices,
- Hermitian random matrices with general, decaying correlations among their entries.

In the remainder of this introduction, we describe the questions about the eigenvalue statistics studied in the present work. In Chapter 2, we then explain the results presented in the final seven chapters, Chapter 3 to Chapter 9. Each of these chapters has been published (or submitted for publication) as a separate paper. Hence, it can be read independently.

When analyzing the eigenvalue density of a large random matrix, the first question one asks is whether there is a deterministic measure that approximates the eigenvalue density of this ensemble. A theorem that answers this question affirmatively is called *global law* and the deterministic measure is referred to as the *self-consistent density of states*.

This deterministic measure is typically determined solely by the first two moments of the random matrix ensemble and it can be computed by solving the *Dyson equation*

$$-m(z)^{-1} = z\mathbb{1} - a + S[m(z)] \tag{1}$$

on $\mathbb{C}^{N\times N}$ under the constraint that $\operatorname{Im} m(z) := \frac{1}{2\mathrm{i}}(m(z)-m(z)^*)$ is positive definite. Here, z lies in the complex upper half-plane, $\mathbb{1}$ is the identity matrix in $\mathbb{C}^{N\times N}$, a is a self-adjoint element of $\mathbb{C}^{N\times N}$ and S is a positivity-preserving operator on $\mathbb{C}^{N\times N}$. The matrix a and the operator S encode the first and the second moment of the random matrix ensemble, respectively.

In many cases, the global law can be strengthened to a *local law* which asserts that the eigenvalue density is well approximated by the self-consistent density of states not only globally but also on smaller mecoscopic scales. A local law is called *optimal* if it holds on all scales slightly above the typical eigenvalue spacing. We remark that local laws have played a pivotal role in the proof of the so-called *Wigner-Dyson-Mehta* (WDM) *universality conjecture* via the *three-step strategy* [67], see also the recent developments in [66, 105]. The WDM universality conjecture, which is due to Dyson and Mehta [114], formalizes Wigner's conjecture on the eigenvalue gap distribution mentioned above. It predicts that the eigenvalue statistics on the microscopic scale, the scale of the typical eigenvalue spacing, in the bulk, i.e., where the self-consistent density of states is strictly positive, is given by a universal distribution for all random matrices of the same basic symmetry type. Similarly, for each basic symmetry type, there is a universal (Tracy-Widom) distribution that governs the eigenvalue statistics on the microscopic scale at the edge, i.e., at the boundary of the support of the self-consistent density of states. This phenomenon is called *edge universality*.

Thus, there are three strongly connected but mathematically distinct questions, we will study

(a) Analysis of the solution to the Dyson equation, (1),
(b) Proof of the optimal local laws,
(c) Proof of universality of local spectral statistics.

Previously, in [4, 5], some remarkable universal regularity properties of the self-consistent density of states ρ of *Wigner-type matrices* have been proven. Wigner-type matrices are Hermitian random matrices with centered, independent entries (up to the symmetry constraint). They naturally generalize Wigner matrices. Indeed, ρ is shown to be 1/3-Hölder continuous, analytic, where it is positive, and have a square root edge or an internal cubic root cusp, where it vanishes. It is remarkable that despite the high dimensionality and nonlinearity of the Dyson equation, the singularity structure of ρ can be described in such a simple universal form. Such detailed information about ρ is also necessary to establish a local law not only in the bulk but also in the vicinity of the singularities of ρ. For a certain class of Wigner-type matrices, this has been achieved in [7].

We stress that independence of the matrix elements leads to a structurally much simpler Dyson equation. In fact, m in (1) is always a diagonal matrix in this case and, thus, the Dyson equation can be studied in a commutative setup.

In the current thesis, we substantially generalize the results of [4, 5, 7] by dropping the independence condition on the matrix elements. This leads to a conceptually much more involved genuinely noncommutative Dyson equation, in fact, the analysis of the Dyson equation can go beyond matrices and we present it in the more general setup of von Neumann algebras.

The optimal local law and local spectral statistics in the bulk have been proven in the noncommutative matrix setup in [6, 56]. In this thesis, we perform the detailed edge analysis, culminating in the proof of the Tracy-Widom universality for the edge eigenvalues (including all internal edges) for very general random matrices with correlated entries.

We also analyze the corresponding questions, regularity of self-consistent density of states and local law, for random matrices with independent entries. For these non-Hermitian matrices, the eigenvalues concentrate on a domain in the complex plane. Studying whether the eigenvalues of a random matrix concentrate on a deterministic set is an even more elementary question than a global law. Indeed, the latter implies the former and the deterministic set is the support of the self-consistent density of states.

CHAPTER 2

Overview of the results

We now explain the contents of each individual chapter of the thesis in a short, informal way. We also put these results into the historical context and give the most important motivations. For more detailed information about previous results, we refer to the introductions of the individual chapters. Each section in the present chapter is numbered and titled according to the number and title of the chapter summarized in it.

CHAPTER 3: LOCAL SEMICIRCLE LAW FOR RANDOM MATRICES WITH A FOURFOLD SYMMETRY. Wigner introduced Wigner matrices as a model for the Hamiltonian of large atomic nuclei [159]. In this analogy, the eigenvalues of the Wigner matrix correspond to the energy levels of the atomic nucleus. Since then, random matrix theory has found many further applications in physics. In [32], it was argued that a good approximation to the two-dimensional Anderson model is given by a random matrix $H = (h_{ij})_{i,j \in \mathbb{Z}/N\mathbb{Z}}$ which satisfies the *fourfold symmetry*

$$h_{ij} = \bar{h}_{ji} = h_{-i,-j} = \bar{h}_{-j,-i} \tag{2}$$

for all $i, j \in \mathbb{Z}/N\mathbb{Z}$ and possesses a constant diagonal.

Motivated by this application, we study a class of random matrices with the fourfold symmetry, (2), in Chapter 3[†] below. For these matrices, we establish a local law with Wigner's semicircle law as self-consistent density of states, i.e., the *local semicircle law*. Compared to all previous proofs of local semicircle laws, the main difficulty is that the fourfold symmetry requires the simultaneous analysis of two vector self-consistent equations for the diagonal and the counterdiagonal of the resolvent instead of only one equation for the diagonal of the resolvent. In fact, our argument follows the strategy in [60], where the local semicircle law for *generalized Wigner matrices* was shown. A

[†] Chapter 3 is based on the publication [12].

Hermitian matrix $H = (h_{ij})_{i,j=1}^N$ is a generalized Wigner matrix if $\{h_{ij} \colon i \leq j\}$ are independent and centered random variables such that all variances $s_{ij} := \mathbb{E}|h_{ij}|^2$ scale like $1/N$ with upper and lower bounds and the variance matrix $S = (s_{ij})_{i,j=1}^N$ is stochastic, i.e., the entries in each row sum up to 1.

In fact, in Chapter 3, we consider random matrices $H = (h_{ij})_{i,j \in \mathbb{Z}/N\mathbb{Z}}$ whose entries are centered and independent up to the fourfold symmetry (2) for all $i, j \in \mathbb{Z}/N\mathbb{Z}$. Moreover, we assume that all variances $s_{ij} := \mathbb{E}|h_{ij}|^2$ scale like $1/N$ and the variance matrix $S = (s_{ij})_{i,j\in\mathbb{Z}/N\mathbb{Z}}$ is stochastic. We denote by $m_{\mathrm{sc}}(z)$ the Stieltjes transform of the semicircle law ρ_{sc} on $[-2, 2]$ and by $G(z) := (H - z)^{-1}$ the resolvent of H with entries $G_{ij}(z)$. In this situation, we show that, for any $\gamma > 0$, we have

$$\max_{i,j\in\mathbb{Z}/N\mathbb{Z}} |G_{ij}(z) - \delta_{ij} m_{\mathrm{sc}}(z)| \lesssim \frac{1}{\sqrt{N \operatorname{Im} z}} \tag{3}$$

with very high probability[1] for all $z \in \mathbb{C}$ such that $\operatorname{Im} z \geq N^{-1+\gamma}$ and $||\operatorname{Re} z| - 2| \geq \gamma$ (see Theorem 3.2.3 below). We remark that (3) is prototypical for local laws of Hermitian matrices, which are most conveniently formulated as a high probability estimate on the difference between the resolvent and a deterministic matrix. The estimate (3) is an optimal local law since it implies the convergence of the eigenvalue density of H to the semicircle law on all mesoscopic scales. Here, owing to the normalization $s_{ij} \leq 1/N$, the typical eigenvalue spacing is $1/N$ and $\operatorname{Im} z$ selects the mesoscopic scale $\geq N^{-1+\gamma}$. The local law, (3), also implies *eigenvalue rigidity*, i.e.,

$$|\lambda_j - \gamma_j| \lesssim N^{-1} \tag{4}$$

with very high probability[1]for $\delta \leq j/N \leq 1 - \delta$. Here, $\lambda_1 \leq \ldots \leq \lambda_N$ are the eigenvalues of H and $\gamma_1, \ldots, \gamma_N$ are the $1/N$-quantiles of the semicircle distribution ρ_{sc}.

The local semicircle law, solely with the Hermitian symmetry, in [60] was obtained by analyzing a self-consistent equation for the vector $(G_{ii} - m_{\mathrm{sc}})_{i=1}^N$. Compared to the Hermitian symmetry, the fourfold symmetry imposes additional correlations among the entries. Therefore, the proof of (3) in Chapter 3 below requires analyzing an additional

[1]The notation $\lesssim$ in (3) and (4) indicates that the estimates hold true up to an N^ε-factor with arbitrary $\varepsilon > 0$. The probability of the associated event depends on ε. The precise statement is obtained by replacing $\lesssim$ by the stochastic domination $\prec$ (see Definition 3.2.1 below).

self-consistent equation for the vector $(G_{i,-i})_{i\in\mathbb{Z}/N\mathbb{Z}}$ simultaneously to the one for $(G_{ii} - m_{\rm sc})_{i\in\mathbb{Z}/N\mathbb{Z}}$.

Chapter 4: Local law for random Gram matrices. Prior to Wigner matrices, Wishart had introduced another special class of random matrices in 1928 [160]. In applications to mathematical statistics, he used random matrices of the form XX^*, where X is a $p \times n$ matrix with independent, centered Gaussian entries of identical variance. In this situation, XX^* is called a *Wishart matrix*.

Sample covariance matrices are the generalization of Wishart matrices when the assumption of Gaussian distribution of the entries is dropped. Sample covariance matrices play an important role in mathematical statistics. This is because the covariance matrix of n repeated (independent) measurements of a vector $x \in \mathbb{C}^p$ with independent components is usually modeled by a sample covariance matrix XX^* with a $p \times n$ matrix X. In 1967, Marchenko and Pastur obtained the counterpart of Wigner's semicircle law for sample covariance matrices [112]. The *Marchenko-Pastur law* asserts that if n tends to infinity and simultaneously p/n tends to a strictly positive, finite constant $\gamma \in (0, \infty)$ then the eigenvalue density of XX^* converges to a deterministic probability density ρ_γ on $\mathbb{R}$. In Figure 2.1, this result is demonstrated in two cases, in Figure 2.1 (A) for $p/n \to \gamma = 1/2$ and in Figure 2.1 (B) for $p/n \to \gamma = 1$.

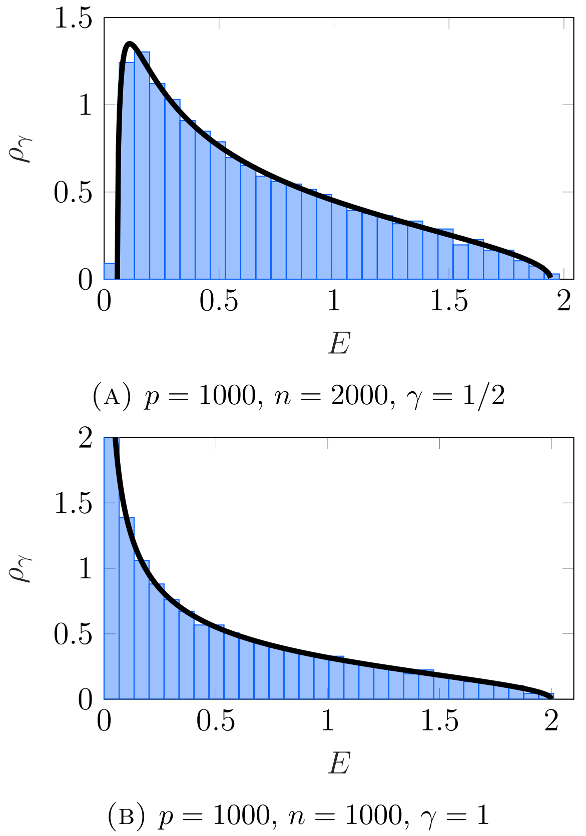


Figure 2.1. Comparison between Marchenko-Pastur law ρ_γ and the eigenvalue density of XX^*, where X is a $p \times n$ matrix with independent, centered Gaussian entries of variance $1/(p+n)$.

By dropping the assumption of identical variances in the definition of sample covariance matrices, we obtain *random Gram matrices* XX^*, where X is a $p \times n$ matrix with independent, centered entries. In the theory of wireless communication, they are used to model systems with multiple transmitting and receiving antennas [90, 150]. The channel capacity of such system is given by an integral with respect to the eigenvalue density of XX^*. Assuming a global or local law for XX^*, this can be approximated by an integral with respect to the self-consistent density of states.

In Chapter 4[†] below, we therefore prove a bulk local law for random Gram matrices and analyze their self-consistent density of states ρ. The main challenge compared to previous works is an additional unstable direction in the defining equation for ρ close to zero. Therefore, the proof of the local law requires very precise information about the behaviour of ρ in the vicinity of zero. In order to obtain this information, we distinguish the cases (i) $p = n$ and (ii) p/n is away from zero, one and infinity. The other main assumption in Chapter 4 is that the variances of the entries of X scale like p with upper and lower bounds. Denoting the variances of the entries of X by s_{ij} and the variance matrix by $S = (s_{ij})_{i,j}$, the self-consistent density of states ρ can be obtained from the unique solution $(m_1, m_2) \in \mathbb{C}^{p+n}$ of the vector Dyson equation

$$\begin{aligned} -\frac{1}{(m_1)_i} &= z + (Sm_2)_i, && \text{for } i = 1, \ldots, p, \\ -\frac{1}{(m_2)_k} &= z + (S^t m_1)_k, && \text{for } k = 1, \ldots, n, \end{aligned} \tag{5}$$

satisfying $\operatorname{Im} m_1(z) > 0$ and $\operatorname{Im} m_2(z) > 0$ for all $z \in \mathbb{C}$ with $\operatorname{Im} z > 0$. In fact, $\operatorname{Im} m_1(E + \mathrm{i}\eta)$ in the limit $\eta \downarrow 0$ determines ρ at E for any $E \in \mathbb{R}$. For a sample covariance matrix, the system (5) reduces to a single scalar quadratic equation that can be solved explicitly [112]. For general S, no explicit solution exists.

For Wigner-type random matrices, the quadratic vector equation (QVE), which is similar to (5), has been analyzed in [4, 7]. One key element in the regularity analysis of the self-consistent density of states and the proof of the local law for Wigner-type matrices and random Gram matrices is to understand the stability properties of the QVE and (5),

[†] Chapter 4 below essentially agrees with the publication [14] which is a joint work with László Erdős and Torben Krüger.

respectively, against small pertubations. The linear stability operator of the QVE has precisely one unstable direction. This instability is directly regularized by the positivity of the self-consistent density of states in the bulk. In contrast to this simpler case, the linear stability operator of (5) has two unstable directions. The first unstable direction is again controlled by the positivity of the self-consistent density of states ρ in the bulk. For the second one, m_1 and m_2 have to be analyzed in detail for $\operatorname{Re} z = 0$. Indeed, we show that (m_1, m_2) avoids this unstable direction for $p = n$ due to an extra symmetry. In Theorem 4.2.8 below, we then conclude that ρ has an inverse square-root blow-up at $E = 0$ in this case. For $|p/n - 1| \geq c$, the support of ρ has a gap around zero and ρ has a point mass at zero if $p > n$ (see Theorem 4.2.10 below). This is used to conclude regularity of the absolutely continuous part of ρ and the local law close to $E = 0$.

CHAPTER 5: SINGULARITIES OF THE DENSITY OF STATES OF RANDOM GRAM MATRICES. In Chapter 5[†] below, we extend the bulk analysis of ρ in Chapter 4 to the vicinity of the singularities of ρ and the local law to the whole real line. In the vicinity of the singularities, the stability is more critical and, owing to the additional unstable direction of the stability operator, the stability analysis has to be adjusted even for $\operatorname{Re} z \neq 0$. More precisely, we prove under some additional assumptions on the variances s_{ij} and away from zero that ρ is 1/3-Hölder continuous, analytic, where it is positive, and has a square root or a cubic root singularity, where it vanishes. Thus, the self-consistent density of states of random Gram matrices has the same regularity properties as the self-consistent density of states of Wigner-type matrices.

In fact, the precise behaviour of ρ close to its singularities is obtained by carefully expanding $\rho(\tau_0 + \omega)$ for small ω around $\tau_0 \in \operatorname{supp} \rho$ satisfying $\rho(\tau_0) = 0$. In [4, 5], it was shown for the Wigner-type setup that this expansion is *stable* in the sense that the coefficients of the cubic and quadratic terms do not vanish at the same time. Owing to this essential property, the expansion is dominated by the cubic or the quadratic term as the coefficient of the linear term vanishes. Hence, we obtain an approximately cubic equation for $\rho(\tau_0 + \omega)$ and only square root or cubic root singularities can occur. We remark that the coefficients in this expansion are basically determined by the linear stability operator

[†] Chapter 5 is based on the publication [11].

of the QVE at $z = \tau_0$, i.e., the analogue of (5) at $z = \tau_0$. Therefore, in the setup of Gram matrices, the stability of this expansion requires a new proof compared to [4, 5] due to the presence of two unstable directions of the stability operator.

CHAPTER 6: LOCAL INHOMOGENEOUS CIRCULAR LAW. Chapter 6[†] below deals with random matrices with independent entries, i.e., without any symmetry. We show the optimal local law for such matrices and analyze the regularity of their self-consistent density of states. The unstable nature of the spectrum of these non-Hermitian and even non-normal matrices requires a much harder simultaneous analysis of a family of Wigner-type matrices with noncentered entries of non-identical variances. This is the main novelty compared to previous works.

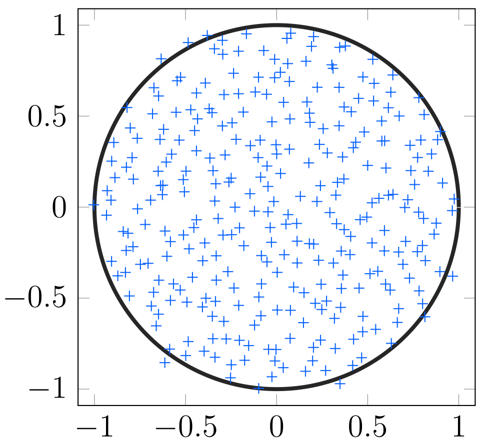


FIGURE 2.2. Eigenvalues of a 300×300 matrix with centered, independent Gaussian entries of variance $1/300$. Almost all eigenvalues are contained in a disk of radius 1.

We now explain our results in Chapter 6 and the difficulties in more detail. Let $X = (x_{ij})_{i,j=1}^N$ be a random matrix with independent and centered entries. We again denote its variance matrix by $S = (s_{ij})_{i,j=1}^N$, $s_{ij} := \mathbb{E}|x_{ij}|^2$, and assume that all variances s_{ij} scale like $1/N$. In Theorem 6.2.6 below, we prove, under additional technical assumptions, that there exists a deterministic function $\sigma\colon \mathbb{C} \to [0, \infty)$ such that the eigenvalue density of X is well approximated by σ on all scales above the typical eigenvalue spacing. The proof holds true inside the disk $D(0, R)$ of radius $R := \sqrt{\rho(S)}$, where $\rho(S)$ is the spectral radius of S. Analogously to the case of identical variances, where σ is the uniform measure on the unit disk (see Figure 2.2), σ is radially symmetric and supported on $D(0, R)$. Moreover, σ is infinitely often differentiable on $D(0, R)$ and has positive upper and lower bounds on $D(0, R)$, i.e., it has a jump discontinuity on the boundary of $D(0, R)$ (see Proposition 6.2.5 below). Furthermore, for every $\varepsilon > 0$,

[†] Chapter 6 below presents the publication [13] which is a joint work with László Erdős and Torben Krüger.

all eigenvalues of X are contained in $D(0, R + \varepsilon)$ with very high probability (see Theorem 6.2.6 below).

Prior to our work, the local law has only been established in the case of identical variances $s_{ij} = 1/N$ [44, 45, 146, 162], which requires a linear stability analysis of a scalar cubic equation. For non-identical variances, a much more challenging linear stability analysis of a $2N$-dimensional vector Dyson equation for the Hermitian random matrix

$$H_\zeta = \begin{pmatrix} 0 & X - \zeta \\ (X - \zeta)^* & 0 \end{pmatrix}, \tag{6}$$

where $\zeta \in \mathbb{C}$ is an additional parameter, is necessary. This Hermitization trick is due to Girko [81]. The global law has been proven in [51]. The proof of a local law necessitates the analysis on much finer scales compared to the one of a global law. Therefore, to obtain our result, the linear stability analysis of the full vector Dyson equation is performed on all scales. The main difficulty is the additional complex parameter ζ in (6), which is not present in the general Hermitian problems studied in [4, 5, 6, 7]. The bounds in the linear stability analysis, also for derivatives with respect to ζ, have to be uniform in ζ. This uniformity is also necessary to obtain the detailed information about σ mentioned above. In particular, the positive lower bound on σ and its smoothness are new results compared to [51].

CHAPTER 7: LOCATION OF THE SPECTRUM OF KRONECKER RANDOM MATRICES. In Chapter 7[†] below, we prove that, for a very big class of Hermitian and non-Hermitian random matrices, the eigenvalues concentrate on deterministic sets. The main difficulty is the lack of a priori control on the self-consistent density of states as we do not impose any irreducibility condition on the variance matrix. Such condition has been present in all previous works. More precisely, we study *Kronecker random matrices.* These are block matrices that consist of a $K \times K$ block structure with blocks of size $N \times N$. Each of these blocks is a linear combination of finitely many Wigner-type matrices and random

[†] Chapter 7 below is a slightly modified version of the publication [16] which was obtained in joint work with László Erdős, Torben Krüger and Yuriy Nemish.

matrices with independent entries. These matrices are assumed to be independent but each matrix is allowed to appear in multiple blocks.

For any Kronecker random matrix X, we provide a monotonically increasing family of deterministic subsets $\mathbb{D}_\varepsilon$, $\varepsilon > 0$, of the complex plane and prove, under some normalization and moment conditions, that for each $\varepsilon > 0$, the spectrum of X is contained in $\mathbb{D}_\varepsilon$,

$$\mathrm{Spec}(X) \subset \mathbb{D}_\varepsilon \tag{7}$$

with very high probability for $N \to \infty$ and fixed K (see Theorem 7.2.4 below). In some situations, it is known that $\cap_{\varepsilon>0}\mathbb{D}_\varepsilon$ actually coincides with the support of the self-consistent density of states obtained from the Dyson equation (cf. Chapters 4, 6 and 9 below). We expect this to be true in much greater generality. Furthermore, we show a global law for any Hermitian Kronecker random matrix in the limit $N \to \infty$ and for fixed K in Theorem 7.2.7 below. Here, we assume that the Hermitian Kronecker matrix satisfies the same normalization and moment conditions as required for the proof of (7).

Owing to the lack of any irreducibility condition for the variance matrix, e.g. a lower bound on the individual variances, and the presence of correlations among the blocks, the self-consistent density of states ρ will not behave nicely in general. However, a sufficient a priori understanding of ρ was essential in all previous arguments. This can be circumvented by a careful analysis of the corresponding Dyson equation (see (8) below) for $z \notin \operatorname{supp}\rho$. On this set, the Dyson equation can still be analyzed and yields enough information to prove (7) and the global law for Hermitian Kronecker matrices.

CHAPTER 8: THE DYSON EQUATION WITH LINEAR SELF-ENERGY: SPECTRAL BANDS, EDGES AND CUSPS. In Chapter 8† below, we study the solution to the *Dyson equation with linear self-energy* (see (8) below) which generalizes the QVE as well as the (vector and matrix) Dyson equations mentioned previously or studied in [4, 5, 6]. We show detailed regularity properties of a measure induced by this solution. This measure is the analogue of the self-consistent density of states. Compared to previous works, the noncommutativity of the underlying algebra requires a novel perturbation expansion around

† Chapter 8 essentially agrees with the preprint [15] which is joint work with László Erdős and Torben Krüger.

a non-self-adjoint operator. Indeed, we consider a von Neumann algebra $\mathcal{A}$ with unit $\mathbb{1}$ and a faithful, normal, tracial state $\langle\,\cdot\,\rangle\colon \mathcal{A} \to \mathbb{C}$. Moreover, let $a = a^* \in \mathcal{A}$ be a self-adjoint element and $S\colon \mathcal{A} \to \mathcal{A}$ a positivity-preserving linear map which is symmetric with respect to the scalar product $(x, y) \mapsto \langle x^* y \rangle$ on $\mathcal{A}$. Here, S is called the *self-energy*. The Dyson equation (with linear self-energy)

$$-\,m(z)^{-1} = z\mathbb{1} - a + S[m(z)] \tag{8}$$

has a unique solution $m\colon \mathbb{H} \to \mathcal{A}$, $\mathbb{H} := \{z \in \mathbb{C}\colon \operatorname{Im} z > 0\}$, such that $\operatorname{Im} m(z) := (m(z) - m(z)^*)/(2\mathrm{i})$ is positive definite for all $z \in \mathbb{H}$ [96]. In fact, m is the Stieltjes transform of a measure on $\mathbb{R}$ with values in the positive semidefinite elements of $\mathcal{A}$ (see Proposition 8.2.1 below). Under suitable assumptions, we show that there is a 1/3-Hölder continuous function $v\colon \mathbb{R} \to \mathcal{A}$ such that

$$m(z) = \int_{\mathbb{R}} \frac{v(\tau)}{\tau - z}\,\mathrm{d}\tau$$

for all $z \in \mathbb{H}$. Furthermore, the function v is real-analytic, where it is positive, and has either a square root edge or a cubic root cusp, where it vanishes (cf. Theorem 8.2.5 below). In Theorem 8.7.1 below, we also obtain precise expansions of v close to all small local minima. The main difficulty compared to the singularity analysis of the QVE in [4] is the noncommutativity of the multiplication in $\mathcal{A}$. This leads to considerably more involved computations compared to [4] but also necessitates a perturbation expansion around a non-self-adjoint operator in place of the self-adjoint unperturbed operator from [4]. We also prove a novel *band mass formula* which relates the mass of $(-\infty, E]$ with respect to the probability density $\rho = \langle v \rangle$ for any $E \in \mathbb{R} \setminus \operatorname{supp} \rho$ to the limit $m(E + \mathrm{i}\eta)$ for $\eta \downarrow 0$ (cf. (8.2.10) below). In many cases, the band mass formula yields quantization results for the mass $\rho(U)$ of a *band* $U \subset \mathbb{R}$, i.e., U is a connected component of $\operatorname{supp} \rho$ (see Proposition 8.2.6 (ii) and Corollary 8.9.4).

The Dyson equation, (8), plays an important role in the analysis of large Hermitian random matrices. Let H be an $N \times N$ Hermitian random matrix with possibly non-centered and correlated entries. In this setup, bulk local laws have been obtained in [6, 56] under general conditions on the correlation decay of the entries of H. In fact, if we choose $\mathcal{A} = \mathbb{C}^{N\times N}$, $\langle\,\cdot\,\rangle$ the normalized trace on $\mathbb{C}^{N\times N}$, $a := \mathbb{E}H$ the expectation of H and $S[x] := \mathbb{E}[(H-a)x(H-a)]$ for $x \in \mathbb{C}^{N\times N}$ in (8) then the local laws in [6, 56] assert that the resolvent of H at $z \in \mathbb{H}$ is close to $m(z)$ as long as z is away from the spectral edges of the spectrum of H. In particular, the eigenvalue density of H is well approximated by the inverse Stieltjes transform ρ of $z \mapsto \langle m(z)\rangle$. Hence, ρ is the self-consistent density of states of H and the main results of Chapter 8 show that, under certain assumptions, $\rho = \langle v\rangle$ has the same regularity properties as the self-consistent density of states of Wigner-type matrices.

Chapter 9: Correlated random matrices: band rigidity and edge universality. In Chapter 9† below, we consider Hermitian random matrices with decaying correlations and general expectation, which generalize Wigner-type matrices. For these random matrices, we prove edge universality at all (possibly internal) regular edges. The edge universality at internal edges requires *band rigidity*, i.e., the absence of whatsoever discrepancy between the number of eigenvalues in a band and its mass, which is the key novelty for these general random matrix models. Even for Wigner-type matrices, self-consistent densities of states with multiple support intervals become ubiquitous. Therefore, band rigidity is necessary to obtain edge universality at all regular edges.

More precisely, we first extend the bulk local laws from [6, 56] to regular spectral edges by applying the results of Chapter 8. Then we use the band mass formula from Chapter 8, the local law and an interpolation argument to establish band rigidity for Hermitian random matrices with decaying correlations (compare Corollary 9.2.5 below). The band rigidity crucially strengthens the customary eigenvalue rigidity (cf. (4)).

In the mid 1990's, Tracy and Widom computed the distribution of the (appropriately rescaled) fluctuation of the largest eigenvalue of the *Gaussian unitary ensemble* around 2 in the limit when the matrix size tends to infinity [148]. The Gaussian unitary ensemble refers to a complex Hermitian Wigner matrix with Gaussian distributed entries. Since then, for many complex Hermitian random matrix ensembles, the eigenvalues at regular spectral edges have been shown to follow this Tracy-Widom distribution. This phenomenon is called edge universality. For the symmetry class of real symmetric random matrices,

† Chapter 9 presents the preprint [17] which was written in joint work with László Erdős, Torben Krüger and Dominik Schröder.

there is a similar development originating from the work of Tracy and Widom in [149]. Combining the edge local law and the band rigidity in Chapter 9 as well as the recent results on the edge statistics of Dyson Brownian motion in [103] implies Tracy-Widom statistics of the extreme eigenvalue at each regular edge (compare Theorem 9.2.7 below).

2.1. Outlook

We complete these introductory chapters with an outlook on two long standing open problems in random matrix theory, the universality for non-Hermitian random matrices and the metal-insulator phase transition for random band matrices.

2.1.1. Universality of local spectral statistics of non-Hermitian random matrices. For Hermitian random matrices with independent entries, the universality of the local spectral statistics is rather well understood. The distributions of various local observables of eigenvalues, e.g. k-point correlation functions and gap statistics of bulk eigenvalues, fluctuations of extreme eigenvalues etc. have been identified for a rich class of these Wigner-type matrices. The common approach to these questions has two part: (i) the eigenvalue distribution is explicitly computed for a model with Gaussian distributed entries, (ii) more general models are shown to exhibit the same eigenvalue distribution as the Gaussian model, i.e., the distribution is *universal*.

Surprisingly, the corresponding questions for random matrices with independent entries without Hermitian symmetry are much harder to answer rigorously. Whereas part (i) of the strategy outlined before for Hermitian matrices can still be completed for many observables, part (ii) has only been obtained rigorously for rather restricted classes of models. For example, even for matrices with i.i.d. entries the universality of the k-point correlation functions has solely been proven under a strong condition of four matching moment with the corresponding Gaussian model [146]. The above mentioned statements for Wigner-type matrices do not need any moment matching conditions; exclusively the correct rescaling is required to obtain a universal distribution for very rich classes of Hermitian random matrices in the large matrix limit. A similar behaviour for non-Hermitian random matrices is also expected but has not been established rigorously yet.

2.1.2. Spectral statistics of random band matrices. A Hermitian $N \times N$ random matrix $H = (h_{ij})_{i,j=1}^N$ is a *random band matrix* of width W, $1 \leq W \leq N$, if $h_{ij} = 0$ for all $i, j \in [N]$ satisfying $|i - j| > W$. There is a dichotomy for the spectral statistics of H depending on the band width W. For large W, the spectral statistics of H agree with the random matrix statistics, e.g. eigenvector delocalization and strong correlations between nearby eigenvalues. This is called the *metal* or *conductor phase*. For small W, the eigenvectors of H are exponentially localized and the eigenvalues are essentially independent of each other. This is the *insulator phase*. Owing to a non-rigorous supersymmetric analysis, a sharp phase transition between these two regimes is expected at $W \approx \sqrt{N}$ [78].

We refrain from providing an exhaustive overview of the literature here and only list the strongest results towards this conjecture; we refer to [42] for a recent more detailed overview. In case the band matrix has Gaussian entries with a special variance and block structure a sharp phase transition on the level of two point correlation function of the characteristic polynomial can be seen at $W \approx \sqrt{N}$ [128, 130]. In the general case, random matrix statistics including eigenvector delocalization has been established for $W \gg N^{3/4}$ in [cite Bourgade Yau Yin]. This is the strongest upper bound on the critical band width. The strongest lower bound has been verified in [126], where eigenvector localization for $W \ll N^{1/8}$ has been proven. For a Gaussian model, this has been improved to $W \ll N^{1/7}$ in [121]. Prior to these results, numerous works have been devoted to upper and lower bounds on the critical band width, which shows that precisely localizing this band width is an intriguing and attractive problem in random matrix theory.

CHAPTER 3

The local semicircle law for random matrices with a fourfold symmetry

In this chapter, we present a slightly modified version of [12]. We consider real symmetric and complex Hermitian random matrices with the additional symmetry $h_{xy} = h_{N-y,N-x}$. The matrix elements are independent (up to the fourfold symmetry) and not necessarily identically distributed. This ensemble naturally arises as the Fourier transform of a Gaussian orthogonal ensemble (GOE). It also occurs as the flip matrix model – an approximation of the two-dimensional Anderson model at small disorder. We show that the density of states converges to the Wigner semicircle law despite the new symmetry type. We also prove the local version of the semicircle law on the optimal scale.

3.1. Introduction

In 1955, Wigner conjectured that the eigenvalues of large random matrices describe the energy levels of large atoms [157]. Therefore, the distribution of the eigenvalues of a random matrix is an interesting and often studied object in random matrix theory. For an $N \times N$ random matrix with eigenvalues $(\lambda_i)_{i=1}^N$, let $\mu_N := N^{-1} \sum_{i=1}^N \delta_{\lambda_i}$ be the *empirical spectral measure*. The celebrated Wigner semicircle law [157] asserts that μ_N converges to the semicircle law given by the density $\sqrt{(4-x^2)_+}/(2\pi)$ in the limit that the matrix size N goes to infinity.

The Wigner-Dyson-Mehta conjecture in [114] asserts that the distribution of the difference between consecutive eigenvalues of a large random matrix only depends on the symmetry type of the matrix and not on the distribution of the entries. This independence of the actual distribution is called universality. The proof of this conjecture by Erdős, Schlein, Yau and Yin in [64, 65] is built upon establishing a local semicircle law in the first step (see [69] for a review). An alternative approach was pursued by Tao and Vu in [144].

Wigner's semicircle law can be used to compute the number of eigenvalues contained in a fixed interval for a large random matrix. With the help of a local semicircle law such prediction can also be made in the case of a variable interval size as long as it is considerably bigger than N^{-1} which is the typical distance of neighbouring eigenvalues. A local semicircle law is most commonly proven by establishing a convergence of the Stieltjes transform $m_N(z) := N^{-1}\sum_{i=1}^N(\lambda_i - z)^{-1}$ of μ_N to the Stieltjes transform m of Wigner's semicircle law. Then an interval size of N^{-1} corresponds to showing the convergence when $\eta = \operatorname{Im} z$ is of this order.

One of the most general versions of a local semicircle law is presented in [60]. They suppose that the random matrix $H = (h_{xy})_{x,y}$ is complex Hermitian (or real symmetric), i.e., $h_{xy} = \bar{h}_{yx}$ for all x and y with real-valued random variables h_{xx} for all x such that $(h_{xy})_{x\le y}$ forms an independent family of centered random variables. Besides assuming that the variances $s_{xy} := \mathbb{E}|h_{xy}|^2$ of a row sum up to one, i.e,

$$\sum_y s_{xy} = 1 \tag{3.1.1}$$

for all x which ensures that the eigenvalues stay of order 1, the most important requirement is the independence of the entries (up to the symmetry constraint).

Many works in random matrix theory start with this independence assumption. However, some naturally arising random matrix models do not fulfill it. An example is the Fourier transform of a Gaussian Orthogonal Ensemble (GOE). For an $N \times N$ matrix $H = (h_{xy})_{x,y=1}^N$ the Fourier transform $\hat{H} = (\hat{h}_{pq})_{p,q\in\mathbb{Z}/N\mathbb{Z}}$ is defined through

$$\hat{h}_{pq} = \frac{1}{N}\sum_{x,y=1}^N h_{xy}\exp\left(-\mathrm{i}\frac{2\pi}{N}(px - qy)\right)$$

for $p, q \in \mathbb{Z}/N\mathbb{Z}$. If $H = (h_{xy})_{x,y=1}^N$ is a real symmetric matrix then $\hat{H} = (\hat{h}_{pq})_{p,q\in\mathbb{Z}/N\mathbb{Z}}$ fulfills the relations

$$\hat{h}_{pq} = \overline{\hat{h}}_{qp} = \hat{h}_{-q,-p} = \overline{\hat{h}}_{-p,-q}$$

for $p, q \in \mathbb{Z}/N\mathbb{Z}$. If the entries of H are, in addition, centered Gaussian distributed random variables such that $\{h_{xy}; x \le y\}$ are independent with $\mathbb{E}h_{xx}^2 = 2\mathbb{E}h_{xy}^2$ for $x \ne y$

then the entries of $\hat{H}$ will be independent up to this symmetry which we call *fourfold symmetry.*

Interestingly, this symmetry also arises in random matrix approximations of the Anderson model. In [32], it is argued that the fourfold symmetry with a constant diagonal – called the flip symmetry – is a good approximation of the two-dimensional Anderson model in the regime of small disorder (see [54] for a review on random matrix models of the Anderson model).

The first local law for Wigner matrices on the optimal scale $\eta \approx N^{-1}$ (with logarithmic corrections) in the bulk has been proven by Erdős, Schlein and Yau in [63]. In [72], Erdős, Yau and Yin proved that $m_N - m$ is of the optimal order $(N\eta)^{-1}$ in the bulk and they could extend this result to the edges in [71]. In the more general case with non-identical variances and the assumption (3.1.1), a local semicircle law on the scale $\eta \approx M^{-1}$ with $M := (\max_{x,y} s_{xy})^{-1}$ has been established by Erdős, Yau and Yin in [70]. For this case, Erdős, Knowles, Yau and Yin obtained the optimal order $(M\eta)^{-1}$ of $m_N - m$ in [60] even at the edge. A more detailed overview of the historical development of the local semicircle law can be found in Section 2.1 of [57].

Our main result is a proof of the local semicircle law for random matrices possessing the fourfold symmetry. Despite the different symmetry type compared to the case in [60] the limiting distribution of the empirical spectral measure will still be Wigner's semicircle law. The basic structure of the proof follows [60]. The main novelty is that not only the diagonal elements of the Green function have to be treated separately from the offdiagonal ones, but elements on the counterdiagonal need to be estimated separately via a new self-consistent equation.

We conclude this introduction with an outline of the structure of the present article. In the following section, we introduce our model and some notation and state our main result. In Section 3.3, we prove that the Fourier transform of a GOE satisfies the assumptions of Theorem 3.2.3. The remaining part is devoted to the proof of our main result. Section 3.4 contains a collection of the tools used in the proof which is given in the subsequent section. In Section 3.6, we show that the fluctuation averaging holds true for the fourfold symmetry as well.

Acknowledgement. I am very grateful to László Erdős for drawing my attention to this question, for suggesting the method and for numerous helpful comments during the preparation of this article. Moreover, I thank Oskari Ajanki and Torben Krüger for useful discussions.

3.2. Main Result

For $N \in \mathbb{N}$ and $x, y \in \mathbb{Z}/N\mathbb{Z}$, let $\zeta_{xy}^{(N)}$ be real or complex valued random variables (in the following we drop the N-dependence in our notation) such that ζ_{xx} is real valued, $\mathbb{E}\zeta_{xy} = 0$ and $\mathbb{E}|\zeta_{xy}|^2 = 1$ for all x, y. Moreover, we assume that for every $p \in \mathbb{N}$ there is a constant μ_p such that

$$\mathbb{E}|\zeta_{xy}|^p \le \mu_p \tag{3.2.1}$$

for all $x, y \in \mathbb{Z}/N\mathbb{Z}$ and $N \in \mathbb{N}$. For fixed $N \in \mathbb{N}$, the entries are supposed to be independent up to the fourfold symmetry $\zeta_{xy} = \bar{\zeta}_{yx} = \zeta_{-y,-x} = \bar{\zeta}_{-x,-y}$ for all $x, y \in \mathbb{Z}/N\mathbb{Z}$.

For $N \in \mathbb{N}$, let $S = (s_{xy})_{x,y \in \mathbb{Z}/N\mathbb{Z}}$ be an $N \times N$-matrix of nonnegative real numbers such that $s_{xy} = s_{yx} = s_{-y,-x} = s_{-x,-y}$ for all x, y and S is stochastic, i.e., for every x we have

$$\sum_y s_{xy} = 1. \tag{3.2.2}$$

Furthermore, we assume that the N-dependent parameter $M := (\max_{x,y} s_{xy})^{-1}$ satisfies

$$N^\delta \le M \le N \tag{3.2.3}$$

for some $\delta > 0$. Note that the first estimate is an assumption on S whereas the second bound follows from the definition of M and (3.2.2).

Defining $h_{xy} := s_{xy}^{1/2}\zeta_{xy}$ we obtain the Hermitian random matrix $H^{(N)} = (h_{xy})_{x,y \in \mathbb{Z}/N\mathbb{Z}}$ which fulfills the following fourfold symmetry

$$h_{xy} = \bar{h}_{yx} = h_{-y,-x} = \bar{h}_{-x,-y} \tag{3.2.4}$$

because of the definition of ζ_{xy} and the conditions on S. By definition, S describes the variances of $H^{(N)}$.

Let ρ denote Wigner's semicircle law and m its Stieltjes transform, i.e.,

$$\rho(x) := \frac{1}{2\pi}\sqrt{(4-x^2)_+}, \qquad m(z) := \frac{1}{2\pi}\int_{-2}^{2}\frac{\sqrt{4-x^2}}{x-z}\mathrm{d}x \tag{3.2.5}$$

for $x \in \mathbb{R}$ and $z \in \mathbb{C}\backslash\mathbb{R}$. For the real and imaginary part of $z \in \mathbb{C}$, we will use the abbreviations E and η, respectively, i.e., $z = E + \mathrm{i}\,\eta$ with $E, \eta \in \mathbb{R}$.

With this definition the complex valued function $m(z)$ is the unique solution of

$$m(z) + \frac{1}{m(z)+z} = 0 \tag{3.2.6}$$

such that $\operatorname{Im} m(z) > 0$ for $\eta > 0$. Denoting the resolvent or Green function of H by

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

and its entries by $G_{ij}(z)$ for $z \in \mathbb{C}\backslash\mathbb{R}$ we obtain for the Stieltjes transform m_N of the empirical spectral measure

$$m_N(z) = \frac{1}{N}\operatorname{Tr} G(z).$$

We use the definitions of stochastic domination and spectral domain given in [60].

Definition 3.2.1 (Stochastic Domination)**.** Let $X = (X^{(N)}(u); u \in U^{(N)}, N \in \mathbb{N})$ and $Y = (Y^{(N)}(u); u \in U^{(N)}, N \in \mathbb{N})$ be two families of nonnegative random variables for a possibly N-dependent parameter set $U^{(N)}$. We say that X is *stochastically dominated* by Y, uniformly in u, if for all $\varepsilon > 0$ and $D > 0$ there is a $N_0(\varepsilon, D) \in \mathbb{N}$ such that

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[X^{(N)}(u) > N^{\varepsilon} Y^{(N)}(u)\right] \le N^{-D}$$

for all $N \ge N_0$. In this case, we use the notation $X \prec Y$. If X is a family consisting of complex valued random variables and $|X| \prec Y$ then we write $X \in O_{\prec}(Y)$.

The definition of stochastic domination implies the following estimate which is important for our arguments

$$|h_{xy}| \prec s_{xy}^{1/2} \le M^{-1/2}. \tag{3.2.7}$$

Definition 3.2.2. An N-dependent family $\mathbf{D} = (\mathbf{D}^{(N)})_{N \in \mathbb{N}}$ of subsets of the complex plane with

$$\mathbf{D}^{(N)} \subset \{z = E + \mathrm{i}\,\eta \in \mathbb{C}; E \in [-10, 10], M^{-1} \le \eta \le 10\}$$

for every $N \in \mathbb{N}$ is called a *spectral domain*.

In analogy to the matrix S, we define $R = (r_{xy}) = (\mathbb{E}h_{xy}^2)_{y\neq -y}^{x\neq -x}$. If N is odd then R is an $(N-1)\times(N-1)$ matrix, otherwise it is an $(N-2)\times(N-2)$ matrix. For $\eta > 0$, we introduce the corresponding two control parameters

$$\Gamma_S(z) := \|(1-m^2(z)S)^{-1}\|_{\ell^\infty\to\ell^\infty}, \qquad \Gamma_R(z) := \|(1-m^2(z)R)^{-1}\|_{\ell^\infty\to\ell^\infty} \tag{3.2.8}$$

and their maximum $\Gamma(z) := \max\{\Gamma_S(z), \Gamma_R(z)\}$ (Note that Γ_S is denoted by Γ in [60]).

For the definition of the spectral domain underlying our estimates, we define

$$\eta_E := \min\left\{\eta;\, \frac{1}{M\eta} \le \min\left\{\frac{M^{-\gamma}}{\Gamma(z)^3}, \frac{M^{-2\gamma}}{\Gamma(z)^4 \operatorname{Im} m(z)}\right\} \text{ for all } z \in [E+\mathrm{i}\,\eta, E+\mathrm{i}\,10]\right\} \tag{3.2.9}$$

for $\gamma \in (0,1/2)$ and $E \in \mathbb{R}$. Then, for $\gamma \in (0,1/2)$ the spectral domain $\mathbf{S} \equiv \mathbf{S}(\gamma) = (\mathbf{S}^{(N)})_{N\in\mathbb{N}}$ is defined as

$$\mathbf{S}^{(N)} := \{E+\mathrm{i}\,\eta; |E| \le 10, \eta_E \le \eta \le 10\}\,. \tag{3.2.10}$$

Note that the spectral domain $\mathbf{S}$ differs from the spectral domain $\mathbf{S}$ in [60] due to the new definition of $\Gamma(z)$. Besides this difference the following main result of this article has the same form as Theorem 5.1 in [60].

Theorem 3.2.3 (Local Semicircle Law)**.** *Let H be a random matrix with the fourfold symmetry* (3.2.4) *such that the conditions* (3.2.1) *and* (3.2.2) *are fulfilled. For $\gamma \in (0,1/2)$, we have*

$$|G_{xy}(z) - \delta_{xy}m(z)| \prec \sqrt{\frac{\operatorname{Im} m(z)}{M\eta}} + \frac{1}{M\eta} \tag{3.2.11}$$

uniformly in x, y and $z \in \mathbf{S}$, as well as

$$|m_N(z) - m(z)| \prec \frac{1}{M\eta} \tag{3.2.12}$$

uniformly in $z \in \mathbf{S}$.

The proof of our main result is based on studying self-consistent equations in the same way as the proof of Theorem 5.1 in [60] which uses one self-consistent equation for $G_{xx}-m$. However, due to the fourfold symmetry it is no longer possible to directly show that the entries $G_{x,-x}$ are small as in [60]. Therefore, we introduce a second, new self-consistent equation for $G_{x,-x}$. While deriving these self-consistent equations we will see that the expressions $G_{xx} - m$ for $x \in \mathbb{Z}/N\mathbb{Z}$ and $G_{x,-x}$ for $x \neq -x$ are connected among each other via $\mathbb{E}|h_{xa}|^2$ and $\mathbb{E}h_{xa}^2$, respectively. Therefore, we introduce the matrix R in an analogous fashion as S is introduced in [60]. The corresponding control parameters Γ_R and Γ_S will appear in our estimates in Section 3.5.3. Whereas the latter control parameter is present in [60] and denoted by Γ in there, the matrix R and the corresponding parameter Γ_R are new in our work. The role of Γ in [60] is filled by the maximum $\Gamma(z) = \max\{\Gamma_S(z), \Gamma_R(z)\}$. Estimates on Γ similar to the ones in [60] are collected in Lemma 3.4.8 and Remark 3.4.9.

Remark 3.2.4. If the random variables h_{xy} are complex valued with $\mathbb{E}h_{xy}^2 = 0$ for all $x \neq y$ then $\Gamma_R(z) \leq C\Gamma_S(z)$ for $z \in \{E + \mathrm{i}\,\eta; E \in [-10, 10], \eta \in (0, 10]\}$ and therefore we can replace Γ by Γ_S in (3.2.9). Thus, in this case, our estimates hold on the spectral domain used in Theorem 5.1 in [60].

To have a shorter notation in the following arguments, we introduce the z-dependent stochastic control parameters

$$\begin{aligned} &\Lambda_d(z) := \max_x |G_{xx}(z) - m(z)|, && \Lambda_g(z) := \max_{x \neq y \neq -x} |G_{xy}(z)|, \\ &\Lambda_-(z) := \max_{x \neq -x} |G_{x,-x}(z)|, && \\ &\Lambda_o(z) := \max\{\Lambda_g(z), \Lambda_-(z)\}, && \Lambda(z) := \max\{\Lambda_d(z), \Lambda_o(z)\}. \end{aligned} \tag{3.2.13}$$

Compared to [60] we added the control parameter Λ_- since the off-diagonal terms $G_{x,-x}$ will be estimated differently than the generic off-diagonal terms.

3.3. Fourier Transform of Random Matrices

In this section, we give an example of a random matrix satisfying the conditions of Theorem 3.2.3, namely the Fourier transform (in the following sense) of a Gaussian orthogonal ensemble.

Definition 3.3.1 (Fourier Transform)**.** Let $H = (h_{xy})_{x,y=1}^N$ be an $N \times N$ matrix. The *Fourier transform* $\hat{H} = (\hat{h}_{pq})_{p,q \in \mathbb{Z}/N\mathbb{Z}}$ is the $N \times N$ matrix whose entries are given by

$$\hat{h}_{pq} = \frac{1}{N} \sum_{x,y=1}^{N} h_{xy} \exp\left(-\mathrm{i}\,\frac{2\pi}{N}(px - qy)\right)$$

for $p, q \in \mathbb{Z}/N\mathbb{Z}$.

In the next Lemma we collect the basic properties of the Fourier transform of a Gaussian orthogonal ensemble which will imply the conditions of Theorem 3.2.3.

Lemma 3.3.2. *Let H be a GOE and $\hat{H}$ its Fourier transform. Then the entries $\hat{h}_{pq}$ and $\hat{h}_{rs}$ are independent if and only if*

$$(p,q) \notin \{(r,s),(s,r),(-r,-s),(-s,-r)\}.$$

Moreover, $\hat{H}$ satisfies the fourfold symmetry (3.2.4) *for all $p, q \in \mathbb{Z}/N\mathbb{Z}$. We have*

$$\mathbb{E}|\hat{h}_{pq}|^2 = N^{-1}, \quad \mathbb{E}\hat{h}_{pr}^2 = 0 \tag{3.3.1}$$

for all q and $p \neq r$.

PROOF. To prove the if-part it suffices to show that $\hat{H}$ satisfies (3.2.4) which is a direct consequence of the fact that H is symmetric.

Since $\hat{h}_{pq}$ and $\hat{h}_{rs}$ are jointly normally distributed and $\mathbb{E}\hat{h}_{pq} = \mathbb{E}\hat{h}_{rs} = 0$, it suffices to prove that $\mathbb{E}\hat{h}_{pq}\overline{\hat{h}_{rs}} = 0$ and $\mathbb{E}\hat{h}_{pq}\hat{h}_{rs} = 0$ in order to show that these random variables are independent. The formula $\mathbb{E}h_{x_1y_1}h_{x_2y_2} = N^{-1}(\delta_{x_1x_2}\delta_{y_1y_2} + \delta_{x_1y_2}\delta_{y_1x_2})$ together with

$$\sum_{x=1}^{N} \exp\left(-\mathrm{i}\,\frac{2\pi}{N}mx\right) = \begin{cases} N, & m = 0, \\ 0, & \text{otherwise} \end{cases}$$

for $m \in \mathbb{Z}/N\mathbb{Z}$ yields $\mathbb{E}\hat{h}_{pq}\hat{h}_{rs} = N^{-1}$ for $(p,q) \in \{(s,r),(-r,-s)\}$ and $\mathbb{E}\hat{h}_{pq}\hat{h}_{rs} = 0$ otherwise. Thus, $\mathbb{E}\hat{h}_{pq}\hat{h}_{rs} \neq 0$ if and only if $(p,q) \in \{(s,r),(-r,-s)\}$. In particular, $\mathbb{E}\hat{h}_{pq}^2 = 0$ for $p \neq q$.

The relation $\overline{\hat{h}_{rs}} = \hat{h}_{sr}$ implies the first part of (3.3.1) and concludes the proof of the only-if part. □

Therefore, the Fourier transform of a Gaussian orthogonal ensemble fulfills all requirements of Theorem 3.2.3 with $s_{pq} := N^{-1}$ and $\zeta_{pq} := N^{-1/2}\hat{h}_{pq}$. Because of the first result in (3.3.1) the condition (3.2.2) is fulfilled. By the second part of (3.3.1) Remark 3.2.4 is applicable. Thus, the local semicircle law holds true for these random matrices.

3.4. Tools

In this section, we collect the tools for the proof of Theorem 3.2.3. We start with listing some resolvent identities which are the basic tool for all our estimates as they encode the dependences between diagonal and off-diagonal entries of the resolvents. Computing the partial expectation of certain terms in expansions of the resolvent entries with respect to a minor will be an important step to derive the self-consistent equations. Thus, we introduce some notation in the second subsection. We conclude with the fluctuation averaging, an important mechanism to improve some bounds, and some estimates on m and Γ which are frequently used in our proofs.

3.4.1. Minors and Resolvent Identities. Let $H = (h_{xy})_{x,y \in \mathbb{Z}/N\mathbb{Z}}$ be a Hermitian matrix and $\mathbb{T} \subset \mathbb{Z}/N\mathbb{Z}$.

Definition 3.4.1. We define the $N \times N$ matrix $H^{(\mathbb{T})}$ and its *resolvent* or *Green function* $G^{(\mathbb{T})}$ through

$$(H^{(\mathbb{T})})_{ij} := \mathbf{1}(i \notin \mathbb{T})\mathbf{1}(j \notin \mathbb{T})h_{ij}, \quad G^{(\mathbb{T})}(z) := (H^{(\mathbb{T})} - z)^{-1}$$

for $i, j \in \mathbb{Z}/N\mathbb{Z}$ and for $z \in \mathbb{C}\backslash\mathbb{R}$. We denote the entries of $G^{(\mathbb{T})}(z)$ by $G_{ij}^{(\mathbb{T})}(z)$. We set

$$\sum_i^{(\mathbb{T})} := \sum_{i; i \notin \mathbb{T}} .$$

In both cases, we write $(a_1, \dots, a_n, \mathbb{T})$ for $(\{a_1, \dots, a_n\} \cup \mathbb{T})$.

Note that $H^{(\mathbb{T})}$ is still a Hermitian $N \times N$ matrix, in particular $G^{(\mathbb{T})}$ exists. To estimate the resolvent entries we make essential use of the following relations.

Lemma 3.4.2 (Resolvent Identities)**.** *For $i, j, k \notin \mathbb{T}$, the following statements hold:*

$$\frac{1}{G_{ii}^{(\mathbb{T})}} = h_{ii} - z - \sum_{a,b}^{(\mathbb{T},i)} h_{ia} G_{ab}^{(\mathbb{T},i)} h_{bi}. \tag{3.4.1}$$

If $i, j \neq k$ then

$$G_{ij}^{(\mathbb{T})} = G_{ij}^{(\mathbb{T},k)} + \frac{G_{ik}^{(\mathbb{T})} G_{kj}^{(\mathbb{T})}}{G_{kk}^{(\mathbb{T})}}, \quad \frac{1}{G_{ii}^{(\mathbb{T})}} = \frac{1}{G_{ii}^{(\mathbb{T},k)}} - \frac{G_{ik}^{(\mathbb{T})} G_{ki}^{(\mathbb{T})}}{G_{ii}^{(\mathbb{T})} G_{ii}^{(\mathbb{T},k)} G_{kk}^{(\mathbb{T})}}. \tag{3.4.2}$$

If $i \neq j$ then

$$G_{ij}^{(\mathbb{T})} = -G_{ii}^{(\mathbb{T})} \sum_{a}^{(\mathbb{T},i)} h_{ia} G_{aj}^{(\mathbb{T},i)} = -G_{jj}^{(\mathbb{T})} \sum_{a}^{(\mathbb{T},j)} G_{ia}^{(\mathbb{T},j)} h_{aj}. \tag{3.4.3}$$

The proof of Schur's complement formula, (3.4.1), and the first identity in (3.4.2) can be found in Lemma 4.2 in [70] and the second identity follows directly from the first one. Lemma 6.10 in [59] contains a proof of (3.4.3).

Moreover, if $\eta > 0$ then the spectral theorem for self-adjoint matrices yields

$$\sum_{l} |G_{kl}^{(\mathbb{T})}(z)|^2 = \frac{1}{\eta} \operatorname{Im} G_{kk}^{(\mathbb{T})}(z). \tag{3.4.4}$$

This identity is sometimes called *Ward identity*.

The functional calculus implies the following estimates on the entries of the resolvent:

$$|G_{ij}^{(\mathbb{T})}(z)| \leq \eta^{-1} \leq M \tag{3.4.5}$$

for $\eta > 0$ and all $i, j \in \mathbb{Z}/N\mathbb{Z}$. The second estimate holds if $z \in \mathbf{D}$ where $\mathbf{D}$ is a spectral domain.

3.4.2. Partial Expectation. For the partial expectation with respect to the σ-algebra generated by $H^{(x,-x)}$, we introduce the following notation.

Definition 3.4.3 (Partial Expectation)**.** Let X be an integrable random variable. For $x \in \mathbb{Z}/N\mathbb{Z}$ we define the random variables $\mathbb{E}_x X$ and $\mathbb{F}_x X$ through

$$\mathbb{E}_x X := \mathbb{E}[X | H^{(x,-x)}], \quad \mathbb{F}_x X := X - \mathbb{E}_x X.$$

The random variable $\mathbb{E}_x X$ is called the *partial expectation* of X with respect to x.

The symbols $\mathbb{E}_x$ and $\mathbb{F}_x$ are the analogues of P_i and Q_i in [60] that were defined by considering the minor $H^{(i)}$. Due to the fourfold symmetry column x, $-x$ and row x, $-x$ contain the same information, so the conditional expectation is taken with respect to the minor $H^{(x,-x)}$. Notice that it may happen that $x = -x$, in which case $H^{(x,-x)}$ is an $(N-1) \times (N-1)$ minor.

Definition 3.4.4 (Independence)**.** We say that the integrable random variable X is *independent* of $\mathbb{T} \subset \mathbb{Z}/N\mathbb{Z}$ if $X = \mathbb{E}_x X$ for all $x \in \mathbb{T}$.

If Y is independent of x then $\mathbb{F}_x(X)Y = XY - \mathbb{E}_x(X\mathbb{E}_x Y) = \mathbb{F}_x(XY)$ and therefore

$$\mathbb{E}\mathbb{F}_x(X)Y = \mathbb{E}\mathbb{F}_x(XY) = \mathbb{E}(XY) - \mathbb{E}\mathbb{E}_x(XY) = 0. \tag{3.4.6}$$

3.4.3. Fluctuation Averaging. Let $\mathbf{D}$ be a spectral domain, H satisfy the requirements of Theorem 3.2.3 and Ψ a deterministic (possibly z-dependent) control parameter which satisfies

$$M^{-1/2} \leq \Psi \leq M^{-c} \tag{3.4.7}$$

for all $z \in \mathbf{D}$ and for some $c > 0$.

The aim of the fluctuation averaging is to estimate linear combinations of the form $\sum_k t_{ik} X_k$ with special random variables X_k and a family of complex weights $T = (t_{ik})$ that satisfy

$$0 \leq |t_{ik}| \leq M^{-1}, \quad \sum_k |t_{ik}| \leq 1. \tag{3.4.8}$$

Note that the family T may be N-dependent. Examples of such weights are given by $t_{ik} = s_{ik} = \mathbb{E}|h_{ik}|^2$, $t_{ik} = N^{-1}$ or $t_{ik} = r_{ik} = \mathbb{E}h_{ik}^2$. Recall that $\Lambda(z) = \max_{x,y}|G_{xy}(z) - \delta_{xy} m(z)|$ which is the basic quantity we want to estimate (cf. (3.2.13)).

Theorem 3.4.5 (Fluctuation Averaging)**.** *Let* $\mathbf{D}$ *be a spectral domain,* Ψ *a deterministic control parameter satisfying* (3.4.7) *and* $T = (t_{ik})$ *a weight satisfying* (3.4.8)*. If* $\Lambda \prec \Psi$ *then*

$$\left|\sum_k t_{ik} \mathbb{F}_k \frac{1}{G_{kk}}\right| \prec \Psi^2, \quad \left|\sum_k t_{ik} \mathbb{F}_k G_{kk}\right| \prec \Psi^2, \quad \left|\sum_{k \neq -k} t_{ik} \mathbb{F}_k G_{k,-k}\right| \prec \Psi^2 \tag{3.4.9}$$

uniformly in i *and* $z \in \mathbf{D}$*. If* $\Lambda \prec \Psi$ *and* T *commutes with* S *then we have*

$$\left|\sum_k t_{ik}(G_{kk} - m)\right| \prec \Gamma_S \Psi^2 \tag{3.4.10}$$

uniformly in i *and* $z \in \mathbf{D}$*. If* $\Lambda \prec \Psi$ *and* T *commutes with* R *then we have*

$$\left|\sum_{k \neq -k} t_{ik} G_{k,-k}\right| \prec \Gamma_R \Psi^2 \tag{3.4.11}$$

uniformly in i *and* $z \in \mathbf{D}$*.*

A similar result was proven in [60], but due to the fourfold symmetry we need the third estimate in (3.4.9) and (3.4.11) which were not present there. For the first estimate in (3.4.9), there is the following stronger bound assuming that there is a stronger a priori bound on the off-diagonal terms, i.e., on $\Lambda_o(z) = \max_{x\neq y}|G_{xy}(z)|$ (cf. (3.2.13)):

Theorem 3.4.6. *Let* $\mathbf{D}$ *be a spectral domain,* Ψ *and* Ψ_o *deterministic control parameters satisfying* (3.4.7) *and* $T = (t_{ik})$ *a weight satisfying* (3.4.8)*. If* $\Lambda \prec \Psi$ *and* $\Lambda_o \prec \Psi_o$ *then*

$$\left|\sum_k t_{ik}\mathbb{F}_k \frac{1}{G_{kk}}\right| \prec \Psi_o^2 \tag{3.4.12}$$

uniformly in i *and* $z \in \mathbf{D}$.

The proof of Theorem 3.4.5 and 3.4.6 can be found in Section 3.6.

3.4.4. Estimates on m and Γ. For convenience, we list some elementary estimates from [60] which are often used in the following proofs.

Lemma 3.4.7. *There is a constant* $c > 0$ *such that for* $z \in \{E + i\eta; E \in [-10, 10], \eta \in (0, 10]\}$ *we have*

$$c \le |m(z)|, \quad |m(z)| \le 1 - c\eta, \quad |m(z)| \le \eta^{-1}, \quad \operatorname{Im} m(z) \ge c\eta. \tag{3.4.13}$$

Since $\Gamma \ge \Gamma_S$ it suffices to prove the following lower bounds on Γ for Γ_S.

Lemma 3.4.8. *There is a constant* $c > 0$ *such that*

$$c \le \Gamma(z), \quad |1 - m^2(z)|^{-1} \le \Gamma(z) \tag{3.4.14}$$

for all $z \in \{E + i\eta; E \in [-10, 10], \eta \in (0, 10]\}$.

Remark 3.4.9. Since $\|R\|_{\ell^\infty\to\ell^\infty} \le 1$ the proof of Proposition A.2 in [60] yields that

$$\Gamma_R(z) \le \frac{C\log N}{1 - \max_\pm\left|\frac{1\pm m^2}{2}\right|} \le \frac{C\log N}{\min\{\eta + E^2, \theta\}}$$

for $z \in \{E + \mathrm{i}\,\eta; -10 \le E \le 10, M^{-1} \le \eta \le 10\}$ with

$$\theta \equiv \theta(z) := \begin{cases} \kappa + \frac{\eta}{\sqrt{\kappa+\eta}}, & \text{if } |E| \le 2, \\ \sqrt{\kappa+\eta}, & \text{if } |E| > 2, \end{cases}$$

and $\kappa := ||E| - 2|$.

3.5. Proof of the Main Result

This section contains the proof of our main result, Theorem 3.2.3. First, we establish the two self-consistent equations which will be the basis of all our estimates. In Section 3.5.2, we bound the error terms in these self-consistent equations so that we can use them to prove a preliminary bound on the central quantity Λ (cf. (3.2.13)) in Section 3.5.3. Finally, we complete the proof of Theorem 3.2.3 in Section 3.5.4 by iteratively improving the preliminary bound from the previous section.

3.5.1. Self-consistent Equations. The goal of this section is to establish the two self-consistent equations for the difference $G_{xx} - m$ and for the off-diagonal terms $G_{x,-x}$. As the matrices are indexed by elements in $\mathbb{Z}/N\mathbb{Z}$ it might happen that $x = -x$ for $x \in \mathbb{Z}/N\mathbb{Z}$, more precisely we have $0 = -0$ in $\mathbb{Z}/N\mathbb{Z}$ and moreover if N is even $N/2 = -N/2$. Since the expansion of the diagonal term G_{xx} by means of the resolvent identities is a bit different for $x = -x$ and in this cases the entry $G_{x,-x}$ is in fact a diagonal term we have to distinguish the two cases, $x \neq -x$ and $x = -x$, in the sequel.

Recall for the following lemma that $s_{xa} = \mathbb{E}|h_{xa}^2|$ and $r_{xa} = \mathbb{E}h_{xa}^2$.

Lemma 3.5.1. *For $v_x := G_{xx} - m$ we have the self-consistent equation*

$$-\sum_a s_{xa} v_a + \Upsilon_x = \frac{1}{v_x + m} - \frac{1}{m} \tag{3.5.1}$$

with the error term

$$\Upsilon_x = \begin{cases} h_{xx} + A_x - Z_x, & x = -x, \\ h_{xx} + A_x + B_x - C_x - Y_x - Z_x, & x \neq -x, \end{cases}$$

and the abbreviations

$$A_x := \sum_a s_{xa} \frac{G_{ax}G_{xa}}{G_{xx}}, \qquad B_x := \sum_a^{(x,-x)} s_{xa} \frac{G^{(x)}_{a,-x}G^{(x)}_{-x,a}}{G^{(x)}_{-x,-x}}, \tag{3.5.2}$$

$$C_x := \left(|h_{x,-x}|^2 - s_{-x,x}\right) G^{(x)}_{-x,-x} + h_{-x,x} \sum_a^{(x,-x)} h_{xa} G^{(x)}_{a,-x} + h_{x,-x} \sum_b^{(x,-x)} G^{(x)}_{-x,b} h_{bx}, \tag{3.5.3}$$

$$Y_x := \left(G^{(x)}_{-x,-x}\right)^{-1} \sum_{a,b}^{(x,-x)} h_{xa} G^{(x)}_{a,-x} G^{(x)}_{-x,b} h_{bx}, \tag{3.5.4}$$

$$Z_x := \begin{cases} \sum_{a,b}^{(x)} \mathbb{F}_x \left[h_{xa} G^{(x)}_{ab} h_{bx}\right], & x = -x, \\ \sum_{a,b}^{(x,-x)} \mathbb{F}_x \left[h_{xa} G^{(x,-x)}_{ab} h_{bx}\right], & x \neq -x. \end{cases} \tag{3.5.5}$$

The self-consistent equation for $G_{x,-x}$ *is given by*

$$G_{x,-x} = m^2 \sum_{a \neq -a} r_{xa} G_{a,-a} + \mathcal{E}_x, \tag{3.5.6}$$

for $x \neq -x$ *where we defined* $\mathcal{E}_x := \mathcal{E}^1_x + \mathcal{E}^2_x - \mathcal{E}^3_x - \mathcal{E}^4_x$ *with the error terms*

$$\begin{aligned} \mathcal{E}^1_x :=& - m^2 \sum_{a \in \{x,-x\}} r_{xa} G_{a,-a} + m^2 \sum_{a=-a} r_{xa} G_{aa} \\ &+ \left(G_{xx} G^{(x)}_{-x,-x} - m^2\right) \sum_a^{(x,-x)} r_{xa} G_{a,-a} - G_{xx} G^{(x)}_{-x,-x} h_{x,-x}, \\ \mathcal{E}^2_x :=& G_{xx} G^{(x)}_{-x,-x} \sum_a^{(x,-x)} \mathbb{F}_x \left[h_{xa} G^{(x,-x)}_{ab} h_{b,-x}\right], \\ \mathcal{E}^3_x :=& G^{(x)}_{-x,-x} \sum_a^{(x,-x)} r_{xa} G_{ax} G_{x,-a}, \qquad \mathcal{E}^4_x := G_{xx} \sum_a^{(x,-x)} r_{xa} G^{(x)}_{a,-x} G^{(x)}_{-x,-a}. \end{aligned}$$

The self-consistent equation (3.5.1) has the same form as (5.9) in [60] and it is proven in a similar way by expanding by means of Schur's complement formula and computing the partial expectation of a term in this expansion. However, we had to replace P_i by $\mathbb{E}_x$ to derive it and the error term Υ_x contains terms which did not appear in (5.8) from [60]. (If $x = -x$ then Υ_x has the same form as in [60].) The term A_x is exactly the same as A_i in (5.8) of [60]. The term Z_x is the analogue of Z_i in [60] but the terms B_x, C_x and Y_x are completely new and will require new estimates.

The self-consistent equation (3.5.6) is new and does not have a counterpart in [60]. Due to the fourfold symmetry there is the necessity to introduce it since in contrast to the symmetry studied in [60] proving directly that the off-diagonal elements $G_{x,-x}$ are small is not possible.

As deriving this self-consistent equation follows the same line as the proof of (3.5.1) – expanding and computing the partial expectation of a term in this expansion – it is not surprising that some error terms in (3.5.6) have counterparts in (3.5.1). Namely, $\mathcal{E}_x^2$ is the counterpart of Z_x. Moreover, $\mathcal{E}_x^3$ and $\mathcal{E}_x^4$ are the error terms corresponding to A_x and B_x, respectively.

Proof. We start with the proof of (3.5.1). For $x = -x$ the derivation of (3.5.1) follows exactly as (5.9) in Section 5.1 of [60] since $\mathbb{E}_x$ and $\mathbb{F}_x$ agree with P_x and Q_x respectively in this case. Similarly, for $x \neq -x$ the self-consistent equation (3.5.1) will be obtained from Schur's complement formula (3.4.1) with $\mathbb{T} = \emptyset$. In this case, its last term can be written in the form

$$\begin{aligned}\sum_{a,b}^{(x)} h_{xa} G_{ab}^{(x)} h_{bx} =& h_{x,-x} G_{-x,-x}^{(x)} h_{-x,x} + \sum_{a}^{(x,-x)} h_{xa} G_{a,-x}^{(x)} h_{-x,x} + \sum_{b}^{(x,-x)} h_{x,-x} G_{-x,b}^{(x)} h_{bx} \\ &+ \sum_{a,b}^{(x,-x)} h_{xa} G_{ab}^{(x,-x)} h_{bx} + \left(G_{-x,-x}^{(x)}\right)^{-1} \sum_{a,b}^{(x,-x)} h_{xa} G_{a,-x}^{(x)} G_{-x,b}^{(x)} h_{bx}\end{aligned} \tag{3.5.7}$$

by applying the resolvent identity (3.4.2). Since the random variables h_{xa} and $h_{-x,b}$ are independent of $H^{(x,-x)}$ we have $\mathbb{E}_x\left[h_{xa} G_{ab}^{(x,-x)} h_{bx}\right] = s_{xa} G_{aa}^{(x,-x)} \delta_{ab}$. Thus,

$$\begin{aligned}\sum_{a,b}^{(x,-x)} \mathbb{E}_x\left[h_{xa} G_{ab}^{(x,-x)} h_{bx}\right] =& \sum_{a}^{(x,-x)} s_{xa} G_{aa}^{(x,-x)} \\ =& \sum_a s_{xa} G_{aa} - \sum_a s_{xa} \frac{G_{ax} G_{xa}}{G_{xx}} - s_{-x,x} G_{-x,-x}^{(x)} - \sum_{a}^{(x,-x)} s_{xa} \frac{G_{a,-x}^{(x)} G_{-x,a}^{(x)}}{G_{-x,-x}^{(x)}},\end{aligned}$$

where we used in the second step the resolvent identity (3.4.2) twice. By splitting the fourth summand on the right-hand side of (3.5.7) according to $\mathbb{E}_x + \mathbb{F}_x = 1$, we get

$$\begin{aligned}\sum_{a,b}^{(x,-x)} h_{xa} G_{ab}^{(x,-x)} h_{bx} =& \sum_{a,b}^{(x,-x)} \mathbb{E}_x\left[h_{xa} G_{ab}^{(x,-x)} h_{bx}\right] + \sum_{a,b}^{(x,-x)} \mathbb{F}_x\left[h_{xa} G_{ab}^{(x,-x)} h_{bx}\right] \\ =& \sum_a s_{xa} G_{aa} - A_x - s_{-x,x} G_{-x,-x}^{(x)} - B_x + Z_x.\end{aligned} \tag{3.5.8}$$

Therefore, the results of (3.5.7) and (3.5.8) allow us to write (3.4.1) in the form

$$\frac{1}{G_{xx}} = -z - m + \Upsilon_x - \sum_a s_{xa} v_a,$$

which implies (3.5.1) using (3.2.6).

We fix $x \neq -x$. To derive (3.5.6) we apply the resolvent identity (3.4.3) twice to get

$$G_{x,-x} = -G_{xx} G^{(x)}_{-x,-x} h_{x,-x} + G_{xx} G^{(x)}_{-x,-x} \sum_{a,b}^{(x,-x)} h_{xa} G^{(x,-x)}_{ab} h_{b,-x}. \tag{3.5.9}$$

Since $\mathbb{E}_x h_{xa} G^{(x,-x)}_{ab} h_{b,-x} = G^{(x,-x)}_{a,-a} r_{xa} \delta_{b,-a}$ splitting up the sum in the second term in (3.5.9) according to $\mathbb{E}_x + \mathbb{F}_x = 1$ yields

$$G_{x,-x} = -G_{xx} G^{(x)}_{-x,-x} h_{x,-x} + G_{xx} G^{(x)}_{-x,-x} \sum_{a}^{(x,-x)} r_{xa} G_{a,-a} + \mathcal{E}^2_x - \mathcal{E}^3_x - \mathcal{E}^4_x \tag{3.5.10}$$

where we used the resolvent identity (3.4.2) twice. We obtain (3.5.6) by adding and substracting $m^2 \sum_a r_{xa} G_{a,-a}$ to the right-hand side of (3.5.10). □

3.5.2. Auxiliary Estimates. The next lemma contains bounds on the resolvent entries of minors of H if there exists an a priori bound on Λ (Recall its definition in (3.2.13)). We will use a deterministic (possibly z-dependent) parameter Ψ which fulfills

$$cM^{-\frac{1}{2}} \leq \Psi \leq M^{-c} \tag{3.5.11}$$

for some $c > 0$ and all large enough N.

Lemma 3.5.2. *Let* $\mathbf{D}$ *be a spectral domain and* φ *the indicator function of a (possibly z-dependent) event. Let* Ψ *be a deterministic control parameter satisfying* (3.5.11). *If* $\varphi\Lambda \prec \Psi$ *and* $\mathbb{T} \subset \mathbb{N}$ *is a fixed finite subset then*

$$\varphi |G^{(\mathbb{T})}_{ij}| \prec \varphi \Lambda_o \prec \Psi, \qquad \varphi |G^{(\mathbb{T})}_{ii}| \prec 1, \qquad \frac{\varphi}{|G^{(\mathbb{T})}_{ii}|} \prec 1,$$

$$\varphi |G^{(\mathbb{T})}_{ii} - m| \prec \varphi \Lambda, \qquad \varphi \operatorname{Im} G^{(\mathbb{T})}_{ii} \prec \operatorname{Im} m + \Lambda$$

uniformly in $z \in \mathbf{D}$ *and in* i, j *for* $i \neq j$ *and* $i, j \notin \mathbb{T}$.

Proof. This result follows by induction on the size of $\mathbb{T}$ using (3.4.13) and (3.4.2). □

Using this result we will establish the first bounds on the error terms in the self-consistent equations in the next lemma. When applying the first part of the following lemma the indicator φ will be defined precisely in such way that the condition $\varphi\Lambda \prec M^{-c}$ holds, i.e., to ensure that $\varphi\Lambda$ is small.

Lemma 3.5.3. *Let* $\mathbf{D}$ *be a spectral domain.*

(i) If φ *is an indicator function such that* $\varphi\Lambda \prec M^{-c}$ *(for some* $c > 0$*) then*

$$\varphi(\Lambda_g + |A_x| + |B_x| + |C_x| + |Y_x| + |Z_x|) \prec \varphi\Lambda^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}, \tag{3.5.12}$$

$$\varphi(|\mathcal{E}_x^1| + |\mathcal{E}_x^2| + |\mathcal{E}_x^3| + |\mathcal{E}_x^4|) \prec \varphi\Lambda^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}} \tag{3.5.13}$$

uniformly in x *and* $z \in \mathbf{D}$.

(ii) For fixed $\eta > 0$ *we have the estimates*

$$\Lambda_- \leq \eta^{-2}\Lambda_- + 2\eta^{-3}\Lambda_-^2 + \epsilon \tag{3.5.14}$$

with $\epsilon \prec M^{-1/2}$ *uniformly in* $z \in \{w \in \mathbb{C}; \operatorname{Im} w = \eta\}$*, and*

$$\Lambda_g \prec M^{-1/2} + \Lambda_-, \tag{3.5.15}$$

$$|A_x| + |B_x| + |C_x| + |Y_x| + |Z_x| \prec M^{-1/2} + \Lambda_o \tag{3.5.16}$$

uniformly in x *and in* $z \in \{w \in \mathbb{C}; \operatorname{Im} w = \eta\}$.

PROOF. In this proof we will occasionally split the index set of a summation into the parts $\{a \neq -a\}$ and $\{a = -a\}$ and use that the latter set contains at most two elements.

In the following proof of the first part Lemma 3.5.2 will be applied several times with $\Psi = M^{-c}$. Note that $M^{-1/2} \prec \sqrt{(\operatorname{Im} m + \Lambda)/(M\eta)}$ because of the fourth estimate in (3.4.13). First, we assume $x \neq -x$. Applying the second estimate in (3.2.7) and (3.2.2) to the definition of A_x in (3.5.2) yields

$$\varphi|A_x| \prec s_{xx}|G_{xx}| + \sum_a^{(x)} s_{xa}\varphi\frac{|G_{xa}G_{ax}|}{|G_{xx}|} \prec M^{-1} + \varphi\Lambda_o^2. \tag{3.5.17}$$

Similarly, using the first estimate in Lemma 3.5.2 we get $\varphi|B_x| \prec \varphi\Lambda_o^2$.

The representation

$$C_x = |h_{x,-x}|^2 G^{(x)}_{-x,-x} - s_{-x,x} G^{(x)}_{-x,-x} - \frac{G_{x,-x}}{G_{xx}} h_{-x,x} - h_{x,-x} \frac{G_{-x,x}}{G_{xx}}, \tag{3.5.18}$$

which follows from the resolvent identity (3.4.3), together with (3.2.7) implies

$$\varphi |C_x| \prec M^{-1/2}. \tag{3.5.19}$$

To estimate Y_x we need the following two auxiliary bounds: We have

$$\varphi \left| \sum_a^{(x,-x)} h^2_{xa} G^{(x,-x)}_{a,-a} \right| \le \sum_{a \neq -a}^{(x,-x)} |h_{xa}|^2 \varphi |G^{(x,-x)}_{a,-a}| + \sum_{a=-a}^{(x,-x)} |h_{xa}|^2 \varphi |G^{(x,-x)}_{aa}| \prec \varphi \Lambda_o + M^{-1}, \tag{3.5.20}$$

where we used (3.2.7) and (3.2.2) in last step. Now, we use the quadratic Large Deviation Bounds from [60] after conditioning on $G^{(x,-x)}$. By applying (C.4) in [60] with $X_k = \zeta_{xk}$ and $a_{kl} = s^{1/2}_{xk} G^{(x,-x)}_{k,-l} s^{1/2}_{xl}$ we get

$$\varphi \left| \sum_{k \neq l}^{(x,-x)} h_{xk} G^{(x,-x)}_{k,-l} h_{xl} \right|^2 \prec \sum_{k \neq l}^{(x,-x)} s_{xk} s_{xl} \varphi |G^{(x,-x)}_{k,-l}|^2 \prec \frac{\varphi}{M\eta} \sum_k^{(x,-x)} s_{xk} \operatorname{Im} G^{(x,-x)}_{kk} \prec \frac{\operatorname{Im} m + \Lambda}{M\eta}, \tag{3.5.21}$$

where we used the second estimate in (3.2.7) and (3.4.4) in the second step. Thus, the representation

$$Y_x = G^{(x)}_{-x,-x} \left(\sum_{a,k}^{(x,-x)} h_{xa} G^{(x,-x)}_{ak} h_{k,-x} \right) \left(\sum_{b,l}^{(x,-x)} h_{-x,l} G^{(x,-x)}_{lb} h_{bx} \right), \tag{3.5.22}$$

which follows from the resolvent identity (3.4.3), yields (after separating the case $k = -a$)

$$\begin{aligned} \varphi |Y_x| &\prec \varphi \left| \sum_a^{(x,-x)} h^2_{xa} G^{(x,-x)}_{a,-a} \right|^2 + \varphi \left| \sum_{a \neq k}^{(x,-x)} h_{xa} G^{(x,-x)}_{a,-k} h_{xk} \right|^2 \\ &\prec \varphi \Lambda_o^2 + \frac{\operatorname{Im} m + \Lambda}{M\eta} \prec \varphi \Lambda_o^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}. \end{aligned} \tag{3.5.23}$$

Before estimating Z_x, we conclude from its definition in (3.5.5) that

$$Z_x := \begin{cases} \sum_a^{(x)} \left(|h_{xa}|^2 - s_{xa} \right) G^{(x)}_{aa} + \sum_{a \neq b}^{(x)} h_{xa} G^{(x)}_{ab} h_{bx}, & x = -x, \\ \sum_a^{(x,-x)} \left(|h_{xa}|^2 - s_{xa} \right) G^{(x,-x)}_{aa} + \sum_{a \neq b}^{(x,-x)} h_{xa} G^{(x,-x)}_{ab} h_{bx}, & x \neq -x. \end{cases}$$

We fix $x \neq -x$ and apply (C.4) in [60] with $X_i = \zeta_{xi}$ and $a_{ij} = s_{xi}^{1/2} G_{ij}^{(x,-x)} s_{jx}^{1/2}$ to get

$$\varphi \left| \sum_{i \neq j}^{(x,-x)} h_{xi} G_{ij}^{(x,-x)} h_{jx} \right|^2 \prec \left(\sum_{i \neq j}^{(x,-x)} s_{xi} s_{jx} \varphi |G_{ij}^{(x,-x)}|^2 \right)^{1/2} \prec \frac{\operatorname{Im} m + \Lambda}{M\eta}, \tag{3.5.24}$$

where the last step follows in the same way as the last step in (3.5.21). Moreover, (C.2) in [60] with $X_i = (|\zeta_{xi}|^2 - 1)(\mathbb{E}|\zeta_{xi}|^4 - 1)^{-1/2}$ and $a_i = (\mathbb{E}|\zeta_{xi}|^4 - 1)^{1/2} s_{xi} G_{ii}^{(x,-x)}$ implies

$$\varphi \left| \sum_{i}^{(x,-x)} \Big(|h_{xi}|^2 - s_{xi} \Big) G_{ii}^{(x,-x)} \right|^2 \prec \sum_{i}^{(x,-x)} s_{xi}^2 (\mathbb{E}|\zeta_{xi}|^4 - 1) \varphi |G_{ii}^{(x,-x)}|^2 \prec M^{-1}, \tag{3.5.25}$$

where we used (3.2.1), the second estimate in (3.2.7) and (3.2.2) in the last step. Therefore, absorbing $M^{-1/2}$ into the second summand we get

$$\varphi |Z_x| \leq \varphi \left| \sum_{i \neq j}^{(x,-x)} h_{xi} G_{ij}^{(x,-x)} h_{jx} \right| + \varphi \left| \sum_{i}^{(x,-x)} \Big(|h_{xi}|^2 - s_{xi} \Big) G_{ii}^{(x,-x)} \right| \prec \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}. \tag{3.5.26}$$

If $x = -x$ then Z_x can be bounded by the right-hand side in (3.5.12) similarly to the previous estimate and for A_x in exactly the same way as in (3.5.17).

To estimate the generic off-diagonal entry G_{xy} under the assumption that all of $x, -x, y, -y$ are different, we use the expansion

$$\begin{aligned} G_{xy} = & - G_{xx}^{(-x,-y)} G_{yy}^{(x,-x,-y)} \left(h_{xy} - \sum_{k,l}^{(x,-x,y,-y)} h_{xk} G_{kl}^{(x,-x,y,-y)} h_{ly} \right) \\ & + \frac{G_{x,-y}^{(-x)} G_{-y,y}^{(-x)}}{G_{-y,-y}^{(-x)}} + \frac{G_{x,-x} G_{-x,y}}{G_{-x,-x}}, \end{aligned} \tag{3.5.27}$$

which follows from applying (3.4.3) twice and afterwards applying the first identity in (3.4.2) twice. Conditioning on $G^{(x,-x,y,-y)}$ and applying (C.3) in [60] with $X_k = \zeta_{xk}$, $Y_l = \zeta_{ly}$ and $a_{kl} = s_{xk}^{1/2} G_{kl}^{(x,-x,y,-y)} s_{ly}^{1/2}$ yield

$$\varphi \left| \sum_{k,l}^{(x,-x,y,-y)} h_{xk} G_{kl}^{(x,-x,y,-y)} h_{ly} \right|^2 \prec \varphi \sum_{k,l}^{(x,-x,y,-y)} s_{xk} |G_{kl}^{(x,-x,y,-y)}|^2 s_{ly} \prec \frac{\operatorname{Im} m + \Lambda}{M\eta}, \tag{3.5.28}$$

where the last step follows exactly as in (3.5.21), which implies

$$\varphi |G_{xy}| \prec M^{-1/2} + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}} + \varphi \Lambda_o^2.$$

If $x = -x$ or $y = -y$ then the proof of the last statement is easier. This finishes the proof of (3.5.12).

Now, we turn to the proof of (3.5.13). The trivial estimate $|\mathbb{E}h_{xy}^2| \leq \mathbb{E}|h_{xy}|^2 = s_{xy} \leq M^{-1}$ implies that the first two terms in $\varphi|\mathcal{E}_x^1|$ are bounded by M^{-1}. By (3.2.7) its last term is bounded by $M^{-1/2}$. Splitting the summation in the third term of $\varphi|\mathcal{E}_x^1|$ into $a \neq -a$ and $a = -a$ and using the estimate on $|\mathbb{E}h_{xy}|^2$ we obtain $\varphi|\mathcal{E}_x^1| \prec \varphi\Lambda\Lambda_- + M^{-1/2}$ due to (3.2.2), (3.4.13), the fourth estimate in Lemma 3.5.2 and (3.2.7). Similarly to the bound on the third term in $\varphi|\mathcal{E}_x^1|$, we get $\varphi|\mathcal{E}_x^3| \prec \varphi\Lambda_o^2$ and $\varphi|\mathcal{E}_x^4| \prec \varphi\Lambda_o^2$. To estimate $\mathcal{E}_x^2$ we calculate the partial expectation in its definition which yields

$$\mathcal{E}_x^2 = G_{xx}G_{-x,-x}^{(x)} \sum_a^{(x,-x)} \left(h_{xa}^2 - r_{xa}\right) G_{a,-a}^{(x,-x)} + G_{xx}G_{-x,-x}^{(x)} \sum_{a\neq b}^{(x,-x)} h_{xa}G_{a,-b}^{(x,-x)}h_{xb}.$$

Similarly to (3.5.25) the first term can be bounded by M^{-1}. Using (3.5.21) for the second term implies

$$\varphi|\mathcal{E}_x^2| \prec \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}$$

which completes the proof of (3.5.13).

Finally, we prove part (ii) of Lemma 3.5.3. In contrast to part (i), we fix $\eta > 0$. Since constants do not matter in the estimates with respect to the stochastic domination we will not keep track of η in such estimates. We start the proof of part (ii) of Lemma 3.5.3 with verifying (3.5.16). First, we remark that applying (3.2.7), (3.4.4) and (3.4.5) yields

$$\begin{aligned} \left|\sum_a^{(\mathbb{T})} h_{xa}G_{ab}^{(\mathbb{T}')}\right| &\leq \left(\sum_a |h_{xa}|^2\right)^{1/2} \left(\sum_a |G_{ab}^{(\mathbb{T}')}|^2\right)^{1/2} \\ &\prec \left(\sum_a s_{xa}\right)^{1/2} \left(\eta^{-1}\operatorname{Im} G_{bb}^{(\mathbb{T}')}\right)^{1/2} \leq \eta^{-1} \end{aligned} \tag{3.5.29}$$

for arbitrary finite subsets $\mathbb{T}, \mathbb{T}' \subset \mathbb{N}$. The resolvent identity (3.4.3) and the previous bound imply

$$|A_x| \leq |s_{xx}G_{xx}| + \sum_a^{(x)} s_{xa}|G_{ax}| \left|\sum_b^{(x)} h_{xb}G_{ba}^{(x)}\right| \prec M^{-1} + \Lambda_o, \tag{3.5.30}$$

where we used (3.2.7) and (3.4.5) in the second step. The estimate

$$|B_x| \leq \sum_a^{(x,-x)} s_{xa} \left| \sum_k^{(x,-x)} G_{ak}^{(x,-x)} h_{k,-x} \right| |G_{-x,a}^{(x)}| \prec M^{-1/2} \tag{3.5.31}$$

is a consequence of (C.2) in [60] with $X_k = \zeta_{k,-x}$ and $a_k = s_{k,-x}^{1/2} G_{ak}^{(x,-x)}$, (3.4.4), (3.4.5) and (3.2.2).

Applying (3.5.29) to the second and third term in (3.5.3) and (3.2.7) to the first term yields $|C_x| \prec M^{-1/2}$.

To estimate Y_x we start from (3.5.22) but (3.5.20) is estimated differently. Using the resolvent identity (3.4.2) twice we get

$$\begin{aligned} \left| \sum_k^{(x,-x)} h_{xk}^2 G_{k,-k}^{(x,-x)} \right| \prec & \sum_{k \neq -k}^{(x,-x)} s_{xk} |G_{k,-k}| + \sum_{k=-k}^{(x,-x)} s_{xk} |G_{kk}| \\ & + \sum_k^{(x,-x)} s_{xk} \frac{|G_{k,-x}^{(x)} G_{-x,-k}^{(x)}|}{|G_{-x,-x}^{(x)}|} + \sum_k^{(x,-x)} s_{xk} \frac{|G_{kx} G_{x,-k}|}{|G_{xx}|} \prec \Lambda_o + M^{-1/2}, \end{aligned}$$

where the last step follows similarly to (3.5.30) and (3.5.31). Combining this with the usage of (3.4.5) instead of Lemma 3.5.2 in (3.5.21) yields $|Y_x| \prec M^{-1/2} + \Lambda_o$. We get $|Z_x| \prec M^{-1/2}$ by similar adjustments of (3.5.26). This completes the proof of (3.5.16).

Before proving (3.5.14) we show

$$\Lambda_g \leq \eta^{-1} \Lambda_- + \tilde{\epsilon} \tag{3.5.32}$$

with some $\tilde{\epsilon} \prec M^{-1/4}$ uniformly for $z \in \{w \in \mathbb{C}; \operatorname{Im} w = \eta\}$. In case all of x, $-x$, y and $-y$ are different it will be derived from the representation in (3.5.27). The first summand in (3.5.27) is bounded by $M^{-1/2}$ due to (3.2.7) and (3.4.5). Using (3.4.5) instead of Lemma 3.5.2 in (3.5.28) yields that the second term in (3.5.27) is dominated by $M^{-1/2}$ as well. For the third summand in (3.5.27) we use the estimate

$$\left| \frac{G_{x,-y}^{(-x)} G_{-y,y}^{(-x)}}{G_{-y,-y}^{(-x)}} \right| = \left| G_{xx}^{(-x)} \sum_a^{(x,-x)} h_{xa} G_{a,-y}^{(x,-x)} \right| \left| \sum_a^{(-y,-x)} h_{-y,a} G_{ay}^{(-y,-x)} \right| \prec M^{-1/2},$$

where we used (C.2) in [60] as in the proof of (3.5.31) for the first factor and (3.5.29) for the second factor. For the fourth term in (3.5.27) we obtain

$$\begin{aligned}\frac{|G_{x,-x}G_{-x,y}|}{|G_{-x,-x}|} &\leq \Lambda_- \left| \sum_a^{(-x)} h_{-x,a} G_{ay}^{(-x)} \right| \leq \Lambda_- \left(\sum_a^{(-x)} |h_{-x,a}|^2 \right)^{1/2} \left(\sum_a^{(-x)} |G_{ay}^{(-x)}|^2 \right)^{1/2} \\ &\leq \Lambda_- \eta^{-1} + \eta^{-2} \left| \sum_a^{(-x)} (|h_{-x,a}|^2 - s_{-x,a}) \right|^{1/2}\end{aligned} \tag{3.5.33}$$

by applying the resolvent identity (3.4.3) and inserting $s_{-x,a}$. In the last step, we applied (3.4.4) and (3.4.5). Note that similarly to (3.5.25) we conclude that the second term is dominated by $M^{-1/4}$.

We denote the sum of the absolute values of the first three summands in (3.5.27) and the second summand in (3.5.33) by $\tilde{\epsilon}_{xy}$ and set $\tilde{\epsilon} := \sup_{x,y} \tilde{\epsilon}_{xy}$. Then the above considerations show $\tilde{\epsilon} \prec M^{-1/4}$ in this case. If $x = -x$ or $y = -y$ then estimating G_{xy} is easier. Thus, (3.5.32) follows.

Without inserting $s_{-x,a}$ in (3.5.33) and instead using (3.2.7) we see that the representation (3.5.27) implies (3.5.15).

To prove (3.5.14) we assume $x \neq -x$ and consider the expansion

$$\begin{aligned}G_{x,-x} = G_{xx} G_{-x,-x}^{(x)} \sum_{a \neq -a}^{(x,-x)} r_{xa} G_{a,-a} + G_{xx} G_{-x,-x}^{(x)} \sum_{a=-a}^{(x,-x)} r_{xa} G_{a,-a} - G_{xx} G_{-x,-x}^{(x)} h_{x,-x} \\ + \mathcal{E}_x^2 - \mathcal{E}_x^3 - \mathcal{E}_x^4.\end{aligned}$$

Obviously, the absolute value of the first summand on the right-hand side is not bigger than $\eta^{-2}\Lambda_-$ and $|\mathcal{E}_x^3| \leq \eta^{-1}\Lambda_g^2$. We call the sum of the second and the third term on the right-hand side $\mathcal{E}_x^5$ and obtain $|\mathcal{E}_x^5| \prec M^{-1/2}$ by (3.2.7). Similarly as before, we get $|\mathcal{E}_x^2| \prec M^{-1/2}$ by using (3.4.5) instead of Lemma 3.5.2. An argument in the fashion of (3.5.31) yields $|\mathcal{E}_x^4| \prec M^{-1/2}$.

Thus, by setting $\epsilon_x := 2\eta^{-1}\tilde{\epsilon}^2 + |\mathcal{E}_x^2| + |\mathcal{E}_x^4| + |\mathcal{E}_x^5|$ and using (3.5.32) we get

$$|G_{x,-x}| \leq \eta^{-2}\Lambda_- + \eta^{-1}\Lambda_g^2 + |\mathcal{E}_x^2| + |\mathcal{E}_x^4| + |\mathcal{E}_x^5| \leq \eta^{-2}\Lambda_- + 2\eta^{-3}\Lambda_-^2 + \epsilon_x.$$

Since $\epsilon_x \prec M^{-1/2}$ uniformly in x the estimate (3.5.14) follows from the definition $\epsilon := \sup_x \epsilon_x$. □

3.5.3. Preliminary Bound on Λ. In this section, we establish a deterministic bound on Λ. The proof will make essential use of the self-consistent equations in Lemma 3.5.1.

Proposition 3.5.4. *We have* $\Lambda \prec M^{-\gamma/3}\Gamma^{-1}$ *uniformly in* $\mathbf{S}$.

Once we have proven the two subsequent lemmas the proof of Proposition 3.5.4 follows exactly as in [60].

Lemma 3.5.5. *We have the estimate* $\mathbf{1}(\Lambda \leq M^{-\gamma/4}\Gamma^{-1})\Lambda \prec M^{-\gamma/2}\Gamma^{-1}$ *uniformly in* $\mathbf{S}$.

PROOF. In this proof, we will use Lemma 3.5.3 (i) several times with $\varphi := \mathbf{1}(\Lambda \leq M^{-\gamma/4}\Gamma^{-1})$. Following the proof of Lemma 5.4 in [60] we get

$$\varphi\Lambda_d \prec \varphi\Gamma_S \left(\Lambda^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}\right)$$

since $|\Upsilon_x| \prec \varphi\Lambda^2 + \sqrt{(\operatorname{Im} m + \Lambda)/M\eta}$ by (3.5.12). Moreover, because of (3.5.12) and the first estimate in (3.4.14) we have

$$\varphi\Lambda_g \prec \varphi\Gamma_S \left(\Lambda^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}\right).$$

Using (3.5.6) we get

$$\sum_{y \neq -y} (1 - m^2 R_{xy}) G_{y,-y} = \mathcal{E}_x$$

for all $x \neq -x$. Inverting $(1 - m^2 R)$ and using (3.5.13) yield

$$\varphi\Lambda_- = \max_{x \neq -x} \varphi|G_{x,-x}| \leq \Gamma_R \max_{x \neq -x} \varphi|\mathcal{E}_x| \prec \varphi\Gamma_R \left(\Lambda^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}\right). \tag{3.5.34}$$

In total, we get

$$\varphi\Lambda \prec \varphi\Gamma \left(\Lambda^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}\right)$$

as in (5.18) of [60]. Employing the definitions of $\mathbf{S}$ and φ as in the proof of Lemma 5.4 in [60] establishes the claim. □

When estimating the off-diagonal terms $G_{x,-x}$ in (3.5.34) the control parameter Γ_R appears naturally as the operator norm of $(1 - m^2 R)^{-1}$ in the same way as Γ_S (which is called Γ in [60]) is used in [60] to bound the differences $G_{xx} - m$.

Lemma 3.5.6. *We have* $\Lambda \prec M^{-1/2}$ *uniformly in* $z \in [-10, 10] + 2\mathrm{i}$.

PROOF. We use the bounds $|G_{ij}^{(\mathbb{T})}| \leq 1/\eta = 1/2$ from (3.4.5) and $|m| \leq 1/\eta = 1/2$ from the third estimate in (3.4.13). In particular, they imply $|v_x| = |G_{xx} - m| \leq 1$ and $|m^{-1}| \geq 2$.

By (3.5.14) with $\eta = 2$ we have

$$\Lambda_- \leq \frac{8}{5}\epsilon \prec M^{-1/2}.$$

Thus, (3.5.15) implies $\Lambda_g \prec M^{-1/2}$. Hence, $\Lambda_o \prec M^{-1/2}$ and therefore $|\Upsilon_x| \prec M^{-1/2}$ by (3.5.16).

Following now the reasoning of the proof of Lemma 5.5 in [60] we get $\Lambda \prec M^{-1/2}$. □

PROOF OF PROPOSITION 3.5.4. The maximum of the two Lipschitz-continuous functions Γ_S and Γ_R is a Lipschitz-continuous function whose Lipschitz-constant is not bigger than the maximum of the original Lipschitz-constants. Therefore, Proposition 3.5.4 can be proven exactly in the same way as Proposition 5.3 in [60]. □

3.5.4. Proof of the Main Result. In the whole section let Ψ be a deterministic control parameter satisfying

$$cM^{-1/2} \leq \Psi \leq M^{-\gamma/3}\Gamma^{-1}. \tag{3.5.35}$$

The following proposition states that such deterministic bound on Λ can always be improved. This self-improving mechanism is also present in Proposition 5.6 of [60].

Proposition 3.5.7. *Let* Ψ *satisfy* (3.5.35) *and fix* $\varepsilon \in (0, \gamma/3)$. *If* $\Lambda \prec \Psi$ *then* $\Lambda \prec F(\Psi)$ *with*

$$F(\Psi) := M^{-\varepsilon}\Psi + \sqrt{\frac{\operatorname{Im} m}{M\eta}} + \frac{M^{\varepsilon}}{M\eta}.$$

PROOF. We will apply the results of Lemma 3.5.3 (i) with $\varphi = 1$. Using (3.5.12) we get

$$\Lambda_g + |\Upsilon_x| \prec \Lambda^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}} \prec \Gamma\Psi^2 + \sqrt{\frac{\operatorname{Im} m + \Psi}{M\eta}} \tag{3.5.36}$$

because of the first estimate in (3.4.14). The self-consistent equation (3.5.6) for $G_{x,-x}$ implies the estimate

$$|G_{x,-x}| \leq |m^2| \left| \sum_{a \neq -a} (\mathbb{E}h_{xa}^2) G_{a,-a} \right| + |\mathcal{E}_x| \prec \Gamma\Psi^2 + \sqrt{\frac{\operatorname{Im} m + \Psi}{M\eta}} \tag{3.5.37}$$

which holds uniformly in x. Here, we applied the fluctuation averaging (3.4.11) for $G_{x,-x}$ with $t_{xa} = \mathbb{E}h_{xa}^2$ and (3.4.13) to the first summand, $|\mathbb{E}h_{xy}^2| \leq M^{-1}$, Lemma 3.5.2 and (3.4.14) to the second summand and (3.5.13) to $|\mathcal{E}_x|$ and employed $\Gamma_R \leq \Gamma$ and (3.4.14) afterwards.

Starting with these estimates the reasoning in the proof of Proposition 5.6 in [60] yields

$$\Lambda \prec \Gamma\Psi^2 + \sqrt{\frac{\operatorname{Im} m + \Lambda}{M\eta}}.$$

The claim follows from applying Young's inequality and the condition $\Psi \leq M^{-\gamma/3}\Gamma^{-1}$ to the right-hand side of the previous estimate. □

In the following lemma we use the notation $[v]$ for the mean of a vector $v = (v_i)_i \in \mathbb{C}^N$, i.e.,

$$[v] = \frac{1}{N}\sum_i v_i.$$

Lemma 3.5.8. *If Ψ is a deterministic control parameter such that $\Lambda \prec \Psi$ then we have $[\Upsilon] \in O_\prec(\Psi^2)$.*

Proof. If $x \neq -x$ then we obtain from Schur's complement formula (3.4.1) and the definition of Υ_x

$$\Upsilon_x = A_x + B_x - s_{x,-x}\mathbb{E}_x G_{-x,-x}^{(x)} - \mathbb{E}_x Y_x + \mathbb{F}_x \frac{1}{G_{xx}}. \tag{3.5.38}$$

The fluctuation averaging (3.4.9) with $t_{ik} = 1/N$ yields $[\mathbb{F}_x G_{xx}^{-1}] \in O_\prec(\Psi^2)$. Obviously, we have $|A_x| \prec \Psi^2$ and $|B_x| \prec \Psi^2$ by Lemma 3.5.2. Lemma 3.6.1, Lemma 3.5.2 and (3.2.7) imply $|s_{x,-x}\mathbb{E}_x G_{-x,-x}^{(x)}| \prec M^{-1} \leq \Psi^2$ due to the first estimate in (3.5.35).

Using (3.5.20) and the first two steps in (3.5.21) with $\varphi = 1$ we obtain

$$\left| \sum_{k,l}^{(x,-x)} h_{xk} G_{kl}^{(x,-x)} h_{l,-x} \right| \prec \Psi. \tag{3.5.39}$$

Thus, the representation of Y_x in (3.5.22) and the application of Lemma 3.5.2 yield $|Y_x| \prec \Psi^2$. Hence, Lemma 3.6.1 implies $|\mathbb{E}_x Y_x| \prec \Psi^2$. For $x = -x$ the relation (3.5.38) without the second to fourth term on the right-hand side and $|A_x| \prec \Psi^2$ hold true and $|[\Upsilon]| \prec \Psi^2$ follows from (3.5.38). □

Proposition 3.5.4, Proposition 3.5.7 and Lemma 3.5.8 imply Theorem 3.2.3 along the same lines as Proposition 5.3, Proposition 5.6 and Lemma 5.7 in [60] complete the proof of Theorem 5.1 in [60].

3.6. Proof of the Fluctuation Averaging

In this section, we verify the fluctuation averaging, i.e., Theorem 3.4.5 and Theorem 3.4.6. To this end, we transfer the proof of the fluctuation averaging given in [60] to our setting. We only highlight the differences due to the special counterdiagonal terms $G_{x,-x}$.

We start with two preparatory lemmas. The following result is the analogue of Lemma B.1 in [60] whose proof works in the current situation as well. Recall that $\mathbb{E}_x X = \mathbb{E}[X|H^{(x,-x)}]$ is the expectation conditioned on the minor $H^{(x,-x)}$ and $\mathbb{F}_x X = X - \mathbb{E}_x X$ for an integrable random variable X (cf. Definition 3.4.1 and Definition 3.4.3).

Lemma 3.6.1. *Let Ψ be a deterministic control parameter satisfying $\Psi \geq N^{-C}$ and let $X(u)$ be nonnegative random variables such that for every $p \in \mathbb{N}$ there exists a constant c_p with $\mathbb{E}[X(u)^p] \leq N^{c_p}$ for all large N. If $X(u) \prec \Psi$ uniformly in u then*

$$\mathbb{E}_x X(u)^n \prec \Psi^n, \qquad \mathbb{F}_x X(u)^n \prec \Psi^n, \qquad \mathbb{E} X(u)^n \prec \Psi^n$$

uniformly in u and in x.

This lemma will be used throughout the following arguments. The trivial condition $\mathbb{E}[X(u)^p] \leq N^{c_p}$ will always be fulfilled. The following lemma which replaces (B.5) in [60] gives an auxiliary bound for estimating high moments of $|\sum_k t_{ik} \mathbb{F}_k G_{kk}^{-1}|$ when there are bounds on $\Lambda = \max_{x,y} |G_{xy} - \delta_{xy} m|$ and $\Lambda_o = \max_{x \neq y} |G_{xy}|$ (cf. (3.2.13)).

Lemma 3.6.2. *Let $\mathbf{D}$ be a spectral domain. Suppose $\Lambda \prec \Psi$ and $\Lambda_o \prec \Psi_o$ for some deterministic control parameters Ψ and Ψ_o which satisfy* (3.5.11)*. Then for fixed $p \in \mathbb{N}$*

we have

$$\left|\mathbb{F}_x\left(G_{xx}^{(\mathbb{T})}\right)^{-1}\right| \prec \Psi_o \tag{3.6.1}$$

uniformly in $\mathbb{T}\subset\mathbb{N}$, $|\mathbb{T}|\le p$, $x\notin\mathbb{T}\cup-\mathbb{T}$ *and* $z\in\mathbf{D}$.

Proof. If $x=-x$ then the proof of (3.6.1) is exactly the same as the proof of (B.5) in [60]. For $x\neq -x$ we start with (3.4.1). Since $x,-x\notin\mathbb{T}$ we obtain as in the proof of (3.5.7) by using the first resolvent identity (3.4.2) that

$$\begin{aligned}\sum_{a,b}^{(\mathbb{T},x)} h_{xa}G_{ab}^{(\mathbb{T},x)}h_{bx} = {}& C_x^{(\mathbb{T})} + \sum_{a,b}^{(\mathbb{T},x,-x)} h_{xa}G_{ab}^{(\mathbb{T},x,-x)}h_{bx}\\ &+\left(G_{-x,-x}^{(\mathbb{T},x)}\right)^{-1}\sum_{a,b}^{(\mathbb{T},x,-x)} h_{xa}G_{a,-x}^{(\mathbb{T},x)}G_{-x,b}^{(\mathbb{T},x)}h_{bx},\end{aligned} \tag{3.6.2}$$

where we used the definition

$$C_x^{(\mathbb{T})} := h_{x,-x}G_{-x,-x}^{(\mathbb{T},x)}h_{-x,x} + \sum_{a}^{(\mathbb{T},x,-x)} h_{xa}G_{a,-x}^{(\mathbb{T},x)}h_{-x,x} + \sum_{b}^{(\mathbb{T},x,-x)} h_{x,-x}G_{-x,b}^{(\mathbb{T},x)}h_{bx}.$$

The assumptions of Lemma 3.6.1 are fulfilled for each term of the expansion in (3.6.2) by (3.2.7) and the second estimate in (3.4.5).

Similar to the proof of (3.5.19) we get $|C_x^{(\mathbb{T})}|\prec M^{-1/2}\le\Psi_o$ by (3.5.11). Using the first step in (3.5.24) and the argument in (3.5.25) we get

$$\begin{aligned}\left|\mathbb{F}_x\sum_{a,b}^{(\mathbb{T},x,-x)} h_{xa}G_{ab}^{(\mathbb{T},x,-x)}h_{bx}\right| &\le \left|\sum_{a\neq b}^{(\mathbb{T},x,-x)} h_{xa}G_{ab}^{(\mathbb{T},x,-x)}h_{bx}\right| + \left|\sum_{a}^{(\mathbb{T},x,-x)}\left(|h_{xa}|^2-s_{xa}\right)G_{aa}^{(\mathbb{T},x,-x)}\right|\\ &\prec \Psi_o\end{aligned}$$

where we used that Ψ_o fulfills (3.5.11). The estimate

$$\left|\sum_{k,l}^{(\mathbb{T},x,-x)} h_{xk}G_{kl}^{(\mathbb{T},x,-x)}h_{l,-x}\right| \prec \Psi_o \tag{3.6.3}$$

which follows from adapting (3.5.20) and the first step in (3.5.21) implies

$$\left|\left(G_{-x,-x}^{(\mathbb{T},x)}\right)^{-1}\sum_{a,b}^{(\mathbb{T},x,-x)} h_{xa}G_{a,-x}^{(\mathbb{T},x)}G_{-x,b}^{(\mathbb{T},x)}h_{bx}\right| \prec \Psi_o^2 \prec \Psi_o \tag{3.6.4}$$

using a similar representation as in (3.5.22) and Lemma 3.5.2. By Lemma 3.6.1 these estimates imply

$$\left| \mathbb{F}_x \sum_{a,b}^{(\mathbb{T},x)} h_{xa} G_{ab}^{(\mathbb{T},x)} h_{bx} \right| \prec \Psi_o.$$

Thus, the claim is obtained by applying Schur's complement formula (3.4.1) to $G_{xx}^{(\mathbb{T})}$ and observing that $|\mathbb{F}_x(h_{xx} - z)| = |h_{xx}| \prec M^{-1/2} \leq \Psi_o$ as h_{xx} is independent of $H^{(x,-x)}$ and $\mathbb{E} h_{xx} = 0$. □

Proof of Theorem 3.4.6. The proof is similar to the proof of Theorem 4.7 on pages 48 to 53 in [60] so we only describe the changes needed to transfer this proof to its version for the fourfold symmetry.

First, we use Lemma 3.6.2 instead of (B.5). Moreover, we have to change some notions introduced in the proof of Theorem 4.7. In the middle of page 49, an equivalence relation on the set $\{1, \ldots, p\}$ is introduced which has to be substituted by the following equivalence relation. Starting with $\mathbf{k} := (k_1, \ldots, k_p) \in (\mathbb{Z}/N\mathbb{Z})^p$ and $r, s \in \{1, \ldots, p\}$ we define $r \sim s$ if and only if $k_r = k_s$ or $k_r = -k_s$. As in [60] the summation over all $\mathbf{k}$ is regrouped with respect to this equivalence relation and the notion of "lone" labels has to be understood with respect to this equivalence relation. We use the same notation $\mathbf{k}_L$ for the set of summation indices corresponding to lone labels. Differing from the definition in [60] we call a resolvent entry $G_{xy}^{(\mathbb{T})}$ with $x, y \notin \mathbb{T}$ *maximally expanded* if $\mathbf{k}_L \cup -\mathbf{k}_L \subset \mathbb{T} \cup \{x, y\}$. Correspondingly, we denote by $\mathcal{A}$ the set of monomials in the off-diagonal entries $G_{xy}^{(\mathbb{T})}$ with $\mathbb{T} \subset \mathbf{k}_L \cup -\mathbf{k}_L$, $x \neq y$ and $x, y \in \mathbf{k} \backslash \mathbb{T}$ (considering $\mathbf{k}$ as a subset of $\mathbb{Z}/N\mathbb{Z}$) and the inverses of diagonal entries $1/G_{xx}^{(\mathbb{T})}$ with $\mathbb{T} \subset \mathbf{k}_L \cup -\mathbf{k}_L$ and $x \in \mathbf{k} \backslash \mathbb{T}$. With these alterations the algorithm can be applied as in [60]. In the proof of (B.15) the assertion $(*)$ has to be replaced by

$(*)$ For each $s \in L$ there exists $r = \tau(s) \in \{1, \ldots, p\} \backslash \{s\}$ such that the monomial $A_{\sigma_r}^r$ contains a resolvent entry with lower index k_s or $-k_s$.

To prove this claim, we suppose by contradiction that there is $s \in L$ such that $A_{\sigma_r}^r$ does not contain k_s and $-k_s$ as lower index for all $r \in \{1, \ldots, p\} \backslash \{s\}$. Without loss of generality we assume $s = 1$. This implies that each resolvent entry in $A_{\sigma_r}^r$ contains k_1 and $-k_1$ as upper index since $A_{\sigma_r}^r$ is maximally expanded for all $r \in \{2, \ldots, p\}$. Therefore,

$A^r_{\sigma_r}$ is independent of k_1 as defined in Definition 3.4.4. Using (3.4.6) and proceeding as in [60] concludes the proof of $(*)$.

Following verbatim the remaining steps in the proof of Theorem 4.7 in [60] establishes the assertion of Theorem 3.4.6. □

Now, we deduce Theorem 3.4.5 from Theorem 3.4.6.

PROOF OF THEOREM 3.4.5. The first estimate in (3.4.9) follows from Theorem 3.4.6 directly by setting $\Psi_o := \Psi$ and using $\Lambda_o \leq \Lambda \prec \Psi_o$.

To verify the second estimate in (3.4.9) we use the fourth estimate in Lemma 3.5.2 which implies

$$|\mathbb{F}_x G^{(\mathbb{T})}_{xx}| = |\mathbb{F}_x \left(G^{(\mathbb{T})}_{xx} - m\right)| \prec \Psi. \tag{3.6.5}$$

Now, following the proof of Theorem 3.4.6 verbatim with $\Psi_o := \Psi$ and replacing the usage of Lemma 3.6.2 by (3.6.5) yield the second estimate in (3.4.9).

Similarly, the third estimate in (3.4.9) is proven by following the proof of Theorem 3.4.6 verbatim with $\Psi_o := \Psi$ and Lemma 3.6.2 replaced by

$$|\mathbb{F}_x G^{(\mathbb{T})}_{x,-x}| \prec \Lambda_o \prec \Psi$$

for $x \neq -x$ which is a consequence of Lemma 3.5.2 and Lemma 3.6.1.

Next, we establish (3.4.10). We start from Schur's complement formula (3.4.1) with $\mathbb{T} = \emptyset$ and use (3.2.6) to get

$$\frac{1}{G_{xx}} = \frac{1}{m} + h_{xx} - \left(\sum_{k,l}^{(x)} h_{xk} G^{(x)}_{kl} h_{lx} - m\right). \tag{3.6.6}$$

Using Lemma 3.5.2 with $\varphi = 1$ and the first estimate in (3.4.13) we get

$$\left|\frac{1}{G_{xx}} - \frac{1}{m}\right| = \left|\frac{G_{xx} - m}{G_{xx} m}\right| \prec |G_{xx} - m| \prec \Psi.$$

Thus, $|h_{xx} - \left(\sum_{k,l}^{(x)} h_{xk} G^{(x)}_{kl} h_{lx} - m\right)| \prec \Psi$ as well. Therefore, we can expand the inverse of the right-hand side of (3.6.6) around $1/m$ which yields

$$v_x := G_{xx} - m = m^2 \left(-h_{xx} + \sum_{k,l}^{(x)} h_{xk} G^{(x)}_{kl} h_{lx} - m\right) + g_x \tag{3.6.7}$$

with error terms g_x such that $|g_x| \prec \Psi^2$ uniformly in x. By (3.5.7), (3.5.8), (3.5.3), (3.5.4) and (3.5.5), we have for $x \neq -x$ the representation

$$\sum_{k,l}^{(x)} h_{xk} G_{kl}^{(x)} h_{lx} = \sum_a s_{xa} G_{aa} - A_x - B_x - s_{-x,x} G_{-x,-x}^{(x)} + Z_x + Y_x + C_x + s_{-x,x} G_{-x,-x}^{(x)}. \quad (3.6.8)$$

Taking the expectation $\mathbb{E}_x$ of (3.6.7) we want to prove that

$$\mathbb{E}_x v_x = m^2 \sum_a s_{xa} v_a + f_x, \quad (3.6.9)$$

where $|f_x| \prec \Psi^2$ uniformly in x. From (3.5.8) we get that the sum of the first four summands on the right-hand side of (3.6.8) is $H^{(x,-x)}$-measurable. Therefore, it suffices to show that all summands except the first one on the right-hand side of (3.6.8) are bounded by Ψ^2 uniformly in x. For A_x and B_x this follows directly from their definitions in (3.5.2). Since $Z_x = \mathbb{F}_x X_x$ for some random variable X_x we get $\mathbb{E}_x Z_x = 0$. The representation (3.5.18) for C_x and Lemma 3.5.2 yield $|C_x| \prec M^{-1} + M^{-1/2}\Psi \prec \Psi^2$ by (3.5.11). The bound (3.6.4) with $\mathbb{T} = \emptyset$ gives $|Y_x| \prec \Psi^2$ uniformly in x. If $x = -x$ then the argumentation in [60] can be applied. This finishes the proof of (3.6.9).

Therefore, since $\mathbb{E}_x + \mathbb{F}_x = 1$ we have

$$\begin{aligned} w_a := \sum_x t_{ax} v_x &= \sum_x t_{ax} \mathbb{E}_x v_x + \sum_x t_{ax} \mathbb{F}_x v_x = m^2 \sum_{x,y} t_{ax} s_{xy} v_y + F_a \\ &= m^2 \sum_{x,y} s_{ax} t_{xy} v_y + F_a = m^2 \sum_x s_{ax} w_x + F_a, \end{aligned} \quad (3.6.10)$$

where we used (3.6.9) with the notation $F_a := \sum_x t_{ax}(f_x + \mathbb{F}_x v_x)$ in the third step and in the fourth step that T and S commute. Note that $|F_a| \prec \Psi^2$ uniformly in a as $|\sum_x t_{ax} \mathbb{F}_x v_x| = |\sum_x t_{ax} \mathbb{F}_x G_{xx}| \prec \Psi^2$ by the second estimate in (3.4.9). Introducing the vectors $\mathbf{w} := (w_a)_{a \in \mathbb{Z}/N\mathbb{Z}}$ and $\mathbf{F} := (F_a)_{a \in \mathbb{Z}/N\mathbb{Z}}$ and writing (3.6.10) in matrix form we get

$$\mathbf{w} = m^2 S \mathbf{w} + \mathbf{F}.$$

Inverting the last equation yields

$$\mathbf{w} = (1 - m^2 S)^{-1} \mathbf{F}.$$

Recalling the definition (3.2.8) we have

$$\|\mathbf{w}\|_\infty \le \Gamma_S \|\mathbf{F}\|_\infty \prec \Gamma_S \Psi^2$$

since $|F_a| \prec \Psi^2$ uniformly in a is equivalent to $\|\mathbf{F}\|_\infty \prec \Psi^2$. This proves (3.4.10).

In order to prove (3.4.11) it suffices to verify

$$\mathbb{E}_x G_{x,-x} = m^2 \sum_{a \neq -a} (\mathbb{E} h_{xa}^2) G_{a,-a} + f_x \tag{3.6.11}$$

with $|f_x| \prec \Psi^2$ uniformly in x. Then (3.4.11) follows from the same reasoning as in the proof of (3.4.10) with S replaced by R and

$$w_x := \sum_{a \neq -a} t_{xa} G_{a,-a}.$$

To compute the partial expectation $\mathbb{E}_x G_{x,-x}$ we use the expansion

$$\begin{aligned} G_{x,-x} = \; & m^2 \sum_a^{(x,-x)} (\mathbb{E} h_{xa}^2) G_{a,-a}^{(x,-x)} + m^2 \sum_{a \neq b}^{(x,-x)} h_{xa} G_{a,-b}^{(x,-x)} h_{xb} \\ & + m^2 \sum_a^{(x,-x)} \left(h_{xa}^2 - \mathbb{E} h_{xa}^2 \right) G_{a,-a}^{(x,-x)} + (m^2 - G_{xx} G_{-x,-x}^{(x)}) h_{x,-x} - m^2 h_{x,-x} \\ & + (G_{xx} G_{-x,-x}^{(x)} - m^2) \sum_{a \neq b}^{(x,-x)} h_{xa} G_{a,-b}^{(x,-x)} h_{xb} \\ & + (G_{xx} G_{-x,-x}^{(x)} - m^2) \sum_a^{(x,-x)} h_{xa}^2 G_{a,-a}^{(x,-x)}, \end{aligned} \tag{3.6.12}$$

which follows from the resolvent identities in a similar way as (3.5.6).

The first summand in (3.6.12) is $H^{(x,-x)}$-measurable. Using (3.4.2) twice and adding the two missing terms we obtain the first summand on the right-hand side of (3.6.11). The error terms originating from the usage of the resolvent identities and the added terms are obviously dominated by Ψ^2. The partial expectations with respect to $H^{(x,-x)}$ of the second and the fifth term vanish. For the remaining terms we use Lemma 3.6.1. First, $|m^2 - G_{xx} G_{-x,-x}^{(x)}| \prec \Psi$ because of the triangle inequality, Lemma 3.5.2 and the second estimate in (3.4.13). Thus, using (3.2.7) and (3.4.7) for the fourth term, the first step in (3.5.21) for the sixth term and (3.5.20) for the seventh term we get that these summands are dominated by Ψ^2. Similarly to (3.5.25) we see that the third summand is dominated by

Ψ^2 using the Large Deviation Bound (C.2) in [60] and the first estimate in Lemma 3.5.2. Lemma 3.6.1 establishes (3.6.11) which finishes the proof of Theorem 3.4.5. □

CHAPTER 4

Local law for random Gram matrices

This chapter consists of a modified version of the publication [14] which was written jointly with László Erdős and Torben Krüger. We prove a local law in the bulk of the spectrum for random Gram matrices XX^*, a generalization of sample covariance matrices, where X is a large matrix with independent, centered entries with arbitrary variances. The limiting eigenvalue density that generalizes the Marchenko-Pastur law is determined by solving a system of nonlinear equations. Our entrywise and averaged local laws are on the optimal scale with the optimal error bounds. They hold both in the square case (hard edge) and in the properly rectangular case (soft edge). In the latter case we also establish a macroscopic gap away from zero in the spectrum of XX^*.

4.1. Introduction

Random matrices were introduced in pioneering works by Wishart [160] and Wigner [157] for applications in mathematical statistics and nuclear physics, respectively. Wigner argued that the energy level statistics of large atomic nuclei could be described by the eigenvalues of a large *Wigner matrix*, i.e., a hermitian matrix $H = (h_{ij})_{i,j=1}^N$ with centered, identically distributed and independent entries (up to the symmetry constraint $H = H^*$). He proved that the empirical spectral measure (or density of states) converges to the *semicircle law* as the dimension of the matrix N goes to infinity. Moreover, he postulated that the statistics of the gaps between consecutive eigenvalues depend only on the symmetry type of the matrix and are independent of the distribution of the entries in the large N limit. The precise formulation of this phenomenon is called the *Wigner-Dyson-Mehta universality conjecture*, see [114].

Historically, the second main class of random matrices is the one of *sample covariance matrices*. These are of the form XX^* where X is a $p \times n$ matrix with centered, identically distributed independent entries. In statistics context, its columns contain n samples of a

p-dimensional data vector. In the regime of high dimensional data, i.e., in the limit when $n, p \to \infty$ in such a way that the ratio p/n converges to a constant, the empirical spectral measure of XX^* was explicitly identified by Marchenko and Pastur [112]. Random matrices of the form XX^* also appear in the theory of wireless communication; the spectral density of these matrices is used to compute the transmission capacity of a Multiple Input Multiple Output (MIMO) channel. This fundamental connection between random matrix theory and wireless communication was established by Telatar [147] and Foschini [76, 77] (see also [150] for a review). In this model, the element x_{ij} of the *channel matrix* X represents the transmission coefficient from the j-th transmitter to the i-th receiver antenna. The received signal is given by the linear relation $y = Xs + w$, where s is the input signal and w is a Gaussian noise with variance σ^2. In case of i.i.d. Gaussian input signals, the channel capacity is given by

$$\mathrm{Cap} = \frac{1}{p} \log\det\Big(I + \sigma^{-2} XX^*\Big). \tag{4.1.1}$$

The assumption in these models that the matrix elements of H or X have identical distribution is a simplification that does not hold in many applications. In Wigner's model, the matrix elements h_{ij} represent random quantum transition rates between physical states labelled by i and j and their distribution may depend on these states. Analogously, the transmission coefficients in X may have different distributions. This leads to the natural generalizations of both classes of random matrices by allowing for general variances, $s_{ij} := \mathbb{E}|h_{ij}|^2$ and $s_{ij} := \mathbb{E}|x_{ij}|^2$, respectively. We will still assume the independence of the matrix elements and their zero expectation. Under mild conditions on the variance matrix $S = (s_{ij})$, the limiting spectral measure depends only on the second moments, i.e., on S, and otherwise it is independent of the fine details of the distributions of the matrix elements. However, in general there is no explicit formula for the limiting spectral measure. In fact, the only known way to find it in the general case is to solve a system of nonlinear deterministic equations, known as the Dyson (or Schwinger-Dyson) equation in this context, see [34, 82, 99, 156].

For the generalization of Wigner's model, the Dyson equation is a system of equations of the form

$$-\frac{1}{m_i(z)} = z + \sum_{j=1}^{N} s_{ij} m_j(z), \qquad \text{for } i = 1, \ldots, N, \qquad z \in \mathbb{H}, \tag{4.1.2}$$

where z is a complex parameter in the upper half plane $\mathbb{H} := \{z \in \mathbb{C} \ : \ \operatorname{Im} z > 0\}$. The average $\langle m(z)\rangle = \frac{1}{N}\sum_i m_i(z)$ in the large N limit gives the Stieltjes transform of the limiting spectral density, which then can be computed by inverting the Stieltjes transform. In fact, $m_i(z)$ approximates individual diagonal matrix elements $G_{ii}(z)$ of the resolvent $G(z) = (H - z)^{-1}$, thus the solution of (4.1.2) gives much more information on H than merely the spectral density. In the case when S is a stochastic matrix, i.e., $\sum_j s_{ij} = 1$ for every i, the solution $m_i(z)$ to (4.1.2) is independent of i and the density is still the semicircle law. The corresponding generalized Wigner matrix was introduced in [70] and the optimal local law was proven in [71, 72]. For the general case, a detailed analysis of (4.1.2) and the shapes of the possible density profiles was given in [4, 5] with the optimal local law in [7].

Considering the XX^* model with a general variance matrix for X, we note that in statistical applications the entries of X within the same row still have the same variance, i.e., $s_{ik} = s_{il}$ for all i and all k, l. However, beyond statistics, for example modeling the capacity of MIMO channels, applications require to analyze the spectrum of XX^* with a completely general variance profile for X [52, 92]. These are called *random Gram matrices*, see e.g. [82, 90]. The corresponding Dyson equation is (see [52, 82, 150] and references therein)

$$-\frac{1}{m_i(\zeta)} = \zeta - \sum_{k=1}^{n} s_{ik} \frac{1}{1 + \sum_{j=1}^{p} s_{jk} m_j(\zeta)}, \qquad \text{for } i = 1, \ldots, p, \qquad \zeta \in \mathbb{H}. \tag{4.1.3}$$

We have $m_i(\zeta) \approx (XX^* - \zeta)^{-1}_{ii}$ and the average of $m_i(\zeta)$ yields the Stieltjes transform of the spectral density exactly as in case of the Wigner-type ensembles. In fact, there is a direct link between these two models: Girko's symmetrization trick reduces (4.1.3) to studying (4.1.2) on $\mathbb{C}^N$ with $N = n + p$, where S and H are replaced by

$$\boldsymbol{S} = \begin{pmatrix} 0 & S \\ S^t & 0 \end{pmatrix}, \quad \boldsymbol{H} = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}, \tag{4.1.4}$$

respectively, and $z^2 = \zeta$.

The limiting spectral density, also called the *global law*, is typically the first question one asks about random matrix ensembles. It can be strengthened by considering its *local* versions. In most cases, it is expected that the deterministic density computed via the Dyson equation accurately describes the eigenvalue density down to the smallest possible scale which is slightly above the typical eigenvalue spacing (we choose the standard normalization such that the spacing in the bulk spectrum is of order $1/N$). This requires to understand the trace of the resolvent $G(z)$ at a spectral parameter very close to the real axis, down to the scales $\operatorname{Im} z \gg 1/N$. Additionally, *entry-wise local laws* and *isotropic local laws*, i.e., controlling individual matrix elements $G_{ij}(z)$ and bilinear forms $\langle v, G(z)w\rangle$, carry important information on eigenvectors and allow for perturbation theory. Moreover, effective error bounds on the speed of convergence as N goes to infinity are also of great interest.

Local laws have also played a crucial role in the recent proofs of the Wigner-Dyson-Mehta conjecture. The three-step approach, developed in a series of works by Erdős, Schlein, Yau and Yin [64, 65] (see [69] for a review), was based on establishing the local law as the first step. Similar input was necessary in the alternative approach by Tao and Vu in [141, 144].

In this paper, we establish the optimal local law for random Gram matrices with a general variance matrix S in the bulk spectrum; edge analysis and local spectral universality is deferred to a forthcoming work. We show that the empirical spectral measure of XX^* can be approximated by a deterministic measure ν on $\mathbb{R}$ with a continuous density away from zero and possibly a point mass at zero. The convergence holds locally down to the smallest possible scale and with an optimal speed of order $1/N$. In the special case when X is a square matrix, $n = p$, the measure ν does not have a point mass but the density has an inverse square-root singularity at zero (called the *hard edge case*). In the *soft edge case*, $n \neq p$, the continuous part of ν is supported away from zero and it has a point mass of size $1 - n/p$ at zero if $p > n$. All these features are well-known for the classical Marchenko-Pastur setup, but in the general case we need to demonstrate them without any explicit formula.

We now summarize some previous related results on Gram matrices. If each entry of X has the same variance, local Marchenko-Pastur laws have first been proven in [65, 122] for the soft edge case; and in [44, 46] for the hard edge case. The isotropic local law was given in [36]. Relaxing the assumption of identical variances to a doubly stochastic variance matrix of X, the optimal local Marchenko-Pastur law has been established in [3] for the hard edge case. Sample correlation matrices in the soft edge case were considered in [28].

Motivated by the linear model in multivariate statistics and to depart from the identical distribution, random matrices of the form TZZ^*T^* have been extensively studied where T is a deterministic matrix and the entries of Z are independent, centered and have unit variance. If T is diagonal, then they are generalizations of sample covariance matrices as $TZZ^*T^* = XX^*$ and the elements of $X = TZ$ are also independent. With this definition, all entries within one row of X have the same variance since $s_{ij} = \mathbb{E}|x_{ij}|^2 = (TT^*)_{ii}$, i.e., it is a special case of our random Gram matrix. In this case the Dyson system of equations (4.1.3) can be reduced to a single equation for the average $\langle m(z)\rangle$, i.e., the limiting density can still be obtained from a *scalar self-consistent equation*. This is even true for matrices of the form XX^* with $X = TZ\widetilde{T}$, where both T and $\widetilde{T}$ are deterministic, investigated for example in [53]. For general T the elements of $X = TZ$ are not independent, so general sample covariance matrices are typically not Gram matrices. The global law for TZZ^*T^* has been proven by Silverstein and Bai in [134]. Knowles and Yin showed optimal local laws for a general deterministic T in [101].

Finally, we review some existing results on random Gram matrices with general variance S, when (4.1.3) cannot be reduced to a simpler scalar equation. The global law, even with nonzero expectation of X, has been determined by Girko [82] via (4.1.3) who also established the existence and uniqueness of the solution to (4.1.3). More recently, motivated by the theory of wireless communication, Hachem, Loubaton and Najim initiated a rigorous study of the asympotic behaviour of the channel capacity (4.1.1) with a general variance matrix S [88, 92], This required to establish the global law under more general conditions than Girko; see also [90] for a review from the point of view of applications. Hachem *et al.* have also established Gaussian fluctuations of the channel capacity (4.1.1) around a deterministic limit in [91] for the centered case. For a nonzero expectation of X, a similar result was obtained in [89], where S was restricted to a product form. Very recently in [33], a special k-fold clustered matrix XX^* was considered, where the samples came from k different clusters with possibly different distributions. The Dyson equation in this case reduces to a system of k equations. In an information-plus-noise model of the form $(R+X)(R+X)^*$, the effect of adding a noise matrix to X with identically distributed entries was studied knowing the limiting density of RR^* [55].

In all previous works concerning general Gram matrices, the spectral parameter z was fixed, in particular $\operatorname{Im} z$ had a positive lower bound independent of the dimension of the matrix. Technically, this positive imaginary part provided the necessary contraction factor in the fixed point argument that led to the existence, uniqueness and stability of the solution to the Dyson equation, (4.1.3). For local laws down to the optimal scales $\operatorname{Im} z \gg 1/N$, the regularizing effect of $\operatorname{Im} z$ is too weak. In the bulk spectrum $\operatorname{Im} z$ is effectively replaced with the local density, i.e., with the average imaginary part $\operatorname{Im} \langle m(z) \rangle$. The main difficulty with this heuristics is its apparent circularity: the yet unknown solution itself is necessary for regularizing the equation. This problem is present in all existing proofs of any local law. This circularity is broken by separating the analysis into three parts. First, we analyze the behavior of the solution $m(z)$ as $\operatorname{Im} z \to 0$. Second, we show that the solution is stable under small perturbations of the equation and the stability is provided by $\operatorname{Im} \langle m(E+i0) \rangle$ for any energy E in the bulk spectrum. Finally, we show that the diagonal elements of the resolvent of the random matrix satisfy a perturbed version of (4.1.3), where the perturbation is controlled by large deviation estimates. Stability then provides the local law.

While this program could be completed directly for the Gram matrix and its Dyson equation, (4.1.3), the argument appears much shorter if we used Girko's linearization (4.1.4) to reduce the problem to a Wigner-type matrix and use the comprehensive analysis of (4.1.2) from [4, 5] and the local law from [7]. There are two major obstacles to this naive approach.

First, the results of [4, 5] are not applicable as $\boldsymbol{S}$ does not satisfy the uniform primitivity assumption imposed in these papers (recall that a matrix A is primitive if there is a positive integer L such that all entries of A^L are strictly positive). This property is crucial for many proofs in [4, 5] but $\boldsymbol{S}$ in (4.1.4) is a typical example of a nonprimitive matrix. It is not a mere technical subtlety, in fact in the current paper, the stability estimates of (4.1.2) require a completely different treatment, culminating in the key technical bound, the Rotation-Inversion lemma (see Lemma 4.3.6 later).

Second, Girko's transformation is singular around $z \approx 0$ since it involves a $z^2 = \zeta$ change in the spectral parameter. This accounts for the singular behavior near zero in the limiting density for Gram matrices, while the corresponding Wigner-type matrix has no singularity at zero. Thus, we need to perform a more accurate analysis near zero. If $p \neq n$, the soft edge case, we derive and analyze two new equations for the first coefficients in the expansion of m around zero. Indeed, the solutions to these new equations describe the point mass at zero and provide information about the gap above zero in the support of the approximating measure. In the hard edge case, $n = p$, an additional symmetry allows us to exclude a point mass at zero.

Acknowledgement. The authors thank Zhigang Bao for helpful discussions.

Notation. For vectors $v, w \in \mathbb{C}^l$, the operations product and absolute value are defined componentwise, i.e., $vw = (v_i w_i)_{i=1}^l$ and $|v| = (|v_i|)_{i=1}^l$. Moreover, for $w \in (\mathbb{C} \setminus \{0\})^l$, we set $1/w := (1/w_i)_{i=1}^l$. For vectors $v, w \in \mathbb{C}^l$, we define $\langle w \rangle = l^{-1} \sum_{i=1}^l w_i$, $\langle v\,, w \rangle = l^{-1} \sum_{i=1}^l \overline{v_i} w_i$, $\|w\|_2^2 = l^{-1} \sum_{i=1}^l |w_i|^2$ and $\|w\|_\infty = \max_{i=1,\ldots,l} |w_i|$, $\|v\|_1 := \langle |v| \rangle$. Note that $\langle w \rangle = \langle 1\,, w \rangle$ where we used the convention that 1 also denotes the vector $(1, \ldots, 1) \in \mathbb{C}^l$. For a matrix $A \in \mathbb{C}^{l \times l}$, we use the short notations $\|A\|_\infty := \|A\|_{\infty \to \infty}$ and $\|A\|_2 := \|A\|_{2 \to 2}$ if the domain and the target are equipped with the same norm whereas we use $\|A\|_{2 \to \infty}$ to denote the matrix norm of A when it is understood as a map $(\mathbb{C}^l, \|\cdot\|_2) \to (\mathbb{C}^l, \|\cdot\|_\infty)$.

4.2. Main results

Let $X = (x_{ik})_{i,k}$ be a $p \times n$ matrix with independent, centered entries and variance matrix $S = (s_{ik})_{i,k}$, i.e.,

$$\mathbb{E}x_{ik} = 0, \quad s_{ik} := \mathbb{E}|x_{ik}|^2$$

for $i = 1, \ldots, p$ and $k = 1, \ldots, n$.

Assumptions 4.2.1. (A) The variance matrix S is *flat*, i.e., there is $s_* > 0$ such that

$$s_{ik} \leq \frac{s_*}{p+n}$$

for all $i = 1, \ldots, p$ and $k = 1, \ldots, n$.

(B) There are $L_1, L_2 \in \mathbb{N}$ and $\psi_1, \psi_2 > 0$ such that

$$[(SS^t)^{L_1}]_{ij} \geq \frac{\psi_1}{p+n}, \quad [(S^tS)^{L_2}]_{kl} \geq \frac{\psi_2}{p+n}$$

for all $i, j = 1, \ldots, p$ and $k, l = 1, \ldots, n$.

(C) All entries of X have bounded moments in the sense that there are $\mu_m > 0$ for $m \in \mathbb{N}$ such that

$$\mathbb{E}|x_{ik}|^m \leq \mu_m s_{ik}^{m/2}$$

for all $i = 1, \ldots, p$ and $k = 1, \ldots, n$.

(D) The dimensions of X are comparable with each other, i.e., there are constants $r_1, r_2 > 0$ such that

$$r_1 \leq \frac{p}{n} \leq r_2.$$

In the following, we will assume that s_*, L_1, L_2, ψ_1, ψ_2, r_1, r_2 and the sequence $(\mu_m)_m$ are fixed constants which we will call, together with some constants introduced later, *model parameters*. The constants in all our estimates will depend on the model parameters without further notice. We will use the notation $f \lesssim g$ if there is a constant $c > 0$ that depends on the model parameter only such that $f \leq cg$ and their counterparts $f \gtrsim g$ if $g \lesssim f$ and $f \sim g$ if $f \lesssim g$ and $f \gtrsim g$. The model parameters will be kept fixed whereas the parameters p and n are large numbers which will eventually be sent to infinity.

We start with a theorem about the deterministic density.

Theorem 4.2.2. *(i) If (A) holds true, then there is a unique holomorphic function $m\colon \mathbb{H} \to \mathbb{C}^p$ satisfying*

$$-\frac{1}{m(\zeta)} = \zeta - S\frac{1}{1+S^t m(\zeta)} \tag{4.2.1}$$

for all $\zeta \in \mathbb{H}$ such that $\operatorname{Im} m(\zeta) > 0$ for all $\zeta \in \mathbb{H}$. Moreover, there is a probability measure ν on $\mathbb{R}$ whose support is contained in $[0, 4s_*]$ such that

$$\langle m(\zeta)\rangle = \int_{\mathbb{R}} \frac{1}{\omega-\zeta}\nu(\mathrm{d}\omega)$$

for all $\zeta \in \mathbb{H}$.

(ii) Assume (A), (B) and (D). The measure ν is absolutely continuous wrt. the Lebesgue measure apart from a possible point mass at zero, i.e., there are a number $\pi_* \in [0,1]$ and a locally Hölder-continuous function $\pi\colon (0,\infty) \to [0,\infty)$ such that $\nu(\mathrm{d}\omega) = \pi_*\delta_0(\mathrm{d}\omega) + \pi(\omega)\mathbf{1}(\omega > 0)\mathrm{d}\omega$.

Part (i) of this theorem has already been proven in [92] and we will see that it also follows directly from [4, 5]. We included this part only for completeness. Part (ii) is a new result.

For $\zeta \in \mathbb{C} \setminus \mathbb{R}$, we denote the resolvent of XX^* at ζ by

$$R(\zeta) := (XX^* - \zeta)^{-1}$$

and its entries by $R_{ij}(\zeta)$ for $i, j = 1, \ldots, p$.

We state our main result, the local law, i.e., optimal estimates on the resolvent R, both in entrywise and in averaged form. In both cases, we provide different estimates when the real part of the spectral parameter ζ is in the bulk and when it is away from the spectrum. As there may be many zero eigenvalues, hence, a point mass at zero in the density ν, our analysis for spectral parameters ζ in the vicinity of zero requires a special treatment. We thus first prove the local law under the general assumptions (A) – (D) for ζ away from zero. Some additional assumptions in the following subsections will allow us to extend our arguments to all ζ.

All of our results are uniform in the spectral parameter ζ which is contained in some spectral domain

$$\mathbb{D}_\delta := \{\zeta \in \mathbb{H} \colon \delta \leq |\zeta| \leq 10 s_*\} \tag{4.2.2}$$

for some $\delta \geq 0$. In the first result, we assume $\delta > 0$. In the next section, under additional assumptions on S, we will work on the bigger spectral domain $\mathbb{D}_0$ that also includes a neighbourhood of zero.

Theorem 4.2.3 (Local Law for Gram matrices)**.** *Let $\delta, \varepsilon_* > 0$ and $\gamma \in (0,1)$. If X is a random matrix satisfying (A) – (D) then for every $\varepsilon > 0$ and $D > 0$ there is a constant $C_{\varepsilon,D} > 0$ such that*

$$\mathbb{P}\Big(\exists \zeta \in \mathbb{D}_\delta, i, j \in \{1, \ldots, p\} \colon \operatorname{Im}\zeta \geq p^{-1+\gamma},\ \pi(\operatorname{Re}\zeta) \geq \varepsilon_*,\\ |R_{ij}(\zeta) - m_i(\zeta)\delta_{ij}| \geq \frac{p^\varepsilon}{\sqrt{p \operatorname{Im}\zeta}}\Big) \leq \frac{C_{\varepsilon,D}}{p^D}, \tag{4.2.3a}$$

$$\mathbb{P}\Big(\exists \zeta \in \mathbb{D}_\delta, i, j \in \{1, \ldots, p\} \colon \operatorname{dist}(\zeta, \operatorname{supp}\nu) \geq \varepsilon_*,\\ |R_{ij}(\zeta) - m_i(\zeta)\delta_{ij}| \geq \frac{p^\varepsilon}{\sqrt{p}}\Big) \leq \frac{C_{\varepsilon,D}}{p^D}, \tag{4.2.3b}$$

for all $p \in \mathbb{N}$. Furthermore, for any sequences of deterministic vectors $w \in \mathbb{C}^p$ satisfying $\|w\|_\infty \leq 1$, we have

$$\mathbb{P}\Big(\exists \zeta \in \mathbb{D}_\delta \colon \operatorname{Im}\zeta \geq p^{-1+\gamma},\ \pi(\operatorname{Re}\zeta) \geq \varepsilon_*,\\ \Big|\frac{1}{p}\sum_{i=1}^p w_i \left[R_{ii}(\zeta) - m_i(\zeta)\right]\Big| \geq \frac{p^\varepsilon}{p \operatorname{Im}\zeta}\Big) \leq \frac{C_{\varepsilon,D}}{p^D}, \tag{4.2.4a}$$

$$\mathbb{P}\Big(\exists \zeta \in \mathbb{D}_\delta \colon \operatorname{dist}(\zeta, \operatorname{supp}\nu) \geq \varepsilon_*,\\ \Big|\frac{1}{p}\sum_{i=1}^p w_i \left[R_{ii}(\zeta) - m_i(\zeta)\right]\Big| \geq \frac{p^\varepsilon}{p}\Big) \leq \frac{C_{\varepsilon,D}}{p^D}, \tag{4.2.4b}$$

for all $p \in \mathbb{N}$. In particular, choosing $w_i = 1$ for all $i = 1, \ldots, p$ in (4.2.4) *yields that $p^{-1}\operatorname{Tr} R(\zeta)$ is close to $\langle m(\zeta)\rangle$.*

The constant $C_{\varepsilon,D}$ depends, in addition to ε and D, only on the model parameters and on γ, δ and ε_*.

These results are optimal up to the arbitrarily small tolerance exponents $\gamma > 0$ and $\varepsilon > 0$. We remark that under stronger (e.g. subexponential) moment conditions in (C), one may replace the p^γ and p^ε factors with high powers of $\log p$.

Owing to the symmetry of the assumptions (A) – (D) in X and X^*, we can exchange X and X^* in Theorem 4.2.3 and obtain a statement about X^*X instead of XX^* as well.

For the results in the up-coming subsections, we need the following notion of a sequence of high probability events.

Definition 4.2.4 (Overwhelming probability)**.** Let $N_0\colon (0,\infty) \to \mathbb{N}$ be a function that depends on the model parameters and the tolerance exponent γ only. For a sequence $A = (A^{(p)})_p$ of random events, we say that A holds true *asymptotically with overwhelming probability* (a.w.o.p.) if for all $D > 0$

$$\mathbb{P}(A^{(p)}) \geq 1 - p^D$$

for all $p \geq N_0(D)$.

We denote the eigenvalues of XX^* by $\lambda_1 \leq \ldots \leq \lambda_p$ and define

$$i(\chi) := \left\lceil p \int_{-\infty}^{\chi} \nu(\mathrm{d}\omega) \right\rceil, \qquad \text{for } \chi \in \mathbb{R}. \tag{4.2.5}$$

For a spectral parameter $\chi \in \mathbb{R}$ in the bulk, the nonnegative integer $i(\chi)$ is the index of an eigenvalue expected to be close to χ.

Theorem 4.2.5. *Let $\delta, \varepsilon_* > 0$ and X be a random matrix satisfying (A) – (D).*

(i) (Bulk rigidity away from zero) For every $\varepsilon > 0$ and $D > 0$, there exists a constant $C_{\varepsilon,D} > 0$ such that

$$\mathbb{P}\left(\exists\, \tau \in (\delta, 10 s_*] : \pi(\tau) \geq \varepsilon_*, |\lambda_{i(\tau)} - \tau| \geq \frac{p^\varepsilon}{p} \right) \leq \frac{C_{\varepsilon,D}}{p^D} \tag{4.2.6}$$

holds true for all $p \in \mathbb{N}$.

The constant $C_{\varepsilon,D}$ depends, in addition to ε and D, only on the model parameters as well as on δ and ε_*.

(ii) Away from zero, all eigenvalues lie in the vicinity of the support of ν, i.e., a.w.o.p.

$$\operatorname{Spec}(XX^*) \cap \{\tau; |\tau| \geq \delta,\ \operatorname{dist}(\tau, \operatorname{supp}\nu) \geq \varepsilon_*\} = \varnothing. \tag{4.2.7}$$

In the following two subsections, we distinguish between square Gram matrices, $n = p$, and properly rectangular Gram matrices, $|p/n - 1| \geq d_* > 0$, in order to extend the local law, Theorem 4.2.3, to include zero in the spectral domain $\mathbb{D}$. Since the density of states behaves differently around zero in these two cases, separate statements and proofs are necessary.

4.2.1. Square Gram matrices. The following concept is well-known in linear algebra. For understanding singularities of the density of states in random matrix theory, it was introduced in [4].

Definition 4.2.6 (Fully indecomposable matrix)**.** A $K \times K$ matrix $T = (t_{ij})_{i,j=1}^K$ with nonegative entries is called *fully indecomposable* if for any two subsets $I, J \subset \{1, \ldots, K\}$ such that $\#I + \#J \geq K$, the submatrix $(t_{ij})_{i \in I, j \in J}$ contains a nonzero entry.

For square Gram matrices, we add the following assumptions.

(E1) The matrix X is square, i.e., $n = p$.

(F1) The matrix S is *block fully indecomposable*, i.e., there are constants $\varphi > 0$, $K \in \mathbb{N}$, a fully indecomposable matrix $Z = (z_{ij})_{i,j=1}^K$ with $z_{ij} \in \{0, 1\}$ and a partition $(I_i)_{i=1}^K$ of $\{1, \ldots, p\}$ such that

$$\#I_i = \frac{p}{K}, \quad s_{xy} \geq \frac{\varphi}{p+n} z_{ij}, \quad x \in I_i \text{ and } y \in I_j$$

for all $i, j = 1, \ldots, K$.

The constants φ and K in (F1) are considered model parameters as well.

Remark 4.2.7. Clearly, (E1) yields (D) with $r_1 = r_2 = 1$. Moreover, adapting the proof of Theorem 2.2.1 in [29], we see that (F1) implies (B) with L_1, L_2, ψ_1 and ψ_2 explicitly depending on φ and K.

Theorem 4.2.8 (Local law for square Gram matrices)**.** *If X satisfies (A), (C), (E1) and (F1), then*

(i) The conclusions of Theorem 4.2.3 are valid with the following modifications: (4.2.3b) *and* (4.2.4) *hold true for $\delta = 0$ (cf.* (4.2.2)) *while instead of* (4.2.3a),

we have

$$\mathbb{P}\Big(\exists \zeta \in \mathbb{D}_0, \exists i, j : \operatorname{Im}\zeta \geq p^{-1+\gamma},\ \pi(\operatorname{Re}\zeta) \geq \varepsilon_*,\\ |R_{ij}(\zeta) - m_i(\zeta)\delta_{ij}| \geq p^{\varepsilon}\sqrt{\frac{\langle \operatorname{Im} m(\zeta)\rangle}{p \operatorname{Im}\zeta}}\Big) \leq \frac{C_{\varepsilon,D}}{p^D}. \tag{4.2.8}$$

(ii) $\pi_* = 0$ *and the limit* $\lim_{\omega\downarrow 0}\pi(\omega)\sqrt{\omega}$ *exists and lies in* $(0,\infty)$.

(iii) (Bulk rigidity down to zero) For every $\varepsilon_* > 0$ *and every* $\varepsilon > 0$ *and* $D > 0$, *there exists a constant* $C_{\varepsilon,D} > 0$ *such that*

$$\mathbb{P}\left(\exists\, \tau \in (0, 10s_*] : \pi(\tau) \geq \varepsilon_*, |\lambda_{i(\tau)} - \tau| \geq \frac{p^{\varepsilon}}{p}\left(\sqrt{\tau} + \frac{1}{p}\right)\right) \leq \frac{C_{\varepsilon,D}}{p^D} \tag{4.2.9}$$

for all $p \in \mathbb{N}$. *The constant* $C_{\varepsilon,D}$ *depends, in addition to* ε *and* D, *only on the model parameters and on* ε_*.

(iv) There are no eigenvalues away from the support of ν, *i.e.,* (4.2.7) *holds true with* $\delta = 0$.

We remark that the bound of the individual resolvent entries (4.2.8) deteriorates as ζ gets close to zero since $\langle \operatorname{Im} m(\zeta)\rangle \sim |\zeta|^{-1/2}$ in this regime while the averaged version, (4.2.4), with $\delta = 0$, does not show this behaviour.

4.2.2. Properly rectangular Gram matrices.

(E2) The matrix X is properly rectangular, i.e., there is $d_* > 0$ such that

$$\left|\frac{p}{n} - 1\right| \geq d_*.$$

(F2) The matrix elements of S are bounded from below, i.e., there is a $\varphi > 0$ such that

$$s_{ik} \geq \frac{\varphi}{n+p}$$

for all $i = 1, \dots, p$ and $k = 1, \dots, n$.

The constants d_* and φ in (E2) and (F2), respectively, are also considered as model parameters. Note that (F2) is a simpler version of (F1). For properly rectangular Gram matrices we work under the stronger condition (F2) for simplicity but our analysis could be adjusted to some weaker condition as well.

Remark 4.2.9. Note that (F2) immediately implies condition (B) with $L = 1$.

We introduce the lower edge of the absolutely continuous part of the distribution ν for properly rectangular Gram matrices

$$\delta_\pi := \inf\{\omega > 0 \colon \pi(\omega) > 0\}. \tag{4.2.10}$$

Theorem 4.2.10 (Local law for properly rectangular Gram matrices)**.** *Let X be a random matrix satisfying (A), (C), (D), (E2) and (F2). We have*

(i) The gap between zero and the lower edge is macroscopic $\delta_\pi \sim 1$.

(ii) (Bulk rigidity down to zero) The estimate (4.2.6) *holds true with $\delta = 0$.*

(iii) There are no eigenvalues away from the support of ν, i.e., (4.2.7) *holds true with $\delta = 0$.*

(iv) If $p > n$, then $\pi_* = 1 - n/p$ and $\dim\ker(XX^*) = p - n$ a.w.o.p.

(v) If $p < n$, then $\pi_* = 0$ and $\dim\ker(XX^*) = 0$ a.w.o.p.

(vi) (Local law around zero) For every $\varepsilon_* \in (0, \delta_\pi)$, every $\varepsilon > 0$ and $D > 0$, there exists a constant $C_{\varepsilon,D} > 0$, such that

$$\mathbb{P}\Big(\exists\, \zeta \in \mathbb{H}, i, j \in \{1, \ldots, p\} \colon |\zeta| \le \delta_\pi - \varepsilon_*, \\ |R_{ij}(\zeta) - m_i(\zeta)\delta_{ij}| \ge \frac{p^\varepsilon}{|\zeta|\sqrt{p}}\Big) \le \frac{C_{\varepsilon,D}}{p^D}, \tag{4.2.11}$$

for all $p \in \mathbb{N}$ if $p > n$ and

$$\mathbb{P}\Big(\exists\, \zeta \in \mathbb{H}, i, j \in \{1, \ldots, p\} \colon |\zeta| \le \delta_\pi - \varepsilon_*, \\ |R_{ij}(\zeta) - m_i(\zeta)\delta_{ij}| \ge \frac{p^\varepsilon}{\sqrt{p}}\Big) \le \frac{C_{\varepsilon,D}}{p^D}, \tag{4.2.12}$$

for all $p \in \mathbb{N}$ if $p < n$. Moreover, in both cases

$$\mathbb{P}\Big(\exists\, \zeta \in \mathbb{H} \colon |\zeta| \le \delta_\pi - \varepsilon_*, \left|\frac{1}{p}\sum_{i=1}^{p}[R_{ii}(\zeta) - m_i(\zeta)]\right| \ge \frac{p^\varepsilon}{p}\Big) \le \frac{C_{\varepsilon,D}}{p^D}, \tag{4.2.13}$$

for all $p \in \mathbb{N}$.

The constant $C_{\varepsilon,D}$ depends, in addition to ε and D, only on the model parameters and on ε_*.

If $p > n$, then the Stieltjes transform of the empirical spectral measure of XX^* has a term proportional to $1/\zeta$ due to the macroscopic kernel of XX^*. This is the origin of the additional factor $1/|\zeta|$ in (4.2.11).

Remark 4.2.11. As a consequence of Theorem 4.2.8 and Theorem 4.2.10 and under the same conditions, the standard methods in [36] and [7] can be used to prove an anisotropic law and delocalization of eigenvectors in the bulk.

4.3. Quadratic vector equation

For the rest of the paper, without loss of generality, we will assume that $s_* = 1$ in (A), which can be achieved by a simple rescaling of X. In the whole section, we will assume that the matrix S satisfies (A), (B) and (D) without further notice.

4.3.1. Self-consistent equation for resolvent entries. We introduce the random matrix $\boldsymbol{H}$ and the deterministic matrix $\boldsymbol{S}$ defined through

$$\boldsymbol{H} = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}, \qquad \boldsymbol{S} = \begin{pmatrix} 0 & S \\ S^t & 0 \end{pmatrix}. \tag{4.3.1}$$

Note that both matrices, $\boldsymbol{H}$ and $\boldsymbol{S}$ have dimensions $(p+n) \times (p+n)$. We denote their entries by $\boldsymbol{H} = (h_{xy})_{x,y}$ and $\boldsymbol{S} = (\sigma_{xy})_{x,y}$, respectively, where $\sigma_{xy} = \mathbb{E}|h_{xy}|^2$ with $x, y = 1, \ldots, n+p$.

It is easy to see that condition (B) implies

(B') There are $L \in \mathbb{N}$ and $\psi > 0$ such that

$$\sum_{k=1}^{L} (\boldsymbol{S}^k)_{xy} \geq \frac{\psi}{n+p} \tag{4.3.2}$$

for all $x, y = 1, \ldots, n+p$.

In the following, a crucial part of the analysis will be devoted to understanding the resolvent of $\boldsymbol{H}$ at $z \in \mathbb{H}$, i.e., the matrix

$$\boldsymbol{G}(z) := (\boldsymbol{H} - z)^{-1} \tag{4.3.3}$$

whose entries are denoted by $G_{xy}(z)$ for $x, y = 1, \dots, n+p$. For $V \subset \{1, \dots, n+p\}$, we use the notation $G^{(V)}_{xy}$ to denote the entries of the resolvent $\boldsymbol{G}^{(V)}(z) = (\boldsymbol{H}^{(V)} - z)^{-1}$ of the matrix $\boldsymbol{H}^{(V)}_{xy} = h_{xy}\mathbf{1}(x \notin V)\mathbf{1}(y \notin V)$ where $x, y = 1, \dots, n+p$.

The Schur complement formula and the resolvent identities applied to $\boldsymbol{G}(z)$ yield the self-consistent equations

$$-\frac{1}{g_{1,i}(z)} = z + \sum_{k=1}^{n} s_{ik} g_{2,k}(z) + d_{1,i}(z), \tag{4.3.4a}$$

$$-\frac{1}{g_{2,k}(z)} = z + \sum_{i=1}^{p} s_{ik} g_{1,i}(z) + d_{2,k}(z), \tag{4.3.4b}$$

where $g_{1,i}(z) := G_{ii}(z)$ for $i = 1, \dots, p$ and $g_{2,k}(z) := G_{k+p,k+p}(z)$ for $k = 1, \dots, n$ with the error terms

$$d_{1,r} := \sum_{k,l=1, k\neq l}^{n} x_{rk} G^{(r)}_{kl} \overline{x}_{rl} + \sum_{k=1}^{n} \left(|x_{rk}|^2 - s_{rk} \right) G^{(r)}_{k+n,k+n} - \sum_{k=1}^{n} s_{rk} \frac{G_{k+n,r} G_{r,k+n}}{g_{1,r}},$$

$$d_{2,m} := \sum_{i,j=1, i\neq j}^{p} \overline{x}_{im} G^{(m+p)}_{ij} x_{jm} + \sum_{i=1}^{p} \left(|x_{im}|^2 - s_{im} \right) G^{(m+p)}_{ii} - \sum_{i=1}^{p} s_{im} \frac{G_{i,m+p} G_{m+p,i}}{g_{2,m}}$$

for $r = 1, \dots, p$ and $m = 1, \dots, n$.

We will prove a local law which states that $g_{1,i}(z)$ and $g_{2,k}(z)$ can be approximated by $(m_1(z))_i$ and $(m_2(z))_k$, respectively, where $m_1 \colon \mathbb{H} \to \mathbb{C}^p$ and $m_2 \colon \mathbb{H} \to \mathbb{C}^n$ are the unique solution of

$$-\frac{1}{m_1} = z + S m_2, \tag{4.3.5a}$$

$$-\frac{1}{m_2} = z + S^t m_1, \tag{4.3.5b}$$

which satisfy $\operatorname{Im} m_1(z) > 0$ and $\operatorname{Im} m_2(z) > 0$ for all $z \in \mathbb{H}$.

The system of self-consistent equations for g_1 and g_2 in (4.3.4) can be seen as a perturbation of the system (4.3.5). With the help of $\boldsymbol{S}$, equations (4.3.5a) and (4.3.5b) can be combined to a vector equation for $\boldsymbol{m} = (m_1, m_2)^t \in \mathbb{H}^{p+n}$, i.e.,

$$-\frac{1}{\boldsymbol{m}} = z + \boldsymbol{S}\boldsymbol{m}. \tag{4.3.6}$$

Since $\boldsymbol{S}$ is symmetric, has nonnegative entries and fulfills (A) with $s_* = 1$, Theorem 2.1 in [4] is applicable to (4.3.6). Here, we take $a = 0$ in Theorem 2.1 of [4]. This theorem implies that (4.3.6) has a unique solution $\boldsymbol{m}$ with $\operatorname{Im} \boldsymbol{m}(z) > 0$ for any $z \in \mathbb{H}$. Moreover, by this theorem, $\boldsymbol{m}_x$ is the Stieltjes transform of a symmetric probability measure on $\mathbb{R}$ whose support is contained in $[-2, 2]$ for all $x = 1, \ldots, n+p$ and we have

$$\|\boldsymbol{m}(z)\|_2 \leq \frac{2}{|z|} \tag{4.3.7}$$

for all $z \in \mathbb{H}$. The function $\langle \boldsymbol{m} \rangle$ is the Stieltjes transform of a symmetric probability measure on $\mathbb{R}$ which we denote by ρ, i.e.,

$$\langle \boldsymbol{m}(z) \rangle = \int_{\mathbb{R}} \frac{1}{t-z} \rho(\mathrm{d}t) \tag{4.3.8}$$

for $z \in \mathbb{H}$. Its support is contained in $[-2, 2]$.

We combine (4.3.4a) and (4.3.4b) to obtain

$$-\frac{1}{\boldsymbol{g}} = z + \boldsymbol{S}\boldsymbol{g} + \boldsymbol{d}, \tag{4.3.9}$$

where $\boldsymbol{g} = (g_1, g_2)^t$ and $\boldsymbol{d} = (d_1, d_2)^t$. We think of (4.3.9) as a perturbation of (4.3.6) and most of the subsequent subsection is devoted to the study of (4.3.9) for an arbitrary perturbation $\boldsymbol{d}$.

Before we start studying (4.3.6) we want to indicate how m and R are related to $\boldsymbol{m} = (m_1, m_2)^t$ and $\boldsymbol{G}$, respectively. The Stieltjes transforms as well as the resolvents are essentially related via the same transformation of the spectral parameter. If $\boldsymbol{G}_{11}(z)$ denotes the upper left $p \times p$ block of $\boldsymbol{G}(z)$ then $R(z^2) = (XX^* - z^2)^{-1} = \boldsymbol{G}_{11}(z)/z$. In the proof of Theorem 4.2.2 in Subsection 4.3.4, we will see that m and m_1 are related via $m(\zeta) = m_1(\sqrt{\zeta})/\sqrt{\zeta}$. (We always choose the branch of the square root satisfying $\operatorname{Im} \sqrt{\zeta} > 0$ for $\operatorname{Im} \zeta > 0$.) Assuming this relation and introducing $\widetilde{m}_2(\zeta) := m_2(\sqrt{\zeta})/\sqrt{\zeta}$, we obtain

$$\begin{aligned} -\frac{1}{m(\zeta)} &= \zeta(1 + S\widetilde{m}_2(\zeta)), \\ -\frac{1}{\widetilde{m}_2(\zeta)} &= \zeta(1 + S^t m(\zeta)) \end{aligned} \tag{4.3.10}$$

from (4.3.5). Solving the second equation for $\widetilde{m}_2$ and plugging the result into the first one yields (4.2.1) immediately. In fact, $\widetilde{m}_2$ is the analogue of m corresponding to X^*X, i.e, the Stieltjes transform of the deterministic measure approximating the eigenvalue density of X^*X.

4.3.2. Structure of the solution. We first notice that the inequality $s_{ik} \le 1/(n+p)$ implies

$$\|S^t w\|_\infty = \max_{k=1,\dots,n} \sum_{i=1}^p s_{ik}|w_i| \le \max_{k=1,\dots,n} \left(p \sum_{i=1}^p s_{ik}^2 \right)^{1/2} \left(\frac{1}{p} \sum_{i=1}^p |w_i|^2 \right)^{1/2} \le \|w\|_2 \tag{4.3.11}$$

for all $w \in \mathbb{C}^p$, i.e., $\|S^t\|_{2\to\infty} \le 1$. Now, we establish some preliminary estimates on the solution of (4.3.6).

Lemma 4.3.1. *Let $z \in \mathbb{H}$ and $x \in \{1, \dots, n+p\}$. We have*

$$|\boldsymbol{m}_x(z)| \le \frac{1}{\operatorname{dist}(z, \operatorname{supp} \rho)}, \tag{4.3.12a}$$

$$\operatorname{Im} \boldsymbol{m}_x(z) \le \frac{\operatorname{Im} z}{\operatorname{dist}(z, \operatorname{supp} \rho)^2}. \tag{4.3.12b}$$

If $z \in \mathbb{H}$ and $|z| \le 10$ then

$$|z| \lesssim |\boldsymbol{m}_x(z)| \le \|\boldsymbol{m}(z)\|_\infty \lesssim \frac{|z|^{2-2L}}{\langle \operatorname{Im} \boldsymbol{m}(z) \rangle} \tag{4.3.13a}$$

$$|z|^{2L} \langle \operatorname{Im} \boldsymbol{m}(z) \rangle \lesssim \operatorname{Im} \boldsymbol{m}_x(z). \tag{4.3.13b}$$

In particular, the support of the measures representing $\boldsymbol{m}_x$ is independent of x away from zero.

The proof essentially follows the same line of arguments as the proof of Lemma 5.4 in [4]. However, instead of using the lower bound on the entries of S^L as in [4] we have to make use of the lower bound on the entries of $\sum_{k=1}^L \boldsymbol{S}^k$.

To prove another auxiliary estimate on $\boldsymbol{S}$, we define the vectors $\boldsymbol{S}_x = (\sigma_{xy})_{y=1,\dots,n+p} \in \mathbb{R}^{n+p}$ for $x = 1, \dots, n+p$. Since (4.3.2) implies

$$\psi \le \sum_{k=1}^L \sum_{y=1}^{n+p} (\boldsymbol{S}^k)_{xy} \le \sum_{k=1}^L \sum_{v=1}^{n+p} \sigma_{xv} \max_{t=1,\dots,n+p} \sum_{y=1}^{n+p} (\boldsymbol{S}^{k-1})_{ty} \le L \sum_{v=1}^{n+p} \sigma_{xv}$$

for any fixed $x = 1, \dots, n+p$, where we used $\|\boldsymbol{S}^{k-1}\|_\infty \le \|\boldsymbol{S}\|_\infty^{k-1} \le 1$ by (A), we obtain

$$\inf_{x=1,\dots,n+p} \|\boldsymbol{S}_x\|_1 \ge \frac{\psi}{L}. \tag{4.3.14}$$

In particular, together with (A), this implies

$$\sum_{j=1}^{p} s_{jk} \sim 1, \quad \sum_{l=1}^{n} s_{il} \sim 1, \qquad i = 1, \ldots, p, \quad k = 1, \ldots, n. \tag{4.3.15}$$

In the study of the stability of (4.3.6) when perturbed by a vector $\boldsymbol{d}$, as in (4.3.9), the linear operator

$$\boldsymbol{F}(z)\boldsymbol{w} := |\boldsymbol{m}(z)|\boldsymbol{S}(|\boldsymbol{m}(z)|\boldsymbol{w}) \tag{4.3.16}$$

for $\boldsymbol{w} \in \mathbb{C}^{n+p}$ plays an important role. Before we collect some properties of operators of this type in the next lemma, we first recall the definition of the gap of an operator from [4].

Definition 4.3.2. Let T be a compact self-adjoint operator on a Hilbert space. The *spectral gap* $\mathrm{Gap}(T) \geq 0$ is the difference between the two largest eigenvalues of $|T|$ (defined by spectral calculus). If the operator norm $\|T\|$ is a degenerate eigenvalue of $|T|$, then $\mathrm{Gap}(T) = 0$.

In the next lemma, we study matrices of the form $\widehat{\boldsymbol{F}}(r)_{xy} := r_x \sigma_{xy} r_y$ where $r \in (0,\infty)^{n+p}$ and $x, y = 1, \ldots, n+p$. If $\inf_x r_x > 0$ then (4.3.2) implies that all entries of $\sum_{k=1}^{L} \widehat{\boldsymbol{F}}(r)^k$ are strictly positive. Therefore, by the Perron-Frobenius theorem, the eigenspace corresponding to the largest eigenvalue $\widehat{\lambda}(r)$ of $\widehat{\boldsymbol{F}}(r)$ is one-dimensional and spanned by a unique non-negative vector $\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{f}}(r)$ such that $\langle \widehat{\boldsymbol{f}}, \widehat{\boldsymbol{f}} \rangle = 1$.

The block structure of $\boldsymbol{S}$ implies that there is a matrix $\widehat{F}(r) \in \mathbb{R}^{p\times n}$ such that

$$\widehat{\boldsymbol{F}}(r) = \begin{pmatrix} 0 & \widehat{F}(r) \\ \widehat{F}(r)^t & 0 \end{pmatrix}. \tag{4.3.17}$$

However, for this kind of operator, we obtain $\mathrm{Spec}\,(\widehat{\boldsymbol{F}}(r)) = -\,\mathrm{Spec}\,(\widehat{\boldsymbol{F}}(r))$, i.e., $\mathrm{Gap}(\widehat{\boldsymbol{F}}(r)) = 0$ by above definition. Therefore, we will compute $\mathrm{Gap}(\widehat{F}(r)\widehat{F}(r)^t)$, instead. We will apply these observations for $\boldsymbol{F}(z)$ where the blocks $\widehat{F}(|\boldsymbol{m}(z)|)$ will be denoted by $F(z)$.

Lemma 4.3.3. *For a vector $r \in (0,\infty)^{n+p}$ which is bounded by constants $r_+ \in (0,\infty)$ and $r_- \in (0,1]$, i.e.,*

$$r_- \leq r_x \leq r_+$$

for all $x = 1, \ldots, n+p$, we define the matrix $\widehat{\boldsymbol{F}}(r)$ with entries $\widehat{\boldsymbol{F}}(r)_{xy} := r_x \sigma_{xy} r_y$ for $x, y = 1, \ldots, n+p$. Then the eigenspace corresponding to $\widehat{\lambda}(r) := \|\widehat{\boldsymbol{F}}(r)\|_{2\to 2}$ is one-dimensional and $\widehat{\lambda}(r)$ satisfies the estimates

$$r_-^2 \lesssim \widehat{\lambda}(r) \lesssim r_+^2. \tag{4.3.18}$$

There is a unique eigenvector $\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{f}}(r)$ corresponding to $\widehat{\lambda}(r)$ satisfying $\widehat{\boldsymbol{f}}_x \geq 0$ and $\|\widehat{\boldsymbol{f}}\|_2 = 1$. Its components satisfy

$$\frac{r_-^{2L}}{r_+^4} \min\left\{\widehat{\lambda}(r), \widehat{\lambda}(r)^{-L+2}\right\} \lesssim \widehat{\boldsymbol{f}}_x \lesssim \frac{r_+^4}{\widehat{\lambda}(r)^2}, \qquad \textit{for all } x = 1, \ldots, n+p. \tag{4.3.19}$$

Moreover, $\widehat{F}(r)\widehat{F}(r)^t$ has a spectral gap

$$\operatorname{Gap}\left(\widehat{F}(r)\widehat{F}(r)^t\right) \gtrsim \frac{r_-^{8L}}{r_+^{16}} \min\left\{\widehat{\lambda}(r)^6, \widehat{\lambda}(r)^{-8L+10}\right\}. \tag{4.3.20}$$

The estimates in (4.3.18) and (4.3.19) can basically be proven following the proof of Lemma 5.6 in [4] where S^L is replaced by $\sum_{k=1}^L \boldsymbol{S}^k$ and $(\widehat{F}/\widehat{\lambda})^L$ by $\sum_{k=1}^L (\widehat{\boldsymbol{F}}/\widehat{\lambda})^k$. Therefore, we will only show (4.3.20) assuming the other estimates.

Proof. We write $\widehat{\boldsymbol{f}} = (\widehat{f}_1, \widehat{f}_2)^t$ for $\widehat{f}_1 \in \mathbb{C}^p$ and $\widehat{f}_2 \in \mathbb{C}^n$ and define a linear operator on $\mathbb{C}^p$ through

$$T := \sum_{k=1}^{L} \left(\frac{\widehat{F}\widehat{F}^t}{\widehat{\lambda}^2}\right)^k.$$

Thus, $\|T\|_2 = L$ as $T\widehat{f}_1 = L\widehat{f}_1$. Using (B') we first estimate the entries t_{ij} by

$$t_{ij} \geq \sum_{k=1}^{L} \frac{r_-^{4k}}{\widehat{\lambda}^{2k}} \left((SS^t)^k\right)_{ij} \geq r_-^{4L} \min\left\{\widehat{\lambda}^{-2}, \widehat{\lambda}^{-2L}\right\} \frac{\psi}{n+p}, \qquad \text{for } i, j = 1, \ldots, p.$$

Estimating $\|\widehat{f}_1\|_2$ and $\|\widehat{f}_1\|_\infty$ from (4.3.19) and applying Lemma 5.6 in [5] or Lemma 5.7 in [4] yield

$$\operatorname{Gap}(T) \geq \frac{\|\widehat{f}_1\|_2^2}{\|\widehat{f}_1\|_\infty^2} p \inf_{i,j} t_{ij} \gtrsim \frac{r_-^{8L}}{r_+^{16}} \min\left\{\widehat{\lambda}^4, \widehat{\lambda}^{-8L+8}\right\}.$$

Here we used (D) and note that the factor $\inf_{i,j} t_{ij}$ in Lemma 5.6 in [5] is replaced by $p \inf_{i,j} t_{ij}$ as t_{ij} are considered as the matrix entries of T and not as the kernel of an integral operator on $L^2(\{1, \ldots, p\})$ where $\{1, \ldots, p\}$ is equipped with the uniform probability measure. As $q(x) := x + x^2 + \ldots + x^L$ is a monotonously increasing, differentiable function on $[0,1]$ and $\mathrm{Spec}(\widehat{F}\widehat{F}^t/\widehat{\lambda}^2) \subset [0,1]$ we obtain $\mathrm{Gap}(T) \sim \mathrm{Gap}(\widehat{F}\widehat{F}^t)/\widehat{\lambda}^2$ which concludes the proof. ☐

Lemma 4.3.4. *The matrix* $\boldsymbol{F}(z)$ *defined in* (4.3.16) *with entries*

$$\boldsymbol{F}_{xy}(z) = |\boldsymbol{m}_x(z)|\sigma_{xy}|\boldsymbol{m}_y(z)|$$

has the norm

$$\|\boldsymbol{F}(z)\|_2 = 1 - \frac{\mathrm{Im}\, z\langle \boldsymbol{f}(z)|\boldsymbol{m}(z)|\rangle}{\langle \boldsymbol{f}(z)\mathrm{Im}\,\boldsymbol{m}(z)|\boldsymbol{m}(z)|^{-1}\rangle}, \tag{4.3.21}$$

where $\boldsymbol{f}(z)$ *is the unique eigenvector of* $\boldsymbol{F}(z)$ *associated to* $\|\boldsymbol{F}(z)\|_2$. *In particular, we obtain*

$$(1 - \|\boldsymbol{F}(z)\|_2)^{-1} \lesssim \frac{1}{|z|} \min\left\{\frac{1}{\mathrm{Im}\, z}, \frac{1}{|z|\,\mathrm{dist}(z, \mathrm{supp}\,\rho)^2}\right\} \tag{4.3.22}$$

for $z \in \mathbb{H}$ *satisfying* $|z| \leq 10$.

Proof. The derivation of (4.3.21) follows the same steps as the proof of (4.4) in [5] (compare Lemma 5.5 in [4] as well). We take the imaginary part of (4.3.6), multiply the result by $|\boldsymbol{m}|$ and take the scalar product with $\boldsymbol{f}$. Thus, we obtain

$$\left\langle \boldsymbol{f}\,,\, \frac{\mathrm{Im}\,\boldsymbol{m}}{|\boldsymbol{m}|}\right\rangle = \mathrm{Im}\, z\langle \boldsymbol{f}|\boldsymbol{m}|\rangle + \|\boldsymbol{F}\|_2 \left\langle \boldsymbol{f}\,,\, \frac{\mathrm{Im}\,\boldsymbol{m}}{|\boldsymbol{m}|}\right\rangle, \tag{4.3.23}$$

where we used the symmetry of $\boldsymbol{F}$ and $\boldsymbol{F}\boldsymbol{f} = \|\boldsymbol{F}\|_2\boldsymbol{f}$. Solving (4.3.23) for $\|\boldsymbol{F}\|_2$ yields (4.3.21).

Now, (4.3.22) is a direct consequence of Lemma 4.3.1 and (4.3.21). ☐

4.3.3. Stability away from the edges and continuity. All estimates of $\boldsymbol{m} - \boldsymbol{g}$, when $\boldsymbol{m}$ and $\boldsymbol{g}$ satisfy (4.3.6) and (4.3.9), respectively, are based on inverting the linear operator

$$\boldsymbol{B}(z)\boldsymbol{w} := \frac{|\boldsymbol{m}(z)|^2}{\boldsymbol{m}(z)^2}\boldsymbol{w} - \boldsymbol{F}(z)\boldsymbol{w}$$

for $\boldsymbol{w} \in \mathbb{C}^{n+p}$. The following lemma bounds $\boldsymbol{B}^{-1}(z)$ in terms of $\langle \mathrm{Im}\,\boldsymbol{m}(z)\rangle$ if z is away from zero. For $\delta > 0$, we use the notation $f \lesssim_\delta g$ if and only if there is an $r > 0$ which is allowed to depend on model parameters such that $f \lesssim \delta^{-r} g$.

Lemma 4.3.5. *There is a universal constant $\kappa \in \mathbb{N}$ such that for all $\delta > 0$ we have*

$$\|\boldsymbol{B}^{-1}(z)\|_2 \lesssim_\delta \min\left\{\frac{1}{(\operatorname{Re} z)^2 \langle \operatorname{Im} \boldsymbol{m}(z)\rangle^\kappa}, \frac{1}{\operatorname{Im} z}, \frac{1}{\operatorname{dist}(z, \operatorname{supp} \rho)^2}\right\}, \tag{4.3.24}$$

$$\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim_\delta \min\left\{\frac{1}{(\operatorname{Re} z)^2 \langle \operatorname{Im} \boldsymbol{m}(z)\rangle^{\kappa+2}}, \frac{1}{(\operatorname{Im} z)^3}, \frac{1}{\operatorname{dist}(z, \operatorname{supp} \rho)^4}\right\} \tag{4.3.25}$$

for all $z \in \mathbb{H}$ satisfying $\delta \le |z| \le 10$.

For the proof of this result, we will need the two following lemmata. We recall that by the Perron-Frobenius theorem an irreducible matrix with nonnegative entries has a unique ℓ^2-normalized eigenvector with positive entries corresponding to its largest eigenvalue. By the definition of the spectral gap, Definition 4.3.2, we observe that if AA^* is irreducible then $\operatorname{Gap}(AA^*) = \|AA^*\|_2 - \max(\operatorname{Spec}(AA^*) \setminus \{\|AA^*\|_2\})$.

Lemma 4.3.6 (Rotation-Inversion Lemma)**.** *There exists a positive constant C such that for all $n, p \in \mathbb{N}$, unitary matrices $U_1 \in \mathbb{C}^{p\times p}$, $U_2 \in \mathbb{C}^{n\times n}$ and $A \in \mathbb{R}^{p\times n}$ with nonnegative entries such that A^*A and AA^* are irreducible and $\|A^*A\|_2 \in (0, 1]$, the following bound holds:*

$$\left\| \begin{pmatrix} U_1 & A \\ A^* & U_2 \end{pmatrix}^{-1} \right\|_2 \le \frac{C}{\operatorname{Gap}(AA^*)|1 - \|A^*A\|_2 \langle v_1\,, U_1 v_1\rangle \langle v_2\,, U_2 v_2\rangle|}, \tag{4.3.26}$$

where $v_1 \in \mathbb{C}^p$ and $v_2 \in \mathbb{C}^n$ are the unique positive, normalized eigenvectors with $AA^*v_1 = \|A^*A\|_2 v_1$ and $A^*Av_2 = \|A^*A\|_2 v_2$. The norm on the left-hand side of (4.3.26) *is infinite if and only if the right-hand side of* (4.3.26) *is infinite, i.e., in this case the inverse does not exist.*

This lemma is proven in the Section 4.5 below.

Lemma 4.3.7. *Let $R\colon \mathbb{C}^{n+p} \to \mathbb{C}^{n+p}$ be a linear operator and $D\colon \mathbb{C}^{n+p} \to \mathbb{C}^{n+p}$ a diagonal operator. If $R - D$ is invertible and $D_{xx} \neq 0$ for all $x = 1, \dots, n+p$ then*

$$\|(R-D)^{-1}\|_\infty \le \left(\inf_{x=1}^{n+p} |D_{xx}|\right)^{-1} \Big(1 + \|R\|_{2\to\infty} \|(R-D)^{-1}\|_2\Big). \tag{4.3.27}$$

The proof of (4.3.27) follows a similar way as the proof of (5.28) in [4].

Proof of Lemma 4.3.5. The bound on $\|\boldsymbol{B}^{-1}(z)\|_\infty$, (4.3.25), follows from (4.3.24) by employing (4.3.27). We use (4.3.27) with $R = \boldsymbol{F}(z)$ and $D = |\boldsymbol{m}(z)|^2/\boldsymbol{m}(z)^2$ and observe that $\|\boldsymbol{F}(z)\|_{2\to\infty} \le \|\boldsymbol{m}\|_\infty^2 \|\boldsymbol{S}\|_{2\to\infty}$. Therefore, (4.3.25) follows from (4.3.24) as $\|\boldsymbol{m}\|_\infty \lesssim \min\{\langle \operatorname{Im} \boldsymbol{m}\rangle^{-1}, (\operatorname{Im} z)^{-1}, \operatorname{dist}(z, \operatorname{supp}\rho)^{-1}\}$ by (4.3.13a) and $\min\{\langle \operatorname{Im} \boldsymbol{m}\rangle^{-1}, (\operatorname{Im} z)^{-1}, \operatorname{dist}(z, \operatorname{supp}\rho)^{-1}\} \gtrsim_\delta 1$ by (4.3.13a) and $\delta \le |z| \le 10$.

Now we prove (4.3.24). Our first goal is the following estimate

$$\|\boldsymbol{B}^{-1}(z)\|_2 \lesssim_\delta \frac{1}{\operatorname{Gap}(F(z)F(z)^t)(\operatorname{Re} z)^2 \langle \operatorname{Im} \boldsymbol{m}(z)\rangle^\kappa} \tag{4.3.28}$$

for some universal $\kappa \in \mathbb{N}$ which will be a consequence of Lemma 4.3.6. We apply this lemma with

$$\begin{pmatrix} 0 & F(z) \\ F(z)^t & 0 \end{pmatrix} = \boldsymbol{F}(z) := \widehat{\boldsymbol{F}}(|\boldsymbol{m}(z)|), \quad \mathfrak{U} := \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} = \operatorname{diag}\left(\frac{|\boldsymbol{m}(z)|^2}{\boldsymbol{m}(z)^2}\right)$$

and $v_1 := f_1/\|f_1\|_2$ and $v_2 := f_2/\|f_2\|_2$ where $\boldsymbol{f} = (f_1, f_2)^t \in \mathbb{C}^{p+n}$. Note that $\lambda(z) := \widehat{\lambda}(|\boldsymbol{m}(z)|) = \|\boldsymbol{F}(z)\|_2$ in Lemma 4.3.3 and $F(z) = \widehat{F}(|\boldsymbol{m}(z)|)$ in the notation of (4.3.17). In Lemma 4.3.3, we choose $r_- := \inf_x |\boldsymbol{m}_x(z)|$ and $r_+ := \|\boldsymbol{m}(z)\|_\infty$ and use the bounds $r_- \gtrsim |z|$ and $r_+ \lesssim |z|^{2-2L}/\langle \operatorname{Im} \boldsymbol{m}(z)\rangle$ by (4.3.13a). Moreover, we have

$$|z|^2 \lesssim \|\boldsymbol{F}(z)\|_2 \le 1 \tag{4.3.29}$$

by (4.3.13a), (4.3.18) and (4.3.21).

We write $\mathfrak{U} = \operatorname{diag}(\mathrm{e}^{-\mathrm{i}\,2\boldsymbol{\psi}})$, i.e., $\mathrm{e}^{\mathrm{i}\boldsymbol{\psi}} = \boldsymbol{m}/|\boldsymbol{m}|$, and $\boldsymbol{\psi} = (\psi_1, \psi_2)^t \in \mathbb{R}^{p+n}$ to obtain

$$\langle v_1\,, U_1 v_1\rangle = \langle v_1\,, (\cos\psi_1 - \mathrm{i}\,\sin\psi_1)^2 v_1\rangle \;=\; \langle v_1\,, (1 - 2(\sin\psi_1)^2 - 2\mathrm{i}\,\cos\psi_1 \sin\psi_1) v_1\rangle$$

and a similar relation holds for $\langle v_2\,, U_2 v_2\rangle$. Thus, we compute

$$\begin{aligned} &\operatorname{Re}\Big(1 - \|F(z)^t F(z)\|_2 \langle v_1\,, (1 - 2(\sin\psi_1)^2 - 2\mathrm{i}\,\cos\psi_1\sin\psi_1)v_1\rangle \\ &\qquad \times \langle v_2\,, (1 - 2(\sin\psi_2)^2 - 2\mathrm{i}\,\cos\psi_2\sin\psi_2)v_2\rangle\Big) \\ &\qquad\qquad = 1 - \|F(z)^t F(z)\|_2 (1 - 2\langle v_1\,, (\sin\psi_1)^2 v_1\rangle - 2\langle v_2\,, (\sin\psi_2)^2 v_2\rangle \\ &\qquad\qquad\quad + 4\langle v_1\,, (\sin\psi_1)^2 v_1\rangle\langle v_2\,, (\sin\psi_2)^2 v_2\rangle) \end{aligned}$$

Using $2a+2b-4ab \geq (a+b)(2-a-b)$ for $a,b\in\mathbb{R}$, and estimating the absolute value by the real part yields

$$\begin{aligned}
&\Big|1-\|F(z)^tF(z)\|_2\langle v_1\,,U_1v_1\rangle\langle v_2\,,U_2v_2\rangle\Big| \\
&\qquad\geq 1-\|F(z)^tF(z)\|_2+\|F(z)^tF(z)\|_2\left(\langle v_1,(\sin\psi_1)^2\,v_1\rangle+\langle v_2,(\sin\psi_2)^2\,v_2\rangle\right)\\
&\qquad\quad\times\left(\langle v_1,(\cos\psi_1)^2\,v_1\rangle+\langle v_2,(\cos\psi_2)^2\,v_2\rangle\right)\\
&\qquad\gtrsim |z|^4\langle \boldsymbol{f},(\sin\boldsymbol{\psi})^2\,\boldsymbol{f}\rangle\langle \boldsymbol{f},(\cos\boldsymbol{\psi})^2\,\boldsymbol{f}\rangle\\
&\qquad\gtrsim_\delta \left(\inf_{x=1,\dots,n+p}\boldsymbol{f}_x^4\right)\left\langle\left(\frac{\operatorname{Im}\boldsymbol{m}}{|\boldsymbol{m}|}\right)^2\right\rangle\left\langle\left(\frac{\operatorname{Re}\boldsymbol{m}}{|\boldsymbol{m}|}\right)^2\right\rangle,
\end{aligned} \tag{4.3.30}$$

where we used $1\geq\|F(z)^tF(z)\|_2=\|\boldsymbol{F}\|_2^2\gtrsim|z|^4$ by (4.3.29) and

$$\langle \boldsymbol{f},(\sin\boldsymbol{\psi})^2\,\boldsymbol{f}\rangle\langle \boldsymbol{f},(\cos\boldsymbol{\psi})^2\,\boldsymbol{f}\rangle\leq 1$$

in the second step. In order to estimate the last expression in (4.3.30), we use $r_-\gtrsim|z|$ and $\|\boldsymbol{F}(z)\|_2\leq 1$ by (4.3.29) as well as (4.3.13a), (4.3.18) and (4.3.19) to get for the first factor

$$\inf_{x=1,\dots,n+p}\boldsymbol{f}_x^4\gtrsim r_-^{8L+8}r_+^{-16}\gtrsim_\delta\langle\operatorname{Im}\boldsymbol{m}\rangle^{16}. \tag{4.3.31}$$

To estimate the last factor in (4.3.30), we multiply the real part of (4.3.6) with $|\boldsymbol{m}|$ and obtain

$$(1+\boldsymbol{F})\frac{\operatorname{Re}\boldsymbol{m}}{|\boldsymbol{m}|}=-\tau|\boldsymbol{m}|$$

if $z=\tau+\mathrm{i}\,\eta$ for $\tau,\eta\in\mathbb{R}$. Estimating $\|\cdot\|_2$ of the last equation yields

$$|\tau|\|\boldsymbol{m}\|_2\leq 2\left\|\frac{\operatorname{Re}\boldsymbol{m}}{|\boldsymbol{m}|}\right\|_2$$

by (4.3.29). As $\|\boldsymbol{m}\|_2\geq\|\operatorname{Im}\boldsymbol{m}\|_2\geq\langle\operatorname{Im}\boldsymbol{m}\rangle$ we get

$$2\left\|\frac{\operatorname{Re}\boldsymbol{m}}{|\boldsymbol{m}|}\right\|_2\geq|\tau|\langle\operatorname{Im}\boldsymbol{m}\rangle. \tag{4.3.32}$$

Finally, we use (4.3.31) for the first factor in (4.3.30) and (4.3.32) for the last factor and apply the last estimate in (4.3.13a) and Jensen's inequality, $\langle(\operatorname{Im}\boldsymbol{m})^2\rangle\geq\langle\operatorname{Im}\boldsymbol{m}\rangle^2$, to estimate the second factor which yields

$$\Big|1 - \|F(z)^t F(z)\|_2 \langle v_1\,, U_1 v_1\rangle\langle v_2\,, U_2 v_2\rangle\Big| \gtrsim_\delta |\tau|^2 \langle \mathrm{Im}\,\boldsymbol{m}\rangle^\kappa. \tag{4.3.33}$$

This completes the proof of (4.3.28).

Next, we bound $\mathrm{Gap}(F(z)F(z)^t)$ from below by applying Lemma 4.3.3 with $r_- :=$ $\inf_x |\boldsymbol{m}_x(z)|$ and $r_+ := \|\boldsymbol{m}(z)\|_\infty$. As $F(z) = \widehat{F}(|\boldsymbol{m}(z)|)$ we have

$$\mathrm{Gap}(F(z)F(z)^t) \gtrsim_\delta \langle \mathrm{Im}\,\boldsymbol{m}(z)\rangle^{16},$$

where we used the estimates in (4.3.13a) and (4.3.29). Combining this estimate on $\mathrm{Gap}(F(z)F(z)^t)$ with (4.3.28) and (4.3.22) and increasing κ, we obtain

$$\|\boldsymbol{B}^{-1}(z)\|_2 \lesssim_\delta \min\left\{\frac{1}{(\mathrm{Re}\,z)^2\langle \mathrm{Im}\,\boldsymbol{m}(z)\rangle^\kappa}, \frac{1}{\mathrm{Im}\,z}, \frac{1}{\mathrm{dist}(\mathrm{Re}\,z, \mathrm{supp}\,\rho)^2}\right\}$$

as $\|\boldsymbol{B}^{-1}(z)\|_2 \le (1 - \|\boldsymbol{F}(z)\|_2)^{-1}$ and $\delta \le |z| \le 10$. □

Lemma 4.3.8 (Continuity of the solution)**.** *If $\boldsymbol{m}$ is the solution of the QVE* (4.3.6) *then $z \mapsto \langle \boldsymbol{m}(z)\rangle$ can be extended to a locally Hölder-continuous function on $\overline{\mathbb{H}}\backslash\{0\}$. Moreover, for every $\delta > 0$ there is a constant c depending on δ and the model parameters such that*

$$|\langle \boldsymbol{m}(z_1)\rangle - \langle \boldsymbol{m}(z_2)\rangle| \le c|z_1 - z_2|^{1/(\kappa+1)} \tag{4.3.34}$$

for all $z_1, z_2 \in \overline{\mathbb{H}}\backslash\{0\}$ such that $\delta \le |z_1|, |z_2| \le 10$ where κ is the universal constant of Lemma 4.3.5.

Proof. In a first step, we prove that $z \mapsto \langle \mathrm{Im}\,\boldsymbol{m}(z)\rangle$ is locally Hölder-continuous. Taking the derivative of (4.3.6) with respect to $z \in \mathbb{H}$ yields

$$(1 - \boldsymbol{m}^2(z)\boldsymbol{S})\partial_z \boldsymbol{m}(z) = \boldsymbol{m}(z)^2.$$

By using that $\partial_z \phi = \mathrm{i}\,2\partial_z \mathrm{Im}\,\phi$ for every analytic function ϕ and taking the average, we get

$$\mathrm{i}\,2\partial_z \langle \mathrm{Im}\,\boldsymbol{m}\rangle = \langle |\boldsymbol{m}|, \boldsymbol{B}^{-1}|\boldsymbol{m}|\rangle.$$

Here, we suppressed the z-dependence of $\boldsymbol{B}^{-1}$. We apply Cauchy-Schwarz inequality and use (4.3.7), (4.3.24) and (4.3.13a) to obtain

$$|\partial_z \langle \operatorname{Im} \boldsymbol{m} \rangle| \leq \|\boldsymbol{m}\|_2 \|\boldsymbol{B}^{-1}\|_{2\to 2} \|\boldsymbol{m}\|_2 \lesssim_\delta \min\{(\operatorname{Re} z)^{-2} \langle \operatorname{Im} \boldsymbol{m} \rangle^{-\kappa}, (\operatorname{Im} z)^{-1}\} \lesssim_\delta \langle \operatorname{Im} \boldsymbol{m} \rangle^{-\kappa}$$

for all $z \in \mathbb{H}$ satisfying $\delta \leq |z| \leq 10$. This implies that $z \mapsto \langle \operatorname{Im} \boldsymbol{m}(z) \rangle$ is Hölder-continuous with Hölder-exponent $1/(\kappa+1)$ on $z \in \mathbb{H}$ satisfying $\delta \leq |z| \leq 10$. Moreover, it has a unique continuous extension to $I_\delta := \{\tau \in \mathbb{R}; \delta/3 \leq |\tau| \leq 10\}$. Multiplying this continuous function on I_δ by π^{-1} yields a Lebesgue-density of the measure ρ (cf. (4.3.8)) restricted to I_δ.

We conclude that the Stieltjes transform $\langle \boldsymbol{m} \rangle$ has the same regularity by decomposing ρ into a measure supported around zero and a measure supported away from zero and using Lemma A.7 in [4]. □

For estimating the difference between the solution $\boldsymbol{m}$ of the QVE and a solution $\boldsymbol{g}$ of the perturbed QVE (4.3.9), we introduce the deterministic control parameter

$$\vartheta(z) := \langle \operatorname{Im} \boldsymbol{m}(z) \rangle + \operatorname{dist}(z, \operatorname{supp} \rho), \quad z \in \mathbb{H}.$$

Lemma 4.3.9 (Stability of the QVE)**.** *Let $\delta \gtrsim 1$. Suppose there are some functions $\boldsymbol{d}\colon \mathbb{H} \to \mathbb{C}^{p+n}$ and $\boldsymbol{g}\colon \mathbb{H} \to (\mathbb{C}\backslash\{0\})^{n+p}$ satisfying* (4.3.9)*. Then there exist universal constants $\kappa_1, \kappa_2 \in \mathbb{N}$ and a function $\lambda_*\colon \mathbb{H} \to (0,\infty)$, independent of n and p, such that $\lambda_*(10\mathrm{i}) \geq 1/5$, $\lambda_*(z) \gtrsim_\delta \vartheta(z)^{\kappa_1}$ and*

$$\|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty \mathbf{1}\Big(\|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty \leq \lambda_*(z)\Big) \lesssim_\delta \vartheta(z)^{-\kappa_2} \|\boldsymbol{d}(z)\|_\infty \tag{4.3.35}$$

for all $z \in \mathbb{H}$ satisfying $\delta \leq |z| \leq 10$. Moreover, there are a universal constant $\kappa_3 \in \mathbb{N}$ and a matrix-valued function $\boldsymbol{T}\colon \mathbb{H} \to \mathbb{C}^{(p+n)\times(p+n)}$, depending only on S and satisfying $\|\boldsymbol{T}(z)\|_{\infty\to\infty} \lesssim 1$, such that

$$\begin{aligned} &|\langle \boldsymbol{w}, \boldsymbol{g}(z) - \boldsymbol{m}(z) \rangle| \cdot \mathbf{1}\Big(\|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty \leq \lambda_*(z)\Big) \\ &\qquad \lesssim_\delta \vartheta(z)^{-\kappa_3} \Big(\|\boldsymbol{w}\|_\infty \|\boldsymbol{d}(z)\|_\infty^2 + |\langle \boldsymbol{T}(z)\boldsymbol{w}, \boldsymbol{d}(z)\rangle|\Big) \end{aligned} \tag{4.3.36}$$

for all $\boldsymbol{w} \in \mathbb{C}^{p+n}$ and $z \in \mathbb{H}$ satisfying $\delta \leq |z| \leq 10$.

Proof. We set $\Phi(z) := \max\{1, \|\boldsymbol{m}(z)\|_\infty\}$, $\Psi(z) := \max\{1, \|\boldsymbol{B}^{-1}(z)\|_\infty\}$ and $\lambda_*(z) := (2\Phi\Psi)^{-1}$. As $\Phi(z) \leq \max\{1, (\operatorname{Im} z)^{-1}\}$ and $\|\boldsymbol{B}^{-1}(z)\|_\infty \leq (1 - \|\boldsymbol{F}(z)\|_\infty)^{-1} \leq (1 - (\operatorname{Im} z)^{-2})^{-1}$ due to $\|\boldsymbol{m}(z)\|_\infty \leq (\operatorname{Im} z)^{-1}$ we obtain $\lambda_*(10\mathrm{i}) \geq 1/5$. Since $\delta \leq |z|$ we obtain $\langle \operatorname{Im} \boldsymbol{m}(z)\rangle^{-1} \gtrsim_\delta 1$ by (4.3.13a). Thus, for $z \in \mathbb{H}$ satisfying $\delta \leq |z| \leq 10$ the first estimate in (4.3.12a), the last estimate in (4.3.13a) and (4.3.25) yield

$$\Phi \lesssim_\delta \vartheta^{-1}, \quad \Psi \lesssim_\delta \vartheta^{-\kappa-2},$$

where κ is the universal constant from Lemma 4.3.5. Therefore, $\lambda_*(z) \gtrsim_\delta \vartheta(z)^{\kappa+3}$ and Lemma 5.11 in [4] yield the assertion as $\|\boldsymbol{w}\|_1 = (p+n)^{-1} \sum_x |\boldsymbol{w}_x| \leq \|\boldsymbol{w}\|_\infty$. □

4.3.4. Proof of Theorem 4.2.2.

Proof of Theorem 4.2.2. We start by proving the existence of the solution m of (4.2.1). Let $\boldsymbol{m} = (m_1, m_2)^t$ be the solution of (4.3.6) satisfying $\operatorname{Im} \boldsymbol{m}(z) > 0$ for $z \in \mathbb{H}$. For $\zeta \in \mathbb{H}$, we set $m(\zeta) := m_1(\sqrt{\zeta})/\sqrt{\zeta}$. Then it is straightforward to check that m satisfies (4.2.1) by solving (4.3.5b) for m_2 and plugging the result into (4.3.5a). Note that $\operatorname{Im} m(\zeta) > 0$ for all $\zeta \in \mathbb{H}$ since $m_{1,i}$ is the Stieltjes transform of a symmetric measure on $\mathbb{R}$ (cf. the explanation before (4.3.7) for the symmetry of this measure).

Next, we show the uniqueness of the solution m of (4.2.1) with $\operatorname{Im} m(\zeta) > 0$ for $\zeta \in \mathbb{H}$ which is a consequence of the uniqueness of the solution of (4.3.6). Therefore, we set $\widetilde{m}_1(\zeta) := m(\zeta)$, $\widetilde{m}_2(\zeta) := -1/(\zeta(1 + S^t\widetilde{m}_1(\zeta)))$ and $\widetilde{\boldsymbol{m}}(\zeta) := (\widetilde{m}_1(\zeta), \widetilde{m}_2(\zeta))^t$ for $\zeta \in \mathbb{H}$. From (4.2.1), we see that

$$|\widetilde{m}_1| = \frac{1}{\left|\zeta - S\frac{1}{1+S^t\widetilde{m}_1}\right|} \leq \frac{1}{\operatorname{Im}\zeta + S\frac{1}{|1+S^t\widetilde{m}_1|}S^t\operatorname{Im}\widetilde{m}_1} \leq \frac{1}{\operatorname{Im}\zeta} \tag{4.3.37}$$

for all $\zeta \in \mathbb{H}$. Since $\widetilde{m}_2$ satisfies

$$-\frac{1}{\widetilde{m}_2(\zeta)} = \zeta + S^t\frac{1}{1+S\widetilde{m}_2}(\zeta) \tag{4.3.38}$$

for $\zeta \in \mathbb{H}$, a similar argument yields $|\widetilde{m}_2| \leq (\operatorname{Im}\zeta)^{-1}$. Combining these two estimates, we obtain $|\widetilde{\boldsymbol{m}}(\zeta)| \leq (\operatorname{Im}\zeta)^{-1}$ for all $\zeta \in \mathbb{H}$. Therefore, multiplying (4.2.1) and (4.3.38) with $\widetilde{m}_1$ and $\widetilde{m}_2$, respectively, yields

$$|1+\mathrm{i}\,\xi\widetilde{\boldsymbol{m}}_x(\mathrm{i}\,\xi)| \leq \|\widetilde{\boldsymbol{m}}(\mathrm{i}\,\xi)\|_\infty \frac{1}{1-\|\widetilde{\boldsymbol{m}}(\mathrm{i}\,\xi)\|_\infty} \leq \frac{1}{\xi-1} \to 0$$

for $\xi \to \infty$ and $x = 1, \ldots, n+p$ where we used $|\widetilde{\boldsymbol{m}}(\zeta)| \leq (\operatorname{Im}\zeta)^{-1}$ in the last but one step. Thus, $\widetilde{\boldsymbol{m}}_x$ is the Stieltjes transform of a probability measure ν_x on $\mathbb{R}$ for all $x = 1, \ldots, n+p$. Multiplying (4.2.1) by $\widetilde{m}_1$, taking the imaginary part and averaging at $\zeta = \chi + \mathrm{i}\,\xi$, for $\chi \in \mathbb{R}$ and $\xi > 0$, yields

$$\begin{aligned}
\chi\langle \operatorname{Im}\widetilde{m}_1\rangle + \xi\langle \operatorname{Re}\widetilde{m}_1\rangle = & -\left\langle \operatorname{Re}\widetilde{m}_1 \,,\, S\frac{1}{|1+S^t\widetilde{m}_1|^2} S^t \operatorname{Im}\widetilde{m}_1 \right\rangle \\
& + \left\langle \operatorname{Im}\widetilde{m}_1 \,,\, S\frac{1}{|1+S^t\widetilde{m}_1|^2} (1+S^t\operatorname{Re}\widetilde{m}_1) \right\rangle \\
= & \left\langle \operatorname{Im}\widetilde{m}_1 \,,\, S\frac{1}{|1+S^t\widetilde{m}_1|^2} \right\rangle \geq 0,
\end{aligned} \tag{4.3.39}$$

where we used the definition of the transposed matrix and the symmetry of the scalar product in the last step. On the other hand, we have

$$\chi\langle \operatorname{Im}\widetilde{m}_1\rangle + \xi\langle \operatorname{Re}\widetilde{m}_1\rangle = \int_{\mathbb{R}} \frac{\xi t}{(t-\chi)^2+\xi^2}\nu(\mathrm{d}t).$$

Assuming that there is a $\chi < 0$ such that $\chi \in \operatorname{supp}\nu$ we obtain that $\chi\langle \operatorname{Im}\widetilde{m}_1\rangle + \xi\langle \operatorname{Re}\widetilde{m}_1\rangle < 0$ for $\xi \downarrow 0$ which contradicts (4.3.39). Therefore $\operatorname{supp}\nu_x \subset [0,\infty)$ for $x = 1, \ldots, p$.

Together with a similar argument for $\widetilde{m}_2$, we get that $\operatorname{supp}\nu_x \subset [0,\infty)$ for all $x = 1, \ldots, n+p$. In particular, we can assume that $\widetilde{\boldsymbol{m}}$ is defined on $\mathbb{C} \setminus [0,\infty)$. We set $m_1(z) := z\widetilde{m}_1(z^2)$, $m_2(z) := z\widetilde{m}_2(z^2)$ and $\boldsymbol{m}(z) := (m_1(z), m_2(z))^t$ for all $z \in \mathbb{H}$. Hence, we get

$$\operatorname{Im}\boldsymbol{m}_x(\tau+\mathrm{i}\,\eta) = \eta\int_{[0,\infty)} \frac{t+\tau^2+\eta^2}{(t-\tau^2+\eta^2)^2+4\eta^2\tau^2}\nu_x(\mathrm{d}t)$$

as $\operatorname{supp}\nu_x \subset [0,\infty)$. This implies $\operatorname{Im}\boldsymbol{m}(z) > 0$ for $z \in \mathbb{H}$ and thus the uniqueness of solutions of (4.3.6) with positive imaginary part implies the uniqueness of $\widetilde{m}_1$.

Finally, we verify the claim about the structure of the probability measure representing $\langle m\rangle$. By Lemma 4.3.8 and the statements following (4.3.6), $\langle m_1\rangle$ is the Stieltjes transform of $\pi_*\delta_0 + \rho_1(\omega)\mathrm{d}\omega$ for some $\pi_* \in [0,1]$ and some symmetric Hölder-continuous function

$\rho_1 \colon \mathbb{R} \setminus \{0\} \to [0, \infty)$ whose support is contained in $[-2, 2]$. Therefore, m is the Stieltjes transform of $\nu(\mathrm{d}\omega) := \pi_* \delta_0(\mathrm{d}\omega) + \pi(\omega)\mathbf{1}(\omega > 0)\mathrm{d}\omega$ where $\pi(\omega) = \omega^{-1/2}\rho_1(\omega^{1/2})$ for $\omega > 0$. Thus, the support of ν is contained in $[0, 4]$. □

4.3.5. Square Gram matrices. In this subsection, we study the stability of (4.3.6) for $n = p$. Here, we assume (A), (E1) and (F1). These assumptions are strictly stronger than (A), (B) and (D) (cf. Remark 4.2.7).

For the following arguments, it is important that $\boldsymbol{m}$ is purely imaginary for $\operatorname{Re} z = 0$ as $\boldsymbol{m}(-\bar{z}) = -\overline{\boldsymbol{m}(z)}$ for all $z \in \mathbb{H}$. If we set

$$\boldsymbol{v}(z) = \operatorname{Im} \boldsymbol{m}(z) \tag{4.3.40}$$

for $z \in \mathbb{H}$, then $\boldsymbol{v}$ fulfills

$$\frac{1}{\boldsymbol{v}(\mathrm{i}\eta)} = \eta + \boldsymbol{S}\boldsymbol{v}(\mathrm{i}\eta) \tag{4.3.41}$$

for all $\eta \in (0, \infty)$ due to (4.3.6). The study of this equation will imply the stability of the QVE at $z = 0$. The following proposition is the main result of this subsection.

Proposition 4.3.10. *Let $n = p$, i.e., (E1) holds true, and S satisfies (A) as well as (F1).*

(i) There exists a $\widehat{\delta} \sim 1$ such that $|\boldsymbol{m}(z)| \sim 1$ uniformly for all $z \in \mathbb{H}$ satisfying $|z| \leq 10$ and $\operatorname{Re} z \in [-\widehat{\delta}, \widehat{\delta}]$. Moreover, $\langle \operatorname{Im} \boldsymbol{m}(z) \rangle \gtrsim 1$ for all $z \in \mathbb{H}$ satisfying $|z| \leq 10$ and $\operatorname{Re} z \in [-\widehat{\delta}, \widehat{\delta}]$ and there is a $\boldsymbol{v}(0) = (v_1(0), v_2(0))^t \in \mathbb{R}^p \oplus \mathbb{R}^p$ such that $\boldsymbol{v}(0) \sim 1$ and

$$\mathrm{i}\boldsymbol{v}(0) = \lim_{\eta \downarrow 0} \boldsymbol{m}(\mathrm{i}\eta).$$

(ii) (Stability of the QVE at $z = 0$) Suppose that some functions $\boldsymbol{d} = (d_1, d_2)^t \colon \mathbb{H} \to \mathbb{C}^{p+p}$ and $\boldsymbol{g} = (g_1, g_2)^t \colon \mathbb{H} \to (\mathbb{C} \setminus \{0\})^{p+p}$ satisfy (4.3.9) and

$$\langle g_1(z) \rangle = \langle g_2(z) \rangle \tag{4.3.42}$$

for all $z \in \mathbb{H}$. There are numbers $\lambda_*, \widehat{\delta} \gtrsim 1$, depending only on S, such that

$$\|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty \mathbf{1}\Big(\|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty \leq \lambda_*\Big) \lesssim \|\boldsymbol{d}(z)\|_\infty \tag{4.3.43}$$

for all $z \in \mathbb{H}$ *satisfying* $|z| \le 10$ *and* $\operatorname{Re} z \in [-\widehat{\delta}, \widehat{\delta}]$. *Moreover, there is a matrix-valued function* $\boldsymbol{T}\colon \mathbb{H} \to \mathbb{C}^{2p\times 2p}$, *depending only on* S *and satisfying* $\|\boldsymbol{T}(z)\|_\infty \lesssim 1$, *such that*

$$|\langle \boldsymbol{w}, \boldsymbol{g}(z) - \boldsymbol{m}(z)\rangle| \cdot \mathbf{1}\Big(\|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty \le \lambda_*\Big) \\ \lesssim \|\boldsymbol{w}\|_\infty \|\boldsymbol{d}(z)\|_\infty^2 + |\langle \boldsymbol{T}(z)\boldsymbol{w}, \boldsymbol{d}(z)\rangle| \tag{4.3.44}$$

for all $\boldsymbol{w} \in \mathbb{C}^{2p}$ *and* $z \in \mathbb{H}$ *satisfying* $|z| \le 10$ *and* $\operatorname{Re} z \in [-\widehat{\delta}, \widehat{\delta}]$.

The remainder of this subsection will be devoted to the proof of this proposition. Therefore, we will always assume that (A), (E1) and (F1) are satisfied.

Lemma 4.3.11. *The function* $\boldsymbol{v}\colon \mathrm{i}(0,\infty) \to \mathbb{R}^{2p}$ *defined in* (4.3.40) *satisfies*

$$1 \lesssim \inf_{\eta \in (0,10]} \boldsymbol{v}(\mathrm{i}\eta) \le \sup_{\eta>0} \|\boldsymbol{v}(\mathrm{i}\eta)\|_\infty \lesssim 1. \tag{4.3.45}$$

If we write $\boldsymbol{v} = (v_1, v_2)^t$ *for* $v_1, v_2\colon \mathrm{i}(0,\infty) \to \mathbb{R}^p$, *then*

$$\langle v_1(\mathrm{i}\eta)\rangle = \langle v_2(\mathrm{i}\eta)\rangle \tag{4.3.46}$$

for all $\eta \in (0,\infty)$.

The estimate in (4.3.45), with some minor modifications which we will explain next, is shown as in the proof of (6.30) of [4].

Proof. From (4.3.41) and the definition of $\boldsymbol{S}$, we obtain $\eta\langle v_1\rangle - \eta\langle v_2\rangle = \langle v_1, S v_2\rangle - \langle v_2, S^t v_1\rangle = 0$ for all $\eta \in (0,\infty)$ which proves (4.3.46). Differing from [4], the discrete functional $\widetilde{J}$ is defined as follows:

$$\widetilde{J}(u) = \frac{\varphi}{2K} \sum_{i,j=1}^{2K} u(i)\mathcal{Z}_{ij}u(j) - \sum_{i=1}^{2K} \log u(i) \tag{4.3.47}$$

for $u \in (0,\infty)^{2K}$ (we used the notation $u(i)$ to denote the i-th entry of u) where $\mathcal{Z}$ is the $2K \times 2K$ matrix with entries in $\{0,1\}$ defined by

$$\mathcal{Z} = \begin{pmatrix} 0 & Z \\ Z^t & 0 \end{pmatrix}. \tag{4.3.48}$$

Decomposing $u = (u_1, u_2)^t$ for $u_1, u_2 \in (0,\infty)^K$ and writing $u_1(i) = u(i)$ and $u_2(j) = u(K + j)$ for their entries we obtain

$$\widetilde{J}(u) = \frac{\varphi}{K}\sum_{i,j=1}^{K} u_1(i) Z_{ij} u_2(j) - \sum_{i=1}^{K} (\log u_1(i) + \log u_2(i)). \tag{4.3.49}$$

Lemma 4.3.12. *If $\Psi < \infty$ is a constant such that $u = (u_1, u_2)^t \in (0,\infty)^K \times (0,\infty)^K$ satisfies*

$$\widetilde{J}(u) \leq \Psi,$$

where $\widetilde{J}$ is defined in (4.3.47)*, and $\langle u_1 \rangle = \langle u_2 \rangle$, then there is a constant $\Phi < \infty$ depending only on (Ψ, φ, K) such that*

$$\max_{k=1}^{2K} u(k) \leq \Phi.$$

Proof. We define $\widetilde{Z}_{ij} := Z_{i\sigma(j)}$ where σ is a permutation of $\{1, \ldots, K\}$ such that $\widetilde{Z}_{ii} = 1$ for all $i = 1, \ldots, K$ where we use the FID property of Z. Moreover, we set $M_{ij} := u_1(i)\widetilde{Z}_{ij}u_2(\sigma(j))$ and follow the proof of Lemma 6.10 in [4] to obtain

$$u_1(i)u_2(\sigma(j)) \lesssim (M^{K-1})_{ij} \lesssim 1$$

for all $i, j = 1, \ldots, K$. Averaging over i and j yields

$$\langle u_1 \rangle^2 = \langle u_2 \rangle^2 \lesssim 1$$

where we used $\langle u_1 \rangle = \langle u_2 \rangle$. This concludes the proof of Lemma 4.3.12. □

Recalling the function $\boldsymbol{v}$ in Lemma 4.3.11, we set $u = (\langle \boldsymbol{v} \rangle_1, \ldots, \langle \boldsymbol{v} \rangle_{2K})$ with $\langle \boldsymbol{v} \rangle_i = Kp^{-1}\sum_{x\in I_i} \boldsymbol{v}_x$, where $I_i := p + I_{i-K}$ for $i \geq K + 1$. Then we have $\langle u_1 \rangle = \langle u_2 \rangle$ by (4.3.46) and since $I_1, \ldots, I_{2K}$ is an equally sized partition of $\{1, \ldots, 2p\}$. Therefore, the assumptions of Lemma 4.3.12 are met which implies (4.3.45) of Lemma 4.3.11 as in [4]. □

We recall from Lemma 4.3.4 that $\boldsymbol{f} = (f_1, f_2)^t$ is the unique nonnegative, normalized eigenvector of $\boldsymbol{F}$ corresponding to the eigenvalue $\|\boldsymbol{F}\|_2$. Moreover, we define $\boldsymbol{f}_- := (f_1, -f_2)^t$ which clearly satisfies

$$\boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2 \boldsymbol{f}_-. \tag{4.3.50}$$

Since the spectrum of $\boldsymbol{F}$ is symmetric, $\operatorname{Spec}(\boldsymbol{F}) = -\operatorname{Spec}(\boldsymbol{F})$ with multiplicities, and $\|\boldsymbol{F}\|_2$ is a simple eigenvalue of $\boldsymbol{F}$, the same is true for the eigenvalue $-\|\boldsymbol{F}\|_2$ of $\boldsymbol{F}$ and $\boldsymbol{f}_-$ spans its associated eigenspace. We introduce

$$\boldsymbol{e}_- := \begin{pmatrix} 1 \\ -1 \end{pmatrix} \in \mathbb{C}^p \oplus \mathbb{C}^p. \tag{4.3.51}$$

Lemma 4.3.13. *For $\eta \in (0,\infty)$, the derivative of $\boldsymbol{m}$ satisfies*

$$\boldsymbol{m}'(\mathrm{i}\eta) = \frac{\mathrm{d}}{\mathrm{d}z}\boldsymbol{m}(\mathrm{i}\eta) = -\boldsymbol{v}(\mathrm{i}\eta)(1+\boldsymbol{F}(\mathrm{i}\eta))^{-1}\boldsymbol{v}(\mathrm{i}\eta). \tag{4.3.52}$$

Moreover, $|\boldsymbol{m}'(\mathrm{i}\eta)| \lesssim 1$ uniformly for $\eta \in (0,10]$.

Proof. In the whole proof, the quantities $\boldsymbol{v}$, $\boldsymbol{f}$, $\boldsymbol{f}_-$ and $\boldsymbol{F}$ are evaluated at $z = \mathrm{i}\eta$ for $\eta > 0$. Therefore, we will mostly suppress the z-dependence of all quantities. Differentiating (4.3.6) with respect to z and using (4.3.40) yields

$$-(1+\boldsymbol{F})\frac{\boldsymbol{m}'}{\boldsymbol{v}} = \boldsymbol{v}.$$

As $\|\boldsymbol{F}\|_2 < 1$ by (4.3.21), the matrix $(1+\boldsymbol{F})$ is invertible which yields (4.3.52) for all $\eta \in (0,\infty)$.

In order to prove $|\boldsymbol{m}'(\mathrm{i}\eta)| \lesssim 1$ uniformly for $\eta \in (0,\infty)$, we first prove that

$$|\langle \boldsymbol{f}_-(\mathrm{i}\eta)\boldsymbol{v}(\mathrm{i}\eta)\rangle| \le \mathcal{O}(\eta). \tag{4.3.53}$$

We define the auxiliary operator $\boldsymbol{A} := \|\boldsymbol{F}\|_2 + \boldsymbol{F} = 1 + \boldsymbol{F} - \eta\frac{\langle \boldsymbol{f}\boldsymbol{v}\rangle}{\langle \boldsymbol{f}\rangle}$ where we used (4.3.21) and (4.3.40). Note that

$$\boldsymbol{A}\boldsymbol{f}_- = 0, \quad \boldsymbol{A}\boldsymbol{e}_- = \boldsymbol{e}_- + \boldsymbol{F}\boldsymbol{e}_- - \eta\frac{\langle \boldsymbol{f}\boldsymbol{v}\rangle}{\langle \boldsymbol{f}\rangle}\boldsymbol{e}_- = \mathcal{O}(\eta), \tag{4.3.54}$$

where we used $\boldsymbol{F}\boldsymbol{e}_- = -\boldsymbol{e}_- + \eta(v_1, -v_2)^t$ which follows from (4.3.6) and the definition of $\boldsymbol{F}$.

Defining $\boldsymbol{Q}\boldsymbol{u} := \boldsymbol{u} - \langle \boldsymbol{f}_-\boldsymbol{u}\rangle\boldsymbol{f}_-$ for $\boldsymbol{u} \in \mathbb{C}^{2p}$ and decomposing

$$\boldsymbol{e}_- = \left\langle \boldsymbol{f}_-\boldsymbol{e}_-\right\rangle \boldsymbol{f}_- + \boldsymbol{Q}\boldsymbol{e}_-$$

yield $\boldsymbol{AQe}_- = \mathcal{O}(\eta)$ because of (4.3.54). As $|\boldsymbol{m}(\mathrm{i}\eta)| \sim 1$ by (4.3.45) for $\eta \in (0, 10]$ the bound (4.3.20) in Lemma 4.3.3 implies that there is an $\varepsilon \sim 1$ such that for all $\eta \in (0, 10]$ we have

$$\operatorname{Spec}(\boldsymbol{F}) \subset \{-\|\boldsymbol{F}\|_2\} \cup [-\|\boldsymbol{F}\|_2 + \varepsilon, \|\boldsymbol{F}\|_2 - \varepsilon] \cup \{\|\boldsymbol{F}\|_2\}. \tag{4.3.55}$$

Since $-\|\boldsymbol{F}\|_2$ is a simple eigenvalue of $\boldsymbol{F}$ and (4.3.50) the symmetric matrix $\boldsymbol{A} = \|\boldsymbol{F}\|_2 + \boldsymbol{F}$ is invertible on $\boldsymbol{f}_-^\perp$ and $\left\|(\boldsymbol{A}|_{\boldsymbol{f}_-^\perp})^{-1}\right\|_2 = \varepsilon^{-1} \sim 1$. As $\boldsymbol{f}_- \perp \boldsymbol{Qe}_-$ we conclude $\boldsymbol{Qe}_- = \mathcal{O}(\eta)$ and hence

$$(1 - \langle \boldsymbol{f} \rangle)(1 + \langle \boldsymbol{f} \rangle) = 1 - \langle \boldsymbol{f} \rangle^2 = 1 - \left\langle \boldsymbol{f}_- \boldsymbol{e}_- \right\rangle^2 = \|\boldsymbol{Qe}_-\|_2^2 = \mathcal{O}(\eta^2). \tag{4.3.56}$$

Thus, using (4.3.46) and (4.3.56), this implies

$$|\langle \boldsymbol{f}_-(\mathrm{i}\eta)\boldsymbol{v}(\mathrm{i}\eta)\rangle| = \left|\langle \boldsymbol{v}\boldsymbol{e}_-\rangle + \left\langle \boldsymbol{v}\left[\boldsymbol{f}_- - \boldsymbol{e}_-\right]\right\rangle\right| \lesssim \left\|\boldsymbol{f}_- - \boldsymbol{e}_-\right\|_2 = \sqrt{2(1 - \langle \boldsymbol{f} \rangle)} = \mathcal{O}(\eta),$$

which concludes the proof of (4.3.53).

In (4.3.52), we decompose $\boldsymbol{v} = \langle \boldsymbol{f}_- \boldsymbol{v}\rangle \boldsymbol{f}_- + \boldsymbol{Qv}$ and, using $\boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2 \boldsymbol{f}_-$ and (4.3.21), we obtain

$$\boldsymbol{m}' = -\boldsymbol{v}\frac{\langle \boldsymbol{f}_- \boldsymbol{v}\rangle}{\eta}\frac{\langle \boldsymbol{f}\rangle}{\langle \boldsymbol{f}\boldsymbol{v}\rangle}\boldsymbol{f}_- - \boldsymbol{v}(1 + \boldsymbol{F})^{-1}\boldsymbol{Qv}.$$

Using (4.3.55), we see that $\|(1 + \boldsymbol{F})^{-1}\boldsymbol{Qv}\|_2 \sim 1$ uniformly for $\eta \in (0, 10]$. Together with $\langle \boldsymbol{f}_-(\mathrm{i}\eta)\boldsymbol{v}(\mathrm{i}\eta)\rangle = \mathcal{O}(\eta)$ by (4.3.53), this yields $|\boldsymbol{m}'(\mathrm{i}\eta)| \lesssim 1$ uniformly for $\eta \in (0, 10]$. $\square$

The previous lemma, (4.3.41) and Lemma 4.3.11 imply that $\boldsymbol{v}(0) := \lim_{\eta \downarrow 0} \boldsymbol{v}(\mathrm{i}\eta)$ exists and satisfies

$$\boldsymbol{v}(0) \sim 1, \quad 1 = \boldsymbol{v}(0)\boldsymbol{Sv}(0) = \boldsymbol{F}(0)1, \quad \langle v_1(0)\rangle = \langle v_2(0)\rangle, \tag{4.3.57}$$

where $\boldsymbol{v}(0) = (v_1(0), v_2(0))^t$.

In the next lemma, we establish an expansion of $\boldsymbol{m}(z)$ on the upper half-plane around $z = 0$. The proof of this result and later the stability estimates on $\boldsymbol{g} - \boldsymbol{m}$ will be a consequence of the equation

$$\boldsymbol{Bu} = \mathrm{e}^{-\mathrm{i}\boldsymbol{\psi}}\boldsymbol{uFu} + \mathrm{e}^{-\mathrm{i}\boldsymbol{\psi}}\boldsymbol{gd} \tag{4.3.58}$$

where $\boldsymbol{u} = (\boldsymbol{g}-\boldsymbol{m})/|\boldsymbol{m}|$ and $\mathrm{e}^{\mathrm{i}\boldsymbol{\psi}} = \boldsymbol{m}/|\boldsymbol{m}|$ with $\boldsymbol{\psi} \in \mathbb{R}^{2p}$. This quadratic equation in $\boldsymbol{u}$ was derived in Lemma 5.8 in [4].

Lemma 4.3.14. *For $z \in \mathbb{H}$, we have*

$$\boldsymbol{m}(z) = \mathrm{i}\boldsymbol{v}(0) - z\boldsymbol{v}(0)(1+\boldsymbol{F}(0))^{-1}\boldsymbol{v}(0) + \mathcal{O}(|z|^2), \tag{4.3.59a}$$

$$\frac{\boldsymbol{m}(z)}{|\boldsymbol{m}(z)|} = \mathrm{i} - (\operatorname{Re} z)(1+\boldsymbol{F}(0))^{-1}\boldsymbol{v}(0) + \mathcal{O}(|z|^2). \tag{4.3.59b}$$

In particular, there is a $\widehat{\delta} \sim 1$ such that $|\boldsymbol{m}(z)| \sim 1$ uniformly for $z \in \mathbb{H}$ satisfying $\operatorname{Re} z \in [-\widehat{\delta}, \widehat{\delta}]$ *and $|z| \leq 10$. Moreover,*

$$\|\boldsymbol{f}(z) - 1\|_\infty = \mathcal{O}(|z|), \quad \left\|\boldsymbol{f}_-(z) - \boldsymbol{e}_-\right\|_\infty = \mathcal{O}(|z|). \tag{4.3.60}$$

Proof. In order to prove (4.3.59a), we consider (4.3.6) at z as a perturbation of (4.3.6) at $z=0$ perturbed by $\boldsymbol{d} = z$ in the notation of (4.3.9). The solution of the unperturbed equation is $\boldsymbol{m} = \mathrm{i}\boldsymbol{v}(0)$. Following the notation of (4.3.9), we find that (4.3.58) holds with $\boldsymbol{g} = \boldsymbol{m}(z)$ and $\boldsymbol{u}(z) = (\boldsymbol{m}(z) - \mathrm{i}\boldsymbol{v}(0))/\boldsymbol{v}(0)$. We write $\boldsymbol{u}(z) = \theta(z)\boldsymbol{e}_- + \boldsymbol{w}(z)$ with $\boldsymbol{w} \perp \boldsymbol{e}_-$. (We will suppress the z-dependence in our notation.) Plugging this into (4.3.58) and projecting onto $\boldsymbol{e}_-$ yields

$$\theta\langle \boldsymbol{v}(0)\rangle = -\left\langle \boldsymbol{e}_-\boldsymbol{v}(0)\boldsymbol{w}\right\rangle, \tag{4.3.61}$$

where we used that $\boldsymbol{F}(0)1 = 1$, i.e., $\langle \boldsymbol{F}(0)\boldsymbol{w}\rangle = \langle \boldsymbol{w}\rangle$, $\langle \boldsymbol{e}_-\boldsymbol{w}\boldsymbol{F}(0)\boldsymbol{w}\rangle = 0$ and $\langle v_1(0)\rangle = \langle v_2(0)\rangle$. Thus, we have $\theta = \mathcal{O}(\|\boldsymbol{w}\|_\infty)$ because of (4.3.57), so that we conclude $-(1+\boldsymbol{F}(0))\boldsymbol{w} = z\boldsymbol{v}(0) + \mathcal{O}(\|\boldsymbol{w}\|_\infty^2 + |z|\|\boldsymbol{w}\|_\infty)$. As $\boldsymbol{w}$, $(1+\boldsymbol{F}(0))\boldsymbol{w}$ and $\boldsymbol{v}(0)$ are orthogonal to $\boldsymbol{e}_-$, the error term is also orthogonal to it which implies

$$\boldsymbol{w} = -z(1+\boldsymbol{F}(0))^{-1}\boldsymbol{v}(0) + \mathcal{O}(|z|^2) \tag{4.3.62}$$

using that $(1+\boldsymbol{F}(0))^{-1}$ is bounded on $\boldsymbol{e}_-^\perp$.

Observing that $\langle m_1(z)\rangle = \langle m_2(z)\rangle$ for $z \in \mathbb{H}$ by (4.3.6) and differentiating this relation yields $\langle \boldsymbol{m}'(\mathrm{i}\eta)\boldsymbol{e}_-\rangle = 0$ for all $\eta \in (0,\infty)$. Hence,

$$\langle \boldsymbol{e}_-\boldsymbol{v}(0)(1+\boldsymbol{F}(0))^{-1}\boldsymbol{v}(0)\rangle = -\lim_{\eta\downarrow 0}\langle \boldsymbol{e}_-\boldsymbol{m}'(\mathrm{i}\eta)\rangle = 0 \tag{4.3.63}$$

by Lemma 4.3.13.

Plugging (4.3.62) into (4.3.61), we obtain

$$\theta\langle \boldsymbol{v}(0)\rangle = \langle \boldsymbol{e}_- \boldsymbol{v}(0)(1+\boldsymbol{F}(0))^{-1}\boldsymbol{v}(0)\rangle + \mathcal{O}(|z|^2) = \mathcal{O}(|z|^2),$$

where we used (4.3.63). Hence, $\boldsymbol{m}(z) = \boldsymbol{v}(0)(\boldsymbol{u} + \mathrm{i}\boldsymbol{v}(0))$ concludes the proof of (4.3.59a) which immediately implies (4.3.59b).

Using the expansion of $\boldsymbol{m}$ in (4.3.59a) in a similar argument as in the proof of $\|\boldsymbol{f}_-(\mathrm{i}\eta) - \boldsymbol{e}_-\|_2 = \mathcal{O}(\eta)$ in Lemma 4.3.13 yields

$$\|\boldsymbol{f}(z) - 1\|_2 = \|\boldsymbol{f}_-(z) - \boldsymbol{e}_-\|_2 = \mathcal{O}(|z|).$$

Similarly, using (4.3.27), we obtain (4.3.60). □

By a standard argument from perturbation theory and possibly reducing $\widehat{\delta} \sim 1$, we can assume that $\boldsymbol{B}(z)$ has a unique eigenvalue $\beta(z)$ of smallest modulus for $z \in \mathbb{H}$ satisfying $|\mathrm{Re}\, z| \le \widehat{\delta}$ and $|z| \le 10$ such that $|\beta'| - |\beta| \gtrsim 1$ for $\beta' \in \mathrm{Spec}(\boldsymbol{B}(z))$ and $\beta' \neq \beta$. This follows from $|\boldsymbol{m}| \sim 1$ and thus $\mathrm{Gap}(F(z)F(z)^t) \gtrsim 1$ by Lemma 4.3.3. For $z \in \mathbb{H}$ satisfying $|\mathrm{Re}\, z| \le \widehat{\delta}$ and $|z| \le 10$, we therefore find a unique (unnormalized) vector $\boldsymbol{b}(z) \in \mathbb{C}^{2p}$ such that $\boldsymbol{B}(z)\boldsymbol{b}(z) = \beta(z)\boldsymbol{b}(z)$ and $\langle \boldsymbol{f}_-, \boldsymbol{b}(z)\rangle = 1$.

We introduce the spectral projection $\boldsymbol{P}$ onto the spectral subspace associated to the eigenvalue $\beta(z)$ of the operator $\boldsymbol{B}(z)$ acting on $(\mathbb{C}^{2p}, \|\cdot\|_\infty)$. We obtain the relation

$$\boldsymbol{P} = \frac{\langle \bar{\boldsymbol{b}}, \cdot\rangle}{\langle \boldsymbol{b}^2\rangle}\boldsymbol{b}.$$

Note that $\boldsymbol{P}$ is not an orthogonal projection in general. Let $\boldsymbol{Q} := 1 - \boldsymbol{P}$ denote the complementary projection onto the spectral subspace of $\boldsymbol{B}(z)$ not containing $\beta(z)$ (this $\boldsymbol{Q}$ is different from the one in the proof of Lemma 4.3.13). Since $\boldsymbol{B}(z) = -1 - \boldsymbol{F}(z) + \mathcal{O}(|z|)$ we obtain

$$\|\boldsymbol{b}(z) - \boldsymbol{e}_-\|_\infty = \|\overline{\boldsymbol{b}(z)} - \boldsymbol{e}_-\|_\infty = \mathcal{O}(|z|) \tag{4.3.64}$$

for $z \in \mathbb{H}$ satisfying $|\mathrm{Re}\, z| \le \widehat{\delta}$ and $|z| \le 10$.

Lemma 4.3.15. *By possibly reducing $\widehat{\delta}$ from Lemma 4.3.14, but still $\widehat{\delta} \gtrsim 1$, we have*

$$\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim \frac{1}{|z|}, \quad \|\boldsymbol{B}^{-1}(z)\boldsymbol{Q}\|_\infty + \|(\boldsymbol{B}^{-1}(z)\boldsymbol{Q})^*\|_\infty \lesssim 1 \tag{4.3.65}$$

for $z \in \mathbb{H}$ satisfying $|\mathrm{Re}\, z| \leq \widehat{\delta}$ and $|z| \leq 10$.

Proof. Due to $|\boldsymbol{m}(z)| \sim 1$ and using (4.3.27) with $R = \boldsymbol{F}(z)$ and $D = |\boldsymbol{m}(z)|^2/\boldsymbol{m}(z)^2$, it is enough to prove the estimates in (4.3.65) with $\|\cdot\|_\infty$ replaced by $\|\cdot\|_2$. We first remark that $|\boldsymbol{m}(z)| \sim 1$ and arguing similarly as in the proof of Lemma 4.3.4 imply $\|\boldsymbol{B}^{-1}(z)\|_2 \lesssim (\mathrm{Im}\, z)^{-1}$.

Now we prove $\|\boldsymbol{B}^{-1}(z)\|_2 \lesssim (\mathrm{Re}\, z)^{-1}$. We apply Lemma 4.3.6 and recall $U_1 = |m_1|^2/m_1^2$ and $U_2 = |m_2|^2/m_2^2$ to get

$$\begin{aligned} \mathrm{Im}\left(1 - \|F(z)^t F(z)\|_2 \left\langle \frac{f_1}{\|f_1\|_2}\,, U_1 \frac{f_1}{\|f_1\|_2} \right\rangle \left\langle \frac{f_2}{\|f_2\|_2}\,, U_2 \frac{f_2}{\|f_2\|_2} \right\rangle\right) \\ = \frac{\|F(z)^t F(z)\|_2}{\|f_1\|_2 \|f_2\|_2} \langle \boldsymbol{v}(0)\rangle \mathrm{Re}\, z + \mathcal{O}(|z|^2), \end{aligned} \tag{4.3.66}$$

where we used (4.3.59b), (4.3.60) and $\|f_1\|_2, \|f_2\|_2, \|F(z)^t F(z)\|_2 \sim 1$. Since $\boldsymbol{v}(0) \sim 1$ and $\mathrm{Gap}(F(z)F(z)^t) \gtrsim 1$ by Lemma 4.3.3 and $|\boldsymbol{m}(z)| \sim 1$, (4.3.66) and Lemma 4.3.6 yield $\|\boldsymbol{B}^{-1}(z)\|_2 \lesssim (\mathrm{Re}\, z)^{-1}$ and hence $\|\boldsymbol{B}^{-1}(z)\|_2 \lesssim \min\{(\mathrm{Im}\, z)^{-1}, (\mathrm{Re}\, z)^{-1}\} \lesssim |z|^{-1}$.

The estimate $\|\boldsymbol{B}^{-1}(z)\boldsymbol{Q}\|_\infty \lesssim 1$ in (4.3.65) follows from $\mathrm{Gap}(F(z)F(z)^t) \gtrsim 1$ by Lemma 4.3.3, $|\boldsymbol{m}(z)| \sim 1$ and a standard argument from perturbation theory as presented in Lemma 8.1 of [4]. Here, it might be necessary to reduce $\widehat{\delta}$. We remark that $\boldsymbol{B}^* = |\boldsymbol{m}|^2/\overline{\boldsymbol{m}}^2 - \boldsymbol{F}$ and similarly $\boldsymbol{P}^* = \langle \boldsymbol{b}, \cdot\rangle/\langle \overline{\boldsymbol{b}}^2\rangle \overline{\boldsymbol{b}}$, i.e., $\boldsymbol{B}^*$ and $\boldsymbol{P}^*$ emerge by the same constructions where $\boldsymbol{m}$ is replaced by $\overline{\boldsymbol{m}}$. Therefore, we obtain $\|(\boldsymbol{B}^{-1}(z)\boldsymbol{Q})^*\|_\infty \lesssim 1$. ☐

Proof of Proposition 4.3.10. The part (i) follows from the previous lemmata.

The part (ii) has already been proven for $|z| \geq \delta$ in Lemma 4.3.9 and for any $\delta \gtrsim 1$. Therefore, we restrict ourselves to $|z| \leq \delta$ for a sufficiently small $\delta \gtrsim 1$. We recall $\mathrm{e}^{\mathrm{i}\psi} = \boldsymbol{m}/|\boldsymbol{m}|$.

Owing to Lemma 4.3.14 and (4.3.64), there are positive constants $\delta, \Phi, \widehat{\Phi} \sim 1$ which only depend on the model parameters such that

$$\|\boldsymbol{m}(z)\|_\infty \leq \Phi, \qquad \|\boldsymbol{b}(z) - \boldsymbol{e}_-\|_2 \|\boldsymbol{b}\|_\infty + \left\|\mathrm{e}^{-\mathrm{i}\psi} + \mathrm{i}\right\|_\infty \|\boldsymbol{b}\|_\infty^2 \leq \widehat{\Phi} |\langle \boldsymbol{b}^2\rangle| |z| \tag{4.3.67}$$

for all $z \in \mathbb{H}$ satisfying $|z| \leq \delta$. Here, we used $\|\boldsymbol{w}\|_2 \leq \|\boldsymbol{w}\|_\infty$ for all $\boldsymbol{w} \in \mathbb{C}^{2p}$. Note that we employed (4.3.64) for estimating $\|\boldsymbol{b} - \boldsymbol{e}_-\|_2$ as well as to obtain $\|\boldsymbol{b}\|_\infty \sim 1$ and $|\langle \boldsymbol{b}^2 \rangle| \sim 1$ for all $z \in \mathbb{H}$ satisfying $|z| \leq \delta$ if $\delta \gtrsim 1$ is small enough.

Lemma 4.3.15 implies the existence of $\Psi, \widehat{\Psi} \sim 1$ such that

$$\|\boldsymbol{B}^{-1}(z)\|_\infty \leq \Psi |z|^{-1}, \qquad \|\boldsymbol{B}^{-1}(z)\boldsymbol{Q}\|_\infty \leq \widehat{\Psi} \tag{4.3.68}$$

for all $z \in \mathbb{H}$ satisfying $|z| \leq \delta$ if $1 \lesssim \delta \leq \widehat{\delta}$ is sufficiently small. With these definitions, we set

$$\lambda_* := \frac{1}{2\Phi(\Psi\widehat{\Phi} + \widehat{\Psi})}. \tag{4.3.69}$$

The estimate on $\boldsymbol{h} := \boldsymbol{g}(z) - \boldsymbol{m}(z) = \boldsymbol{u}|\boldsymbol{m}|$ will be obtained from inverting $\boldsymbol{B}$ in (4.3.58). In order to control the right-hand side of (4.3.58), we decompose it, according to $1 = \boldsymbol{P} + \boldsymbol{Q}$, as

$$\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{uFu} = \frac{\left\langle \boldsymbol{b}\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{uFu}\right\rangle}{\langle \boldsymbol{b}^2 \rangle}\boldsymbol{b} + \boldsymbol{Q}\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{uFu}, \qquad \mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{gd} = \frac{\left\langle \mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{gdb}\right\rangle}{\langle \boldsymbol{b}^2 \rangle}\boldsymbol{b} + \boldsymbol{Q}\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{gd}.$$

Clearly, as $\|\boldsymbol{S}\|_\infty \leq 1$ we have

$$\|(\boldsymbol{B}^{-1}\boldsymbol{Q})(\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{uFu})\|_\infty \leq \widehat{\Psi}\|\boldsymbol{h}\|_\infty^2, \qquad \|(\boldsymbol{B}^{-1}\boldsymbol{Q})(\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{gd})\|_\infty \leq \widehat{\Psi}\|\boldsymbol{g}\|_\infty\|\boldsymbol{d}\|_\infty$$

due to (4.3.68). Using $\langle \boldsymbol{e}_-\boldsymbol{hSh}\rangle = 0$ and (4.3.67), we obtain

$$\begin{aligned}\left\|\left\langle \boldsymbol{b}\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{uFu}\right\rangle \frac{\boldsymbol{b}}{\langle \boldsymbol{b}^2\rangle}\right\|_\infty &\leq \Big(|{-\mathrm{i}}\langle \boldsymbol{hShe}_-\rangle| + |{-\mathrm{i}}\langle(\boldsymbol{b}-\boldsymbol{e}_-)\boldsymbol{hSh}\rangle| + \Big|\Big\langle\Big(\mathrm{e}^{-\mathrm{i}\psi}+\mathrm{i}\Big)\,\boldsymbol{bhSh}\Big\rangle\Big|\Big) \\ &\quad\times \frac{\|\boldsymbol{b}\|_\infty}{|\langle \boldsymbol{b}^2\rangle|} \\ &\leq \widehat{\Phi}|z|\|\boldsymbol{h}\|_\infty^2.\end{aligned}$$

Similarly, due to (4.3.67) and $\langle \boldsymbol{gde}_-\rangle = \langle g_1(z)d_1(z)\rangle - \langle g_2(z)d_2(z)\rangle = 0$ by the perturbed QVE (4.3.9), we get

$$\begin{aligned}\left\|\left\langle \mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{gdb}\right\rangle \frac{\boldsymbol{b}}{\langle \boldsymbol{b}^2\rangle}\right\|_\infty &\leq \Big(|\langle \boldsymbol{gde}_-\rangle| + |\langle(\boldsymbol{b}-\boldsymbol{e}_-)\boldsymbol{gd}\rangle| + \Big|\Big\langle\Big(\mathrm{e}^{-\mathrm{i}\psi}+\mathrm{i}\Big)\,\boldsymbol{bgd}\Big\rangle\Big|\Big)\frac{\|\boldsymbol{b}\|_\infty}{|\langle \boldsymbol{b}^2\rangle|} \\ &\leq \widehat{\Phi}|z|\|\boldsymbol{g}\|_\infty\|\boldsymbol{d}\|_\infty.\end{aligned}$$

Thus, inverting $\boldsymbol{B}$ in (4.3.58), multiplying the result with $|\boldsymbol{m}|$, taking its norm and using (4.3.68) yield

$$\|\boldsymbol{h}\|_\infty \leq \Phi(\Psi\widehat{\Phi} + \widehat{\Psi})\|\boldsymbol{h}\|_\infty^2 + \Phi(\Psi\widehat{\Phi} + \widehat{\Psi})\|\boldsymbol{g}\|_\infty\|\boldsymbol{d}\|_\infty,$$

which implies

$$\|\boldsymbol{h}\|_\infty \mathbf{1}\Big(\|\boldsymbol{h}\|_\infty \leq \lambda_*\Big) \leq \Phi(1 + 2\Phi(\Psi\widehat{\Phi} + \widehat{\Psi}))\|\boldsymbol{d}\|_\infty$$

by the definition of λ_* in (4.3.69). This concludes the proof of (4.3.43).

For the proof of (4.3.44), inverting $\boldsymbol{B}$ in (4.3.58) and taking the scalar product with $\boldsymbol{w}$ yield

$$\begin{aligned}\langle \boldsymbol{w}\,,\boldsymbol{h}\rangle = {} & \langle \boldsymbol{w}\,,\boldsymbol{B}^{-1}(\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{h}\boldsymbol{S}\boldsymbol{h})\rangle + \frac{\langle \boldsymbol{w}\,,|\boldsymbol{m}|\boldsymbol{B}^{-1}\boldsymbol{b}\rangle}{\langle \boldsymbol{b}^2\rangle}\Big\langle \boldsymbol{h}\boldsymbol{d}\left[(\mathrm{e}^{-\mathrm{i}\psi} + \mathrm{i})\boldsymbol{b} - \mathrm{i}(\boldsymbol{b} - \boldsymbol{e}_-)\right]\Big\rangle \\ & + \langle (\boldsymbol{B}^{-1}\boldsymbol{Q})^*(|\boldsymbol{m}|\boldsymbol{w})\,,\mathrm{e}^{-\mathrm{i}\psi}\boldsymbol{h}\boldsymbol{d}\rangle + \langle \boldsymbol{T}\boldsymbol{w}\,,\boldsymbol{d}\rangle,\end{aligned} \tag{4.3.70}$$

where we used $\langle \boldsymbol{e}_-\boldsymbol{g}\boldsymbol{d}\rangle = 0$ and set

$$\boldsymbol{T}\boldsymbol{w} := \langle \boldsymbol{b}^2\rangle^{-1}\langle |\boldsymbol{m}|\boldsymbol{B}^{-1}\boldsymbol{b}, \boldsymbol{w}\rangle\overline{\boldsymbol{m}}\left[(\mathrm{e}^{\mathrm{i}\psi} - \mathrm{i})\overline{\boldsymbol{b}} + \mathrm{i}(\overline{\boldsymbol{b}} - \boldsymbol{e}_-)\right] + \mathrm{e}^{\mathrm{i}\psi}\overline{\boldsymbol{m}}(\boldsymbol{B}^{-1}\boldsymbol{Q})^*(|\boldsymbol{m}|\boldsymbol{w}).$$

Using (4.3.67) and (4.3.68) as well as a similar argument as in the proof of (4.3.43) for the first term in the definition of $\boldsymbol{T}$ and $\|(\boldsymbol{B}^{-1}\boldsymbol{Q})^*\|_\infty \lesssim 1$ by (4.3.65) for the second term, we obtain $\|\boldsymbol{T}\|_\infty \lesssim 1$. Moreover, as above we see that the first term on the right-hand side of (4.3.70) is $\lesssim \|\boldsymbol{w}\|_\infty\|\boldsymbol{h}\|_\infty^2$. The estimates (4.3.67) and (4.3.68) imply that the second term on the right-hand side of (4.3.70) is $\lesssim \|\boldsymbol{w}\|_\infty\|\boldsymbol{h}\|_\infty\|\boldsymbol{d}\|_\infty$. Applying (4.3.43) to these bounds yields (4.3.44). □

4.3.6. Properly rectangular Gram matrices. In this subsection, we study the behaviour of m_1 and m_2 for z close to zero for p/n different from one. We establish that the density of the limiting distribution is zero around zero – a well-known feature of the Marchenko-Pastur distribution for p/n different from one.

We suppose that the assumptions (A), (C) and (D) are fulfilled and we will study the case $p > n$. More precisely, we assume that

$$\frac{p}{n} \geq 1 + d_* \tag{4.3.71}$$

for some $d_* > 0$ which will imply that each component of m_1 diverges at $z = 0$ whereas each component of m_2 stays bounded at $z = 0$. Later, in the proof of Theorem 4.2.10, we will see that these properties carry over to m and $\widetilde{m}_2$ defined above (4.3.10). We use the notation $D_\delta(w) := \{z \in \mathbb{C} \colon |z - w| < \delta\}$ for $\delta > 0$ and $w \in \mathbb{C}$.

Proposition 4.3.16 (Solution of the QVE close to zero)**.** *If (F2) and* (4.3.71) *are satisfied then there exist a vector* $u \in \mathbb{C}^p$*, a constant* $\delta_* \gtrsim 1$ *and analytic functions* $a \colon D_{\delta_*}(0) \to \mathbb{C}^p$*,* $b \colon D_{\delta_*}(0) \to \mathbb{C}^n$ *such that the unique solution* $\boldsymbol{m} = (m_1, m_2)^t$ *of* (4.3.6) *with* $\operatorname{Im} \boldsymbol{m} > 0$ *fulfills*

$$m_1(z) = za(z) - \frac{u}{z}, \qquad m_2(z) = zb(z) \tag{4.3.72}$$

for all $z \in D_{\delta_*}(0) \cap \mathbb{H}$*. Moreover, we have*

- *(i)* $\sum_{i=1}^p u_i = p - n$ *and* $1 \lesssim u_i \leq 1$ *for all* $i = 1, \ldots, p$*,*
- *(ii)* $b(0) = 1/S^t u \sim 1$*,*
- *(iii)* $\|a(z)\|_\infty + \|b(z)\|_\infty \lesssim 1$ *uniformly for all* $z \in D_{\delta_*}(0)$*,*
- *(iv)* $\lim_{\eta \downarrow 0} \operatorname{Im} m_1(\tau + \mathrm{i}\,\eta) = 0$ *and* $\lim_{\eta \downarrow 0} \operatorname{Im} m_2(\tau + \mathrm{i}\,\eta) = 0$ *locally uniformly for all* $\tau \in (-\delta_*, \delta_*) \backslash \{0\}$*.*

The ansatz (4.3.72) is motivated by the following heuristics. Considering $\boldsymbol{H}$ as an operator $\mathbb{C}^p \oplus \mathbb{C}^n \to \mathbb{C}^n \oplus \mathbb{C}^p$, we expect that the first component described by $X^* \colon \mathbb{C}^p \to \mathbb{C}^n$ has a nontrivial kernel for dimensional reasons whereas the second component has a trivial kernel. Since the nonzero eigenvalues of $\boldsymbol{H}^2$ correspond to the nonzero eigenvalues of XX^* and X^*X, the Marchenko-Pastur distribution indicates that there is a constant $\delta_* \gtrsim 1$ such that $\boldsymbol{H}$ has no nonzero eigenvalue in $(-\delta_*, \delta_*)$. As the first component m_1 of $\boldsymbol{m}$ corresponds to the "first component" of $\boldsymbol{H}$, the term $-u/z$ in (4.3.72) implements the expected kernel. For dimensional reasons, the kernel should be $p - n$ dimensional which agrees with part (i) of Proposition 4.3.16. The factor z in the terms $za(z)$ and $zb(z)$ in (4.3.72) realizes the expected gap in the eigenvalue distribution around zero.

Proof of Proposition 4.3.16. We start with the defining equations for u and b. We assume that $u \in (0, 1]^p$ fulfills

$$\frac{1}{u} = 1 + S\frac{1}{S^t u} \tag{4.3.73}$$

and $b\colon D_{\delta_*}(0)\to\mathbb{C}^p$ fulfills

$$-\frac{1}{b(z)}=z^2-S^t\frac{1}{1+Sb(z)} \tag{4.3.74}$$

for some $\delta_*>0$. We then define $a\colon D_{\delta_*}(0)\to\mathbb{C}^p$ through

$$z^2a(z)=u-\frac{1}{1+Sb(z)} \tag{4.3.75}$$

and set $\widehat{m}_1(z):=za(z)-u/z$ and $\widehat{m}_2(z):=zb(z)$ for $z\in D_{\delta_*}(0)$. Thus, for $z\in D_{\delta_*}(0)$, we obtain

$$z+S^t\widehat{m}_1(z)=z-S^t\frac{1}{1+Sb(z)}=-\frac{1}{zb(z)}=-\frac{1}{\widehat{m}_2(z)},$$

where we used (4.3.75) in the first step and (4.3.74) in the second step. Similarly, solving (4.3.75) for $Sb(z)$ yields

$$z+S\widehat{m}_2(z)=z+z\left(\frac{1}{u-z^2a(z)}-1\right)=-\frac{1}{\widehat{m}_1(z)},\qquad z\in D_{\delta_*}(0). \tag{4.3.76}$$

Thus, $(\widehat{m}_1,\widehat{m}_2)$ satisfy (4.3.6), the defining equation for $\boldsymbol{m}=(m_1,m_2)$ and we will be able to conclude that $\widehat{m}_1=m_1$ and $\widehat{m}_2=m_2$.

For the rigorous argument, we first establish the existence and uniqueness of u and b that follow from the next two lemmata whose proofs are given later.

Lemma 4.3.17. *If (F2) and* (4.3.71) *are satisfied then there is a unique solution of* (4.3.73) *in the set* $u\in(0,1]^p$. *Moreover,*

$$1>u_i\gtrsim 1,\qquad (S^tu)_k\gtrsim 1 \tag{4.3.77}$$

for all $i=1,\dots,p$ *and* $k=1,\dots,n$ *and* $\sum_{i=1}^p u_i=p-n$.

Lemma 4.3.18. *If (F2) and* (4.3.71) *are satisfied, then there are a* $\delta_*\sim 1$ *and a unique holomorphic function* $b\colon D_{\delta_*}(0)\to\mathbb{C}^n$ *satisfying* (4.3.74) *with* $b(0)=1/(S^tu)$, *where* u *is the solution of* (4.3.73). *Moreover, we have* $\|b(z)\|_\infty\lesssim 1$ *and* $\|(1+Sb(z))^{-1}\|_\infty\le 1/2$ *for all* $z\in D_{\delta_*}(0)$, $b(0)\sim 1$, $b'(0)=0$, $\operatorname{Im}(zb(z))>0$ *for all* $z\in D_{\delta_*}(0)$ *with* $\operatorname{Im}z>0$ *and* $\operatorname{Im}(zb(z))=0$ *for* $z\in(-\delta_*,\delta_*)$.

Given u and $b(z)$, the formula (4.3.75) defines $a(z)$ for $z\neq 0$. To extend its definition to $z=0$, we observe that the right-hand side of (4.3.75) is a holomorphic function for all

$z \in D_{\delta_*}(0)$ by Lemma 4.3.18. Since $b(0) = 1/(S^t u)$ and the derivative of the right-hand side of (4.3.75) vanishes as $b'(0) = 0$, the first two coefficients of the Taylor series of the right-hand side on $D_{\delta_*}(0)$ are zero by (4.3.73). Thus, (4.3.75) defines a holomorphic function $a\colon D_{\delta_*}(0) \to \mathbb{C}^p$.

Furthermore, $\operatorname{Im} \widehat{m}_2(z) > 0$ for $\operatorname{Im} z > 0$ by Lemma 4.3.18. Taking the imaginary part of (4.3.76) yields

$$\frac{\operatorname{Im} \widehat{m}_1(z)}{|\widehat{m}_1(z)|^2} = \operatorname{Im} z + S \operatorname{Im} \widehat{m}_2(z), \tag{4.3.78}$$

which implies $\operatorname{Im} \widehat{m}_1(z) > 0$ for $\operatorname{Im} z > 0$ as $\operatorname{Im} \widehat{m}_2(z) > 0$ for $z \in \mathbb{H} \cap D_{\delta_*}(0)$. Since the solution $\boldsymbol{m}$ of (4.3.6) with $\operatorname{Im} \boldsymbol{m}(z) > 0$ for $\operatorname{Im} z > 0$ is unique by Theorem 2.1 in [4], we have $\boldsymbol{m}(z) = \widehat{\boldsymbol{m}}(z) := (\widehat{m}_1(z), \widehat{m}_2(z))^t$ for all $z \in \mathbb{H} \cap D_{\delta_*}(0)$. The statements in (i), (ii) and (iii) follow from Lemma 4.3.17, Lemma 4.3.18 and (4.3.75).

For the proof of (iv), we note that $\lim_{\eta \downarrow 0} \operatorname{Im} m_2(\tau + \mathrm{i}\,\eta) = 0$ for all $\tau \in (-\delta_*, \delta_*)$ locally uniformly by Lemma 4.3.18. Because of (4.3.78) and the locally uniform convergence of $m_1(\tau + \mathrm{i}\,\eta)$ to $\tau a(\tau) - u/\tau$ for $\eta \downarrow 0$ and $\tau \in (-\delta_*, \delta_*) \setminus \{0\}$, we have $\lim_{\eta \downarrow 0} \operatorname{Im} m_1(\tau + \mathrm{i}\,\eta) = 0$ locally uniformly for all $\tau \in (-\delta_*, \delta_*) \setminus \{0\}$ as well, which concludes the proof of (iv). □

We conclude this subsection with the proofs of Lemma 4.3.17 and Lemma 4.3.18.

Proof of Lemma 4.3.17. We will show that the functional

$$J\colon (0,1]^p \to \mathbb{R}, \quad u \mapsto \frac{1}{p} \sum_{j=1}^{n} \log \Big(\sum_{i=1}^{p} s_{ij} u_i \Big) + \frac{1}{p} \sum_{i=1}^{p} (u_i - \log u_i)$$

has a unique minimizer u with $u_i > 0$ for all $i = 1, \dots, p$ which solves (4.3.73). Note that

$$J(1, \dots, 1) = \frac{1}{p} \sum_{j=1}^{n} \log \Big(\sum_{i=1}^{p} s_{ij} \Big) + \frac{p}{p} \leq 1. \tag{4.3.79}$$

We start with an auxiliary bound on the components of u. Using (F2) and Jensen's inequality, we get

$$
\begin{aligned}
J(u) &\geq \frac{1}{p}\sum_{k=1}^{n}\log\left(\sum_{i=1}^{p}\frac{\varphi}{n+p}u_i\right)+\frac{1}{p}\sum_{i=1}^{p}(u_i-\log u_i)\\
&\geq \frac{1}{p}\left(\sum_{i=1}^{p}\frac{n}{p}\log\left(\frac{\varphi}{2}u_i\right)-\sum_{i=1}^{p}\log u_i\right)\\
&\geq -\frac{1}{p}\frac{d_*}{1+d_*}\sum_{i=1}^{p}\log u_i+\frac{n}{p}\log\left(\frac{\varphi}{2}\right),
\end{aligned} \tag{4.3.80}
$$

where we used (4.3.71) in the last step. For any $u \in (0,1]^p$ with $J(u) \leq J(1,\dots,1)$, using (4.3.79), we obtain

$$
\begin{aligned}
1 \geq J(1,\dots,1) \geq J(u) &\geq -\frac{d_*}{p(1+d_*)}\sum_{i=1}^{p}\log u_i+\frac{n}{p}\log\left(\frac{\varphi}{2}\right)\\
&\geq -\frac{d_*}{p(1+d_*)}\log u_i+\frac{1}{r_1}\log\left(\frac{\varphi}{2}\right),
\end{aligned}
$$

for any $i=1,\dots,p$, i.e., $u_i \geq \exp(-p(1+d_*)(1-r_1^{-1}\log(\varphi/2))/d_*) > 0$.

Therefore, taking a minimizing sequence, using a compactness argument and the continuity of J, we obtain the existence of $u^\star \in (0,1]^p$ such that $J(u^\star) = \inf_{u\in(0,1]^p} J(u)$ and

$$
u_i^\star \geq \exp\left(-p\frac{1+d_*}{d_*}\left(1-\frac{1}{r_1}\log\left(\frac{\varphi}{2}\right)\right)\right), \qquad i=1,\dots,p. \tag{4.3.81}
$$

Next, we show that $u_i^\star < 1$ for all $i=1,\dots,p$. Assume that $u_i^\star = 1$ for some $i \in \{1,\dots,p\}$. Consider a vector $\widehat{u}$ that agrees with $u^\star$ except that $u_i^\star$ is replaced by $\lambda \in (0,1)$. An elementary calculation then shows that $J(\widehat{u}) \geq J(u^\star)$ implies $s_{ik} = 0$ for all $k=1,\dots,n$ which contradicts (4.3.15).

Therefore, evaluating the derivative $J(u^\star + \tau h)$ for $h \in \mathbb{R}^p$ at $\tau = 0$, which vanishes since $u^\star \in (0,1)^p$ is a minimizer, we see that $u^\star$ satisfies (4.3.73).

To see the uniqueness of the solution of (4.3.73), we suppose that $u^\star, v^\star \in (0,1]^p$ satisfy (4.3.73), i.e., $u^\star = f(u^\star)$ and $v^\star = f(v^\star)$ where $f\colon (0,1]^p \to (0,1]^p$, $f(u) = (1+S((S^t u)^{-1}))^{-1}$. On $(0,1]^p$ we define the distance function

$$
D(u,v) := \sup_{i=1,\dots,p} d(u_i,v_i) \tag{4.3.82}
$$

where $d(a,b) = (a-b)^2/(ab)$ for $a,b>0$. This function d defined on $(0,\infty)^2$ is the analogue of D defined in (A.6) of [5] on $\mathbb{H}^2$. Therefore, we can apply Lemma A.2 in [5] with the natural substitutions which yields

$$D(u^\star, v^\star) = D(f(u^\star), f(v^\star)) = \left(1 + \frac{1}{S(S^t u^\star)^{-1}}\right)^{-1} \left(1 + \frac{1}{S(S^t v^\star)^{-1}}\right)^{-1} D(u^\star, v^\star)$$
$$\leq cD(u^\star, v^\star).$$

for some number c. Here we used 1. and 2. of Lemma A.2 in [5] in the second step and 3. of Lemma A.2 in [5] in the last step. Since we can choose $c < 1$ by (4.3.81), we conclude $u^\star = v^\star$. This argument applies particularly to minimizers of J on $(0,1]^p$.

In the following, we will denote the unique minimizer of J by u. To compute the sum of the components of u we multiply (4.3.73) by u and sum over $i = 1, \ldots, p$ and obtain

$$p = \sum_{i=1}^p u_i + \sum_{i=1}^p u_i \left(S \frac{1}{S^t u}\right)_i = \sum_{i=1}^p u_i + \sum_{j=1}^n (S^t u)_j \frac{1}{(S^t u)_j} = \sum_{i=1}^p u_i + n,$$

i.e., $\sum_{i=1}^p u_i = p - n$.

Finally, we show that the components of the minimizer u are bounded from below by a positive constant which only depends on the model parameters. For $k \in \{1, \ldots, n\}$, we obtain

$$(S^t u)_k \geq \frac{\varphi}{n+p} \sum_{i=1}^p u_i \geq \frac{\varphi}{2} \langle u \rangle = \frac{\varphi}{2} \left(1 - \frac{n}{p}\right) \geq \frac{\varphi d_*}{2(1+d_*)}, \tag{4.3.83}$$

where we used (F2) in the first step, $n \leq p$ in the second step, $\sum_{i=1}^p u_i = p - n$ in the third step and (4.3.71) in the last step. This implies the third bound in (4.3.77).

Therefore, we obtain for all $i = 1, \ldots, p$ from (4.3.73)

$$\frac{1}{u_i} = 1 + \sum_{k=1}^n s_{ik} \frac{1}{(S^t u)_k} \leq 1 + \frac{2(1+d_*)}{\varphi d_*},$$

where we used (A) with $s_* = 1$ in the last step. This shows that u_i is bounded from below by a positive constant which only depends on the model parameters, i.e., the second bound in (4.3.77). □

Proof of Lemma 4.3.18. Instead of solving (4.3.74) directly, we solve a differential equation with the correctly chosen initial condition in order to obtain b. Note that $b_0 := 1/(S^t u)$ fulfills (4.3.74) for $z = 0$ and $b_0 \sim 1$ by (4.3.77) and (4.3.15).

For any $b \in \mathbb{C}^n$ satisfying $(Sb)_i \neq -1$ for $i = 1, \ldots, p$, we define the linear operator

$$L(b) \colon \mathbb{C}^n \to \mathbb{C}^n, \quad v \mapsto L(b)v := bS^t \frac{1}{(1+Sb)^2} S(bv),$$

where bv is understood as componentwise multiplication. Using the definition of $L(b)$, $b_0 = 1/(S^t u)$ and (4.3.73), we get

$$L(b_0)1 = \frac{1}{S^t u} S^t u^2 S \frac{1}{S^t u} = \frac{1}{S^t u}\left(S^t u - S^t u^2\right) = 1 - \frac{S^t u^2}{S^t u} \leq 1 - \kappa \tag{4.3.84}$$

for some $\kappa \sim 1$. Here we used (4.3.15), $u^2 \gtrsim 1$ and (4.3.77) in the last step. As

$$L(b_0) = \frac{1}{S^t u} S^t u^2 S \left(\frac{1}{S^t u}\,\cdot\right)$$

is symmetric and positivity-preserving, Lemma 4.6 in [4] implies $\|L(b_0)\|_{2\to 2} \leq 1 - \kappa$ because of (4.3.84). Therefore, $(1 - L(b_0))$ is invertible and $\|(1 - L(b_0))^{-1}\|_{2\to 2} \leq \kappa^{-1}$. Moreover, $\|(1 - L(b_0))^{-1}\|_{\infty\to\infty} \leq 1 + \|L(b_0)\|_{2\to\infty}\kappa^{-1}$ by (4.3.27) with $R = L(b_0)$ and $D = 1$. The estimate (4.3.11) and the submultiplicativity of the operator norm $\|\cdot\|_2$ yield $\|L(b_0)\|_{2\to\infty} \lesssim 1$. Thus, we obtain

$$\|(1 - L(b_0))^{-1}\|_\infty \lesssim 1.$$

We introduce the notation $U_{\delta'} := \{b \in \mathbb{C}^n; \|b - b_0\|_\infty < \delta'\}$. If we choose $\delta' \leq (2\|S\|_{\infty\to\infty})^{-1}$ then

$$|(1 + Sb)_i| = |u_i^{-1} + (S(b - b_0))_i| \geq |u_i^{-1}| - \|S\|_{\infty\to\infty}\|b - b_0\|_\infty \geq 1/2$$

for all $i = 1, \ldots, p$, where we used the definition of b_0, (4.3.73) and $u_i \leq 1$. Therefore, $\|(1 + Sb)^{-1}\|_\infty \leq 1/2$ for all $b \in U_{\delta'}$, i.e., $U_{\delta'} \to \mathbb{C}^{n\times n}$, $b \mapsto L(b)$ will be a holomorphic map. In particular,

$$\|L(b) - L(b_0)\|_\infty \lesssim \|b - b_0\|_\infty. \tag{4.3.85}$$

If $D := L(b) - L(b_0)$ and $\|(1 - L(b_0))^{-1}D\|_{\infty\to\infty} \leq 1/2$ then $(1 - L(b))$ will be invertible and

$$(1 - L(b))^{-1} = \left(1 - (1 - L(b_0))^{-1}D\right)^{-1}(1 - L(b_0))^{-1},$$

as well as $\|(1-L(b))^{-1}\|_{\infty\to\infty} \leq 2\|(1-L(b_0))^{-1}\|_{\infty\to\infty}$. Therefore, (4.3.85) implies the existence of $\delta' \sim 1$ such that $(1-L(b))$ is invertible and $\|(1-L(b))^{-1}\|_\infty \lesssim 1$ for all $b \in U_{\delta'}$.

Hence, the right-hand side of the differential equation

$$b' := \frac{\partial}{\partial z} b = 2zb(1-L(b))^{-1}b =: f(z,b) \tag{4.3.86}$$

is holomorphic on $D_{\delta'}(0) \times U_{\delta'}$. As $\delta' \sim 1$ and $\sup\{\|f(z,w)\|_\infty; z \in D_{\delta'}(0), b \in U_{\delta'}\} \lesssim 1$, the standard theory of holomorphic differential equations yields the existence of $\delta_* \gtrsim 1$ and a holomorphic function $b\colon D_{\delta_*}(0) \to \mathbb{C}^n$ which is the unique solution of (4.3.86) on $D_{\delta_*}(0)$ satisfying $b(0) = b_0$.

The solution of the differential equation (4.3.86) is a solution of (4.3.74) since dividing by b, multiplying by $(1-L(b))$ and dividing by b in (4.3.86) yields

$$\frac{b'}{b^2} = 2z + \frac{1}{b} L(b) \frac{b'}{b}.$$

This is the derivative of (4.3.74). Since $b(0) = b_0$ fulfils (4.3.74) for $z = 0$ the unique solution of (4.3.86) with this initial condition is a solution of (4.3.74) for $z \in D_{\delta_*}(0)$. There is only one holomorphic solution of (4.3.74) due to the uniqueness of the solution of (4.3.86). This proves the existence and uniqueness of $b(z)$ in Lemma 4.3.18.

Since b is a holomorphic function on $D_{\delta_*}(0)$ such that $|b(z)| \lesssim 1$ on $D_{\delta_*}(0)$ and $\delta_* \sim 1$ there is a holomorphic function $b_1\colon D_{\delta_*}(0) \to \mathbb{C}^n$ such that

$$b(z) = b_0 + b_1(z)z$$

and $|b_1(z)| \lesssim 1$. Thus, we can assume that $\delta_* \gtrsim 1$ is small enough such that $\operatorname{Im} zb(z) \geq (b_0 - |z||b_1(z)|)\operatorname{Im} z > 0$ for all $z \in D_{\delta_*}(0) \cap \mathbb{H}$.

Taking the imaginary part of (4.3.74) for $\tau \in \mathbb{R}$, we get

$$\frac{\operatorname{Im} b(\tau)}{|b(\tau)|^2} = S^t \frac{1}{|1+Sb(\tau)|^2} S \operatorname{Im} b(\tau)$$

or equivalently, introducing

$$\widetilde{L}(z)\colon \mathbb{C}^n \to \mathbb{C}^n, \quad v \mapsto \widetilde{L}(z)v := |b(z)| S^t |1+Sb(z)|^{-2} S(|b(z)|v)$$

for $z \in D_{\delta_*}(0)$, we have

$$\Big(1 - \widetilde{L}(\tau)\Big) \frac{\operatorname{Im} b(\tau)}{|b(\tau)|} = 0. \tag{4.3.87}$$

As $\|(1 + Sb(z))^{-1}\|_\infty \le 1/2$ for all $z \in D_{\delta_*}(0)$, the linear operator $\widetilde{L}(z)$ is well-defined for all $z \in D_{\delta_*}(0)$. Because $\widetilde{L}(0) = L(b_0)$ and $\|\widetilde{L}(b) - \widetilde{L}(b_0)\|_\infty \lesssim \|b - b_0\|_\infty$ we can assume that $\delta_* \gtrsim 1$ is small enough such that $(1 - \widetilde{L}(z))$ is invertible for all $z \in D_{\delta_*}(0)$. Thus, (4.3.87) implies that $\operatorname{Im} b(\tau) = 0$ for all $\tau \in (-\delta_*, \delta_*)$ and consequently, $\operatorname{Im} \tau b(\tau) = 0$ for all $\tau \in (-\delta_*, \delta_*)$. □

4.4. Local laws

4.4.1. Local law for $\boldsymbol{H}$. In this section, we will follow the approach used in [7] to prove a local law for the Wigner-type matrix $\boldsymbol{H}$. We will not give all details but refer the reader to [7]. Therefore, we consider (4.3.4) as a perturbed QVE of the form (4.3.9) with $\boldsymbol{g} := (g_1, g_2)^t \colon \mathbb{H} \to \mathbb{C}^{p+n}$ and $\boldsymbol{d} := (d_1, d_2)^t \colon \mathbb{H} \to \mathbb{C}^{p+n}$, in particular $\boldsymbol{g}(z) = (G_{xx}(z))_{x=1,\dots,n+p}$ where G_{xx} are the diagonal entries of the resolvent of $\boldsymbol{H}$ defined in (4.3.3). We recall that ρ is the probability measure on $\mathbb{R}$ whose Stieltjes transform is $\langle \boldsymbol{m} \rangle$, cf. (4.3.8), where $\boldsymbol{m}$ is the solution of (4.3.6) satisfying $\operatorname{Im} \boldsymbol{m}(z) > 0$ for $z \in \mathbb{H}$.

Definition 4.4.1 (Stochastic domination)**.** Let $P_0 \colon (0,\infty)^2 \to \mathbb{N}$ be a given function which depends only on the model parameters and the tolerance exponent γ. If $\varphi = (\varphi^{(p)})_p$ and $\psi = (\psi^{(p)})_p$ are two sequences of nonnegative random variables then we will say that φ is *stochastically dominated* by ψ, $\varphi \prec \psi$, if for all $\varepsilon > 0$ and $D > 0$ we have

$$\mathbb{P}\left(\varphi^{(p)} \ge p^\varepsilon \psi^{(p)}\right) \le p^{-D}$$

for all $p \ge P_0(\varepsilon, D)$.

In the following, we will use the convention that $\tau := \operatorname{Re} z$ and $\eta := \operatorname{Im} z$ for $z \in \mathbb{C}$.

Theorem 4.4.2 (Local law for $\boldsymbol{H}$ away from the edges)**.** *Fix any $\delta, \varepsilon_* > 0$ and $\gamma \in (0,1)$ independent of p. If the random matrix X satisfies (A) – (D) then the resolvent entries*

$G_{xy}(z)$ of $\boldsymbol{H}$ defined in (4.3.3) *and* (4.3.1)*, respectively, fulfill*

$$\max_{x,y=1,\ldots,n+p} |G_{xy}(z) - \boldsymbol{m}_x(z)\delta_{xy}| \prec \frac{1}{\sqrt{p\eta}}, \quad \textit{if } \operatorname{Im} z \geq p^{-1+\gamma} \textit{ and } \langle \operatorname{Im} \boldsymbol{m}(z) \rangle \geq \varepsilon_*, \tag{4.4.1a}$$

$$\max_{x,y=1,\ldots,n+p} |G_{xy}(z) - \boldsymbol{m}_x(z)\delta_{xy}| \prec \frac{1}{\sqrt{p}}, \quad \textit{if } \operatorname{dist}(z, \operatorname{supp} \rho) \geq \varepsilon_*, \tag{4.4.1b}$$

uniformly for $z \in \mathbb{H}$ satisfying $\delta \leq |z| \leq 10$. For any sequence of deterministic vectors $\boldsymbol{w} \in \mathbb{C}^{n+p}$ satisfying $\|\boldsymbol{w}\|_\infty \leq 1$, we have

$$|\langle \boldsymbol{w}, \boldsymbol{g}(z) - \boldsymbol{m}(z) \rangle| \prec \frac{1}{p\eta}, \quad \textit{if } \operatorname{Im} z \geq p^{-1+\gamma} \textit{ and } \langle \operatorname{Im} \boldsymbol{m}(z) \rangle \geq \varepsilon_*, \tag{4.4.2a}$$

$$|\langle \boldsymbol{w}, \boldsymbol{g}(z) - \boldsymbol{m}(z) \rangle| \prec \frac{1}{p}, \quad \textit{if } \operatorname{dist}(z, \operatorname{supp} \rho) \geq \varepsilon_*, \tag{4.4.2b}$$

uniformly for $z \in \mathbb{H}$ satisfying $\delta \leq |z| \leq 10$. Here, the threshold function P_0 in the definition of the relation $\prec$ depends on the model parameters as well as δ, ε_* and γ.

Remark 4.4.3. The proof of Theorem 4.4.2 actually shows an explicit dependence of the estimates (4.4.1) and (4.4.2) on ε_*. More precisely, if the right-hand sides of (4.4.1) and (4.4.2) are multiplied by a universal inverse power of ε_* and the right-hand side of the condition $\operatorname{Im} z \geq p^{-1+\gamma}$ is multiplied by the same inverse power of ε_* then Theorem 4.4.2 holds true where the relation $\prec$ does not depend on ε_* any more.

Let $\mu_1 \leq \ldots \leq \mu_{n+p}$ be the eigenvalues of $\boldsymbol{H}$. We define

$$I(\tau) := \left\lceil (n+p) \int_{-\infty}^{\tau} \rho(\mathrm{d}\omega) \right\rceil, \quad \tau \in \mathbb{R}. \tag{4.4.3}$$

Thus, $I(\tau)$ denotes the index of an eigenvalue expected to be close to the spectral parameter $\tau \in \mathbb{R}$.

Corollary 4.4.4 (Bulk rigidity, Absence of eigenvalues outside of $\operatorname{supp} \rho$)**.** *Let $\delta, \varepsilon_* > 0$.*

(i) Uniformly for all $\tau \in [-10, -\delta] \cup [\delta, 10]$ satisfying $\rho(\tau) \geq \varepsilon_*$ or $\operatorname{dist}(\tau, \operatorname{supp} \rho) \geq \varepsilon_*$, we have

$$\left| \#\{j; \mu_j \leq \tau\} - (n+p) \int_{-\infty}^{\tau} \rho(\mathrm{d}\omega) \right| \prec 1. \tag{4.4.4}$$

(ii) Uniformly for all $\tau \in [-10, -\delta] \cup [\delta, 10]$ satisfying $\rho(\tau) \geq \varepsilon_*$, we have

$$|\mu_{I(\tau)} - \tau| \prec \frac{1}{n+p}. \tag{4.4.5}$$

(iii) Asymptotically with overwhelming probability, we have

$$\#\Big(\operatorname{Spec}(\boldsymbol{H}) \cap \{\tau \in [-10, -\delta] \cup [\delta, 10]; \operatorname{dist}(\tau, \operatorname{supp} \rho) \geq \varepsilon_*\}\Big) = 0. \tag{4.4.6}$$

The estimates (4.4.2a) and (4.4.2b) in Theorem 4.4.2 imply Corollary 4.4.4 in the same way as the corresponding results, Corollary 1.10 and Corollary 1.11, in [7] were proven. In fact, inspecting the proofs in [7], rigidity at a particular point τ_0 in the bulk requires only (i) the local law, (4.4.2a), around $\tau_0 = \operatorname{Re} z$, (ii) the local law somewhere outside of the support of ρ, (4.4.2b), and (iii) a uniform global law with optimal convergence rate, (4.4.2b), for any z away from $\operatorname{supp} \rho$.

Proof of Theorem 4.4.2. In the proof, we will use the following shorter notation. We introduce the spectral domain

$$\mathbb{D}_{\boldsymbol{H}} := \{z \in \mathbb{H} \colon \delta \leq |z| \leq 10, \quad \operatorname{Im} z \geq p^{-1+\gamma}, \quad \langle \operatorname{Im} \boldsymbol{m}(z) \rangle \geq \varepsilon_* \text{ or } \operatorname{dist}(z, \operatorname{supp} \rho) \geq \varepsilon_*\}$$

for the parameters $\gamma > 0, \varepsilon_* > 0$ and $\delta > 0$. Moreover, we define the random control parameters

$$\Lambda_d(z) := \|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty, \quad \Lambda_o(z) := \max_{\substack{x,y=1,\ldots,n+p \\ x \neq y}} |G_{xy}(z)|, \quad \Lambda(z) := \max\{\Lambda_d(z), \Lambda_o(z)\}.$$

Before proving (4.4.1) and (4.4.2), we establish the auxiliary estimates: Uniformly for all $z \in \mathbb{D}_{\boldsymbol{H}}$, we have

$$\Lambda_d(z) + \|\boldsymbol{d}(z)\|_\infty \prec \sqrt{\frac{\langle \operatorname{Im} \boldsymbol{m}(z) \rangle}{(n+p)\eta}} + \frac{1}{(n+p)\eta} + \frac{1}{\sqrt{n+p}}, \tag{4.4.7a}$$

$$\Lambda_o(z) \prec \sqrt{\frac{\langle \operatorname{Im} \boldsymbol{m}(z) \rangle}{(n+p)\eta}} + \frac{1}{(n+p)\eta} + \frac{1}{\sqrt{n+p}}. \tag{4.4.7b}$$

Moreover, for every sequence of vectors $\boldsymbol{w} \in \mathbb{C}^{p+n}$ satisfying $\|\boldsymbol{w}\|_\infty \leq 1$, we have

$$|\langle \boldsymbol{w}, \boldsymbol{g}(z) - \boldsymbol{m}(z) \rangle| \prec \frac{\langle \operatorname{Im} \boldsymbol{m}(z) \rangle}{(n+p)\eta} + \frac{1}{(n+p)^2\eta^2} + \frac{1}{n+p} \tag{4.4.8}$$

uniformly for $z \in \mathbb{D}_{\boldsymbol{H}}$.

Now, we show that (4.4.8) follows from (4.4.7a) and (4.4.7b). To that end, we use the following lemma which is proven as Theorem 3.5 in [7].

Lemma 4.4.5 (Fluctuation Averaging)**.** *For any $z \in \mathbb{D}_{\boldsymbol{H}}$ and any sequence of deterministic vectors $\boldsymbol{w} \in \mathbb{C}^{n+p}$ with the uniform bound, $\|\boldsymbol{w}\|_\infty \leq 1$ the following holds true: If $\Lambda_o(z) \prec \Phi$ for some deterministic (n and p-dependent) control parameter Φ with $\Phi \leq (n+p)^{-\gamma/3}$ and $\Lambda(z) \prec (n+p)^{-\gamma/3}$ a.w.o.p., then*

$$|\langle \boldsymbol{w}, \boldsymbol{d}(z)\rangle| \prec \Phi^2 + \frac{1}{n+p}. \tag{4.4.9}$$

By (4.4.7a), the indicator function in (4.3.36) is nonzero a.w.o.p. Moreover, (4.4.7b) ensures the applicability of the fluctuation averaging, Lemma 4.4.5, which implies that the last term in (4.3.36) is stochastically dominated by the right-hand side in (4.4.8). Using (4.4.7a) again, we conclude that the first term of the right-hand side of (4.3.36) is dominated by the right-hand side of (4.4.8).

In order to show (4.4.7a) and (4.4.7b) we use the following lemma whose proof we omit, since it follows exactly the same steps as the proof of Lemma 2.1 in [7].

Lemma 4.4.6. *Let $\lambda_*\colon \mathbb{H} \to (0,\infty)$ be the function from Lemma 4.3.9. We have*

$$\|\boldsymbol{d}(z)\|_\infty \mathbf{1}(\Lambda(z) \leq \lambda_*(z)) \prec \sqrt{\frac{\operatorname{Im}\langle \boldsymbol{g}(z)\rangle}{(n+p)\eta}} + \frac{1}{\sqrt{n+p}}, \tag{4.4.10a}$$

$$\Lambda_o(z) \mathbf{1}(\Lambda(z) \leq \lambda_*(z)) \prec \sqrt{\frac{\operatorname{Im}\langle \boldsymbol{g}(z)\rangle}{(n+p)\eta}} + \frac{1}{\sqrt{n+p}} \tag{4.4.10b}$$

uniformly for all $z \in \mathbb{D}_{\boldsymbol{H}}$.

By (4.3.35) and (4.4.10a), we obtain

$$(\Lambda_d(z) + \|\boldsymbol{d}(z)\|_\infty)\mathbf{1}(\Lambda_d(z) \leq \lambda_*(z)) \prec \sqrt{\frac{\langle \operatorname{Im} \boldsymbol{m}\rangle}{(n+p)\eta}} + (n+p)^{-\varepsilon}\Lambda_d + \frac{(n+p)^{\varepsilon}}{(n+p)\eta} + \frac{1}{\sqrt{n+p}}$$

for any $\varepsilon \in (0,\gamma)$. Here we used $\operatorname{Im} \boldsymbol{g} = \operatorname{Im} \boldsymbol{m} + \mathcal{O}(\Lambda_d)$. We absorbe $(n+p)^{-\varepsilon}\Lambda_d$ into the left-hand side and get

$$(\Lambda_d(z) + \|\boldsymbol{d}(z)\|_\infty)\mathbf{1}(\Lambda_d(z) \le \lambda_*(z)) \prec \sqrt{\frac{\langle \operatorname{Im} \boldsymbol{m}\rangle}{(n+p)\eta}} + \frac{1}{(n+p)\eta} + \frac{1}{\sqrt{n+p}} \tag{4.4.11}$$

as $\varepsilon \in (0,\gamma)$ is arbitrary. From (4.4.10b), we conclude

$$\Lambda_o(z)\mathbf{1}(\Lambda(z) \le \lambda_*(z)) \prec \sqrt{\frac{\langle \operatorname{Im} \boldsymbol{m}\rangle}{(n+p)\eta}} + \frac{1}{(n+p)\eta} + \frac{1}{\sqrt{n+p}}, \tag{4.4.12}$$

where we used $\operatorname{Im} \boldsymbol{g} = \operatorname{Im} \boldsymbol{m} + \mathcal{O}(\Lambda_d)$ and (4.4.11) and the fact that $\Lambda_d \le \Lambda$.

We will conclude the proof by establishing that $\mathbf{1}(\Lambda(z) \le \lambda_*(z)) = 1$ a.w.o.p. due to an application of Lemma A.1 in [7]. Combining (4.4.11) and (4.4.12) and using $\langle \operatorname{Im} \boldsymbol{m}(z)\rangle \lesssim (\operatorname{Im} z)^{-1}$, we obtain

$$\Lambda(z)\mathbf{1}(\Lambda(z) \le \lambda_*(z)) \prec (n+p)^{-\gamma/2} \tag{4.4.13}$$

for $z \in \mathbb{D}_{\boldsymbol{H}}$ by the definition of $\mathbb{D}_{\boldsymbol{H}}$. We define the function $\Phi(z) := (n+p)^{-\gamma/3}$ and note that $\Lambda(z) = \|\boldsymbol{g}(z) - \boldsymbol{m}(z)\|_\infty$ is Hölder-continuous since $\boldsymbol{g}$ and $\boldsymbol{m}$ are Hölder-continuous by

$$\max_{x,y=1,\ldots,n+p} |G_{xy}(z_1) - G_{xy}(z_2)| \le \frac{|z_1 - z_2|}{(\operatorname{Im} z_1)(\operatorname{Im} z_2)} \le (n+p)^2|z_1 - z_2| \tag{4.4.14}$$

for $z_1, z_2 \in \mathbb{D}_{\boldsymbol{H}}$ and Lemma 4.3.8, respectively. We choose $z_0 := 10\mathrm{i}$. Since $|G_{xy}(z)| \le (\operatorname{Im} z)^{-1}$ and $|\boldsymbol{m}_x(z)| \le (\operatorname{Im} z)^{-1}$ we get $\Lambda(10\mathrm{i}) \le 1$ and hence $\mathbf{1}(\Lambda(10\mathrm{i}) \le \lambda_*(10\mathrm{i})) = 1$ by Lemma 4.3.9. Therefore, we conclude $\Lambda(z_0) \le (n+p)^{-\gamma/2} \le \Phi(z_0)$ from (4.4.13). Moreover, (4.4.13) implies $\Lambda \cdot \mathbf{1}(\Lambda \in [\Phi - (n+p)^{-1}, \Phi]) < \Phi - (n+p)^{-1}$ a.w.o.p. uniformly on $\mathbb{D}_{\boldsymbol{H}}$. Thus, we get $\Lambda(z) \le (n+p)^{-\gamma/3}$ a.w.o.p. for all $z \in \mathbb{D}_{\boldsymbol{H}}$ by applying Lemma A.1 in [7] to Λ and Φ on the connected domain $\mathbb{D}_{\boldsymbol{H}}$, i.e., $\mathbf{1}(\Lambda(z) \le \lambda_*(z)) = 1$ a.w.o.p. Therefore, (4.4.11) and (4.4.12) yield (4.4.7a) and (4.4.7b), respectively. As remarked above this also implies (4.4.1a).

For the proof of (4.4.1b) and (4.4.2b), we first notice that

$$G_{xx}(z) = \sum_{a=1}^{n+p} \frac{|\boldsymbol{u}_a(x)|^2}{\mu_a - z}$$

for all $x = 1, \dots, n+p$, where $\boldsymbol{u}_a(x)$ denotes the x-component of a $\|\cdot\|_2$ normalized eigenvector $\boldsymbol{u}_a$ corresponding to the eigenvalue μ_a of $\boldsymbol{H}$. Therefore, we conclude

$$\operatorname{Im} G_{xx}(z) = \eta \sum_{a=1}^{n+p} \frac{|\boldsymbol{u}_a(x)|^2}{(\mu_a - \tau)^2 + \eta^2} \prec \eta \sum_{a=1}^{n+p} \mathbf{1}(A_a) \frac{|\boldsymbol{u}_a(x)|^2}{(\mu_a - \tau)^2 + \eta^2} \prec \eta$$

for all $z \in \mathbb{H}$ satisfying $\delta \le |z| \le 10$ and $\operatorname{dist}(z, \operatorname{supp} \rho) \ge \varepsilon_*$. Here we used that $A_a := \{\operatorname{dist}(\mu_a, \operatorname{supp} \rho) \le \varepsilon_*/2\}$ occurs a.w.o.p by (4.4.6) and thus $1 - \mathbf{1}(A_a) \prec 0$. In particular, we have $\langle \operatorname{Im} \boldsymbol{g} \rangle \prec \eta$. Now, (4.4.10a) and (4.4.10b) yield

$$\|\boldsymbol{d}(z)\|_\infty \mathbf{1}(\Lambda(z) \le \lambda_*(z)) \prec \frac{1}{\sqrt{n+p}}, \tag{4.4.15a}$$

$$\Lambda_o(z) \mathbf{1}(\Lambda(z) \le \lambda_*(z)) \prec \frac{1}{\sqrt{n+p}}. \tag{4.4.15b}$$

Following the previous argument but using (4.4.15a) and (4.4.15b) instead of (4.4.10a) and (4.4.10b), we obtain (4.4.1b) and (4.4.2b) and this completes the proof of Theorem 4.4.2. □

4.4.2. Local law for Gram matrices.

Proofs of Theorem 4.2.3 and Theorem 4.2.5. Splitting the resolvent of $\boldsymbol{H}$ at $z \in \mathbb{C} \setminus \mathbb{R}$ into blocks

$$\boldsymbol{G}(z) = \begin{pmatrix} \boldsymbol{G}_{11}(z) & \boldsymbol{G}_{12}(z) \\ \boldsymbol{G}_{21}(z) & \boldsymbol{G}_{22}(z) \end{pmatrix}$$

and computing the product $\boldsymbol{G}(z)(\boldsymbol{H} - z)$ blockwise, we obtain that $(XX^* - z^2)^{-1} = \boldsymbol{G}_{11}(z)/z$ and $(X^*X - z^2)^{-1} = \boldsymbol{G}_{22}(z)/z$ for $z \in \mathbb{C} \setminus \mathbb{R}$. Therefore, (4.2.3) follows from (4.4.1) as well as $|z| \ge \delta$ and $m(\zeta) = m_1(\sqrt{\zeta})/\sqrt{\zeta}$ for $\zeta \in \mathbb{H}$.

As $p \sim n$ we obtain

$$|\langle w, \operatorname{diag}(XX^* - \zeta)^{-1} - m(\zeta) \rangle| \lesssim \left| \left\langle (w, 0)^t \, , \, \frac{1}{\sqrt{\zeta}} \left(\boldsymbol{g}(\sqrt{\zeta}) - \boldsymbol{m}(\sqrt{\zeta}) \right) \right\rangle \right|$$

for $w \in \mathbb{C}^p$. Using $p \sim n$, this implies (4.2.4) by (4.4.2). This concludes the proof of Theorem 4.2.3.

Theorem 4.2.5 is a consequence of the corresponding result for $\boldsymbol{H}$, namely Corollary 4.4.4. □

Proof of Theorem 4.2.8. As $m(\zeta) = m_1(\sqrt{\zeta})/\sqrt{\zeta}$ for $\zeta \in \mathbb{H}$, Proposition 4.3.10 implies $|m(\zeta)| \lesssim |\zeta|^{-1/2}$. Thus, $\pi_* = 0$. Recalling $\pi(\omega) = \omega^{-1/2}\rho_1(\omega^{1/2})\mathbf{1}(\omega > 0)$, where ρ_1 is the bounded density representing $\langle m_1 \rangle$, yields

$$\lim_{\omega \downarrow 0} \pi(\omega)\sqrt{\omega} = \frac{1}{\pi}\langle v_1(0)\rangle \in (0, \infty)$$

by (4.3.59a) which proves part (ii) of Theorem 4.2.8.

Since $n = p$, in this case we have $\text{Spec}(XX^*) = \text{Spec}(X^*X)$. Thus, $\langle g_1 \rangle = \langle g_2 \rangle$, i.e., (4.3.42) is fulfilled and Proposition 4.3.10 is applicable.

Using Proposition 4.3.10 instead of Lemma 4.3.9 and following the argument in Subsection 4.4.1, we obtain the same result as Theorem 4.4.2 without the restriction $|z| \geq \delta$. As in the proof of Theorem 4.2.3, we obtain

$$|R_{ij}(\zeta) - \delta_{ij} m_i(\zeta)| \prec \frac{\sqrt{\text{Re}\sqrt{\zeta}}}{|\sqrt{\zeta}|\sqrt{p\text{Im}\,\zeta}} \lesssim \sqrt{\frac{\langle \text{Im}\, m(\zeta)\rangle}{p\text{Im}\,\zeta}}.$$

Here, we deviated from the proof of Theorem 4.2.3 since $|z|$ can be arbitrarily small for $z \in \mathbb{D}_0$ and used part (ii) of Theorem 4.2.8 in the last step. This concludes the proof of part (i) of Theorem 4.2.8.

Consequently, a version of Corollary 4.4.4 for $\delta = 0$ holds true. Then, part (iii) and (iv) of the theorem follow immediately. ☐

4.4.3. Proof of Theorem 4.2.10. In this subsection, we will assume that (A), (C), (D) and (F2) as well as

$$\frac{p}{n} \geq 1 + d_* \tag{4.4.16}$$

for some $d_* > 0$ hold true.

Theorem 4.4.7 (Local law for $\boldsymbol{H}$ around $z = 0$)**.** *If (A), (C), (D), (F2) and* (4.4.16) *hold true, then*

(i) The kernel of $\boldsymbol{H}$ and the kernel of $\boldsymbol{H}^2$ have dimension $p - n$ a.w.o.p.
(ii) There is a $\gamma_* \gtrsim 1$ such that

$$|\mu| \geq \gamma_* \tag{4.4.17}$$

a.w.o.p. for all $\mu \in \text{Spec}(\boldsymbol{H})$ such that $\mu \neq 0$.

(iii) For every $\varepsilon_* > 0$, we have

$$\max_{x,y=1,\dots,n+p} |G_{xy}(z) - \boldsymbol{m}_x(z)\delta_{xy}| \prec \frac{1}{|z|\sqrt{n+p}}, \tag{4.4.18a}$$

$$|\langle \boldsymbol{g} \rangle - \langle \boldsymbol{m} \rangle| \prec \frac{|z|}{n+p}. \tag{4.4.18b}$$

uniformly for $z \in \mathbb{H}$ satisfying $|z| \le \sqrt{\delta_\pi} - \varepsilon_*$.

We will prove that the kernel of $\boldsymbol{H}^2$ has dimension $p - n$ by using a result about the smallest nonzero eigenvalue of XX^* from [73]. Since this result requires the entries of X to have the same variance and a symmetric distribution, in order to cover the general case, we employ a continuity argument which replaces x_{ik}, for definiteness, by centered Gaussians with variance $(n+p)^{-1}$. This will immediately imply Theorem 4.4.7 and consequently Theorem 4.2.10.

We recall the definition of δ_π from (4.2.10) and choose δ_* as in Proposition 4.3.16 for the whole section. Note that $\delta_*^2 \le \delta_\pi$.

Lemma 4.4.8. *If* (4.4.16) *holds true then for all $\delta_1, \delta_2 > 0$ such that $\delta_1 < \delta_2 < \delta_*^2/2$, the matrix $\boldsymbol{H}^2$ has no eigenvalues in $[\delta_1, \delta_2]$ a.w.o.p.*

Proof. Part (iii) of Corollary 4.4.4 with $\delta = \delta_1$ and $\varepsilon_* = \min\{\delta_1, \delta_\pi - \delta_2\}$ implies

$$\#\left(\operatorname{Spec}(\boldsymbol{H}) \cap [\sqrt{\delta}_1, \sqrt{\delta}_2]\right) = 0$$

a.w.o.p. because there is a gap in the support of ρ by part (iii) of Proposition 4.3.16. Since $\operatorname{Spec}(\boldsymbol{H}^2) = \operatorname{Spec}(\boldsymbol{H})^2$ this concludes the proof. ☐

For the remainder of the section, let $\widehat{X} = (\widehat{x}_{ik})_{i=1,\dots,p}^{k=1,\dots,n}$ consist of independent centered Gaussians with $\mathbb{E}|\widehat{x}_{ik}|^2 = (n+p)^{-1}$. We set

$$\widehat{\boldsymbol{H}} := \begin{pmatrix} 0 & \widehat{X} \\ \widehat{X}^* & 0 \end{pmatrix}.$$

Lemma 4.4.9. *If* (4.4.16) *holds true then the kernel of $\widehat{X}\widehat{X}^*$ has dimension $p-n$ a.w.o.p., $\ker(\widehat{X}^*\widehat{X}) = \{0\}$ a.w.o.p. and there is a $\widehat{\gamma} \sim 1$ such that*

$$\widehat{\lambda} \ge \widehat{\gamma} \tag{4.4.19}$$

for all $\widehat{\lambda} \in \operatorname{Spec}(\widehat{X}^*\widehat{X})$.

Proof. Let $\widehat{\lambda}_1 \leq \ldots \leq \widehat{\lambda}_p$ be the eigenvalues of $\widehat{X}\widehat{X}^*$. The assertion will follow once we have established that $\widehat{\lambda}_{p-n+1} \gtrsim 1$ a.w.o.p. since $\widehat{X}\widehat{X}^*$ and $\widehat{X}^*\widehat{X}$ have the same nonzero eigenvalues and $\dim\ker \widehat{X}\widehat{X}^* \geq p-n$ for dimensional reasons. Corollary V.2.1 in [73] implies that $\widehat{\lambda}_{p-n+1} \geq \gamma_- - p^{-2/3+\varepsilon}$ a.w.o.p. for each $\varepsilon > 0$ where $\gamma_- := 1-2\sqrt{pn}/(n+p) \gtrsim 1$, thus $\widehat{\lambda}_{p-n+1} \gtrsim 1$ a.w.o.p. In fact, our proof only requires that $\widehat{\lambda}_{p-n+1} \geq \gamma_- - \varepsilon$ for any $\varepsilon > 0$ a.w.o.p, which already follows from the argument in [133]. ☐

Proof of Theorem 4.4.7. We define $\boldsymbol{H}_t := \sqrt{1-t}\boldsymbol{H} + \sqrt{t}\widehat{\boldsymbol{H}}$ for $t \in [0,1]$ and set $\gamma_* := \min\{\delta_*/2, \sqrt{\widehat{\gamma}}\}$, where $\widehat{\gamma}$ is chosen as in (4.4.19). By Lemma 4.4.8 with $\delta_2 := \gamma_*^2$ and $\delta_1 := \gamma_*^2/2$, $\boldsymbol{H}_t^2$ has no eigenvalues in $[\delta_1, \delta_2]$ a.w.o.p. for every $t \in [0,1]$. Clearly, the eigenvalues of $\boldsymbol{H}_t^2$ depend continuously on t. Therefore, $\#(\operatorname{Spec}(\boldsymbol{H}^2) \cap [0,\delta_1)) = \#(\operatorname{Spec}(\widehat{\boldsymbol{H}}^2) \cap [0,\delta_1))$. Thus, we get the chain of inequalities

$$p-n \leq \dim\ker \boldsymbol{H} = \dim\ker \boldsymbol{H}^2 \leq \#\left(\operatorname{Spec}(\boldsymbol{H}^2) \cap [0,\delta_1)\right) = \#\left(\operatorname{Spec}(\widehat{\boldsymbol{H}}^2) \cap [0,\delta_1)\right)$$
$$= \dim\ker \widehat{\boldsymbol{H}}^2 = p-n.$$

Here we used Lemma 4.4.9 in the last step. As the left and the right-hand-side are equal all of the inequalities are equalities which concludes the proof of part (i) and part (ii).

We will omit the proof of part (iii) of Theorem 4.4.7 as it is very similar to the proof of part (vi) of Theorem 4.2.10 below which will be independent of part (iii) of Theorem 4.4.7. ☐

Proof of Theorem 4.2.10. Since δ_* is chosen as in Proposition 4.3.16 we conclude $\delta_\pi \geq \delta_*^2 \gtrsim 1$ from part (iv) of this proposition. Part (ii) and (iii) of the theorem follow immediately from (4.4.17) in Theorem 4.4.7.

If $p > n$, then $\dim\ker XX^* = p-n$ a.w.o.p. as $p-n \leq \dim\ker XX^* \leq \dim\ker \boldsymbol{H}^2 = p-n$ a.w.o.p by part (i) of Theorem 4.4.7. By Proposition 4.3.16, we obtain $\pi_* = \langle u \rangle = 1-n/p$, where u is defined as in this proposition. This proves part (iv). If $p < n$, then part (v) follows from interchanging the roles of X and X^* and following the same steps as in the proof of part (iv).

For the proof of part (vi), we first assume $p > n$. By Proposition 4.3.16 we can uniquely extend $\zeta m(\zeta) = \sqrt{\zeta} m_1(\sqrt{\zeta})$ to a holomorphic function on $D_{\delta_*^2}(0)$. We fix γ_* as in (4.4.17). On the event $\{\lambda_i \geq \gamma_*^2$ for all $i = p-n+1, \ldots, p\}$, which holds true a.w.o.p. by (4.4.17), the function $\zeta R(\zeta)$ can be uniquely extended to a holomorphic function on $D_{\gamma_*^2}(0)$. We set $\delta := \min\{\gamma_*^2/2, \delta_*^2\}$ and assume without loss of generality that $\delta \leq \delta_\pi - \varepsilon_*$. For $\zeta \in \mathbb{H}$ satisfying $\delta \leq |\zeta| \leq \delta_\pi - \varepsilon_*$, (4.2.11) is immediate from (4.2.3b). We apply (4.2.3b) to obtain $\max_{i,j} |R_{ij}(\zeta) - m_i(\zeta)\delta_{ij}| \prec 1/p$ for $\zeta \in \mathbb{H}$ satisfying $|\zeta| = \delta$. By the symmetry of $R(\zeta)$ and $m(\zeta)$ this estimate holds true for all $\zeta \in \mathbb{C}$ satisfying $|\zeta| = \delta$. Thus, the maximum principle implies that $\max_{i,j} |\zeta R_{ij}(\zeta) - \zeta m_i(\zeta)\delta_{ij}| \prec 1/p$ which proves (4.2.11) since $\{\lambda_i \geq 2\delta$ for all $i = p-n+1, \ldots, p\}$ which holds true a.w.o.p. by $2\delta \leq \gamma_*^2$ and (4.4.17). If $p < n$ then XX^* does not have a kernel a.w.o.p. by (v). Therefore, a similar argument yields (4.2.12).

For the proof of (4.2.13), we observe that $\dim \ker(XX^*) = p\pi_*$ a.w.o.p. in both cases by (iv) and (v). Thus,

$$\frac{1}{p}\sum_{i=1}^{p}[R_{ii}(\zeta) - m_i(\zeta)] = \frac{1}{p}\left(\sum_{j\colon \lambda_j \geq \gamma_*^2} \frac{1}{\lambda_j - \zeta} - \sum_{i=1}^{p} a_i(\sqrt{\zeta})\right)$$

a.w.o.p. for $\zeta \in D_\delta(0)$, δ chosen as above, by (4.4.17), where a is the holomorphic function on $D_{\delta_*}(0)$ defined in Proposition 4.3.16. The right-hand side of the previous equation can therefore be uniquely extended to a holomorphic function on $D_{\delta_*}(0)$. As before, the estimate (4.2.3b) can be extended to $\zeta \in \mathbb{H}$ with $|\zeta| \leq \delta$ by the maximum principle. ☐

The local law for ζ around zero needed an extra argument, Theorem 4.2.10, due to the possible singularity at $\zeta = 0$. We note that this separate treatment is necessary even if $p < n$, in which case XX^* does not have a kernel and $R(\zeta)$ is regular at $\zeta = 0$, since we study XX^* and X^*X simultaneously. Our main stability results are formulated and proven in terms of $\boldsymbol{H}$, as defined in (4.3.1). Therefore, these results are not sensitive to whether p or n is bigger which means whether XX^* has a kernel or X^*X.

4.5. Proof of the Rotation-Inversion lemma

In this section, we prove the Rotation-Inversion lemma, Lemma 4.3.6.

PROOF OF LEMMA 4.3.6. In this proof, we will write $\|A\|$ to denote $\|A\|_2$. Moreover, we introduce a few short hand notations,

$$\mathcal{U} := \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}, \quad \mathcal{A} := \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}, \quad a_\pm := \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 \\ \pm v_2 \end{pmatrix}, \quad \rho := \|A^*A\|^{1/2}.$$

In particular, we have $Av_2 = \rho \mathrm{e}^{\mathrm{i}\psi} v_1$ and $A^* v_1 = \rho \mathrm{e}^{-\mathrm{i}\psi} v_2$ for some $\psi \in \mathbb{R}$. By redefining v_1 to be $\mathrm{e}^{\mathrm{i}\psi} v_1$ we may assume that $\psi = 0$ and get $\mathcal{A} a_\pm = \pm \rho a_\pm$ as well.

Let us check that indeed $\mathcal{U} + \mathcal{A}$ is not invertible if the right-hand side of (4.3.26) is infinite, i.e., if

$$\|A^*A\| \langle v_1 , U_1 v_1 \rangle \langle v_2 , U_2 v_2 \rangle \;=\; 1\,.$$

In this case we find $\|A^*A\| = 1$, $\langle v_1 , U_1 v_1 \rangle = \mathrm{e}^{\mathrm{i}\varphi}$ and $\langle v_2 , U_2 v_2 \rangle = \mathrm{e}^{-\mathrm{i}\varphi}$ for some $\varphi \in \mathbb{R}$. Thus, v_1 and v_2 are eigenvectors of U_1 and U_2, respectively. Therefore, both $\mathcal{U}$ and $\mathcal{A}$ leave the 2-dimensional subspace spanned by $(v_1, 0)$ and $(0, v_2)$ invariant and in this basis the restriction of $\mathcal{U} + \mathcal{A}$ is represented by the 2×2-matrix

$$\begin{pmatrix} \mathrm{e}^{\mathrm{i}\varphi} & 1 \\ 1 & \mathrm{e}^{-\mathrm{i}\varphi} \end{pmatrix},$$

which is not invertible.

We will now show that in every other case $\mathcal{U} + \mathcal{A}$ is invertible and its inverse satisfies (4.3.26). To this end we will derive a lower bound on $\|(\mathcal{U} + \mathcal{A})w\|$ for an arbitrary normalized vector $w \in \mathbb{C}^{n+p}$. Any such vector admits a decomposition,

$$w \;=\; \alpha_+ a_+ + \alpha_- a_- + \beta b\,,$$

where $\alpha_\pm \in \mathbb{C}$, $\beta \geq 0$ and b is a normalized vector in the orthogonal complement of the 2-dimensional space spanned by a_+ and a_-. The normalization of w implies

$$|\alpha_+|^2 + |\alpha_-|^2 + \beta^2 \;=\; 1\,. \tag{4.5.1}$$

The case $\beta = 1$ is trivial because the spectral gap of A^*A implies a spectral gap of $\mathcal{A}$ in the sense that

$$\operatorname{Spec}(\mathcal{A}/\rho) \;\subseteq\; \{-1\} \cup [\,-1 + \rho^{-2}\operatorname{Gap}(AA^*), 1 - \rho^{-2}\operatorname{Gap}(AA^*)\,] \cup \{1\}\,. \tag{4.5.2}$$

Thus, we will from now on assume $\beta < 1$.

We will use the notations $\mathcal{P}_\parallel$ and $\mathcal{P}_\perp$ for the orthogonal projection onto the 2-dimensional subspace spanned by $a_\pm$ and its orthogonal complement, respectively. We also introduce

$$\begin{aligned} \lambda &:= \frac{1}{2}\frac{|\alpha_+ + \alpha_-|^2}{|\alpha_+|^2 + |\alpha_-|^2} \in [0,1]\,, \\ \kappa &:= (|\alpha_+|^2 + |\alpha_-|^2)^{-1/2}\|\mathcal{P}_\parallel(1+\mathcal{U}^*\mathcal{A})(\alpha_+ a_+ + \alpha_- a_-)\|\,. \end{aligned} \tag{4.5.3}$$

With this notation we will now prove

$$\|(\mathcal{U}+\mathcal{A})w\| \;\geq\; c_1 \operatorname{Gap}(AA^*)\,\kappa\,, \tag{4.5.4}$$

for some positive numerical constant c_1. The analysis is split into the following regimes:

Regime 1: $\kappa^{1/2} < 10\beta$,
Regime 2: $\kappa^{1/2} \geq 10\beta$ and $\lambda < 1/10$,
Regime 3: $\kappa^{1/2} \geq 10\beta$ and $\lambda > 9/10$,
Regime 4: $\kappa^{1/2} \geq 10\beta$ and $1/10 \leq \lambda \leq 9/10$ and $|\langle v_1\,, U_1 v_1\rangle|^2 + |\langle v_2\,, U_2 v_2\rangle|^2 \leq 2 - \kappa/2$,
Regime 5: $\kappa^{1/2} \geq 10\beta$ and $1/10 \leq \lambda \leq 9/10$ and $|\langle v_1\,, U_1 v_1\rangle|^2 + |\langle v_2\,, U_2 v_2\rangle|^2 > 2 - \kappa/2$.

These regimes can be chosen more carefully in order to optimize the constant c_1 in (4.5.4), but we will not do that here.

Regime 1: In this regime we make use of the spectral gap of A^*A by simply using the triangle inequality,

$$\|(\mathcal{U}+\mathcal{A})w\| \;\geq\; \|w\| - \|\mathcal{A}w\| \;=\; 1 - \sqrt{\rho^2|\alpha_+|^2 + \rho^2|\alpha_-|^2 + \beta^2\|\mathcal{A}b\|^2}.$$

We use the inequality $1-\sqrt{1-\tau} \geq \tau/2$ for $\tau \in [0,1]$ as well as the normalization (4.5.1) and find

$$2\|(\mathcal{U}+\mathcal{A})w\| \;\geq\; 1 - \rho^2 + \rho^2\beta^2 - \beta^2\|\mathcal{A}b\|^2 \;\geq\; \rho\beta^2(\rho - \|\mathcal{A}b\|) \;\geq\; \beta^2 \operatorname{Gap}(AA^*)\,.$$

The last inequality follows from (4.5.2) and because b is orthogonal to $a_\pm$. Since $\beta^2 \geq \kappa/100$, we conclude that in the first regime (4.5.4) is satisfied.

Regime 2: In this regime we project on the second component of $(\mathcal{U}+\mathcal{A})w$.

$$\begin{aligned}\sqrt{2}\|(\mathcal{U}+\mathcal{A})w\| &\geq \|(\alpha_+-\alpha_-)U_2v_2+\sqrt{2}\beta U_2b_2-(\alpha_++\alpha_-)A^*v_1-\sqrt{2}\beta A^*b_1\| \\ &\geq |\alpha_+-\alpha_-|\|U_2v_2\|-\sqrt{2}\beta\|U_2b_2\|-\rho|\alpha_++\alpha_-|\|v_2\|-\sqrt{2}\beta\|A^*b_1\| \\ &\geq \sqrt{2}\sqrt{|\alpha_+|^2+|\alpha_-|^2}(\sqrt{1-\lambda}-\sqrt{\lambda})-2\sqrt{2}\beta\,.\end{aligned}$$

Here we used the notation $b=(b_1,b_2)$ for the components of b. The last inequality holds by the normalization of v_2 and b, by $\rho\leq 1$ and by the definition of λ from (4.5.3), which also implies

$$|\alpha_+-\alpha_-|^2 \,=\, 2(1-\lambda)(|\alpha_+|^2+|\alpha_-|^2)\,.$$

Since $\lambda<1/4$ in this regime and $\kappa\leq 2$ by the definition of κ in (4.5.3) we find $\beta\leq\kappa^{1/2}/10\leq 1/5$ and infer

$$\|(\mathcal{U}+\mathcal{A})w\| \,\geq\, \sqrt{1-\beta^2}(\sqrt{1-\lambda}-\sqrt{\lambda})-2\beta \,\geq\, 1/10 \,\geq\, \kappa/20\,.$$

Regime 3: By the symmetry in $a_\pm$ and $\alpha_\pm$ and therefore in λ and $1-\lambda$ this regime is treated in the same way as Regime 2 by estimating the norm of the first component of $(\mathcal{U}+\mathcal{A})w$ from below.

Regime 4: Here we project onto the orthogonal complement of the subspace spanned by a_+ and a_-,

$$\|(\mathcal{U}+\mathcal{A})w\| \,\geq\, \|\mathcal{P}_\perp(\mathcal{U}+\mathcal{A})w\| \,\geq\, \|\mathcal{P}_\perp\mathcal{U}(\alpha_+a_++\alpha_-a_-)\|-\beta\|\mathcal{P}_\perp(\mathcal{U}+\mathcal{A})b\|\,. \quad (4.5.5)$$

We compute the first term in this last expression more explicitly,

$$\begin{aligned}\|\mathcal{P}_\perp\mathcal{U}(\alpha_+a_++\alpha_-a_-)\|^2 &= \|\alpha_+a_++\alpha_-a_-\|^2-\|\mathcal{P}_\parallel\mathcal{U}(\alpha_+a_++\alpha_-a_-)\|^2 \\ &= |\alpha_+|^2+|\alpha_-|^2-\frac{1}{2}|\alpha_++\alpha_-|^2|\langle v_1\,,U_1v_1\rangle|^2 \\ &\quad -\frac{1}{2}|\alpha_+-\alpha_-|^2|\langle v_2\,,U_2v_2\rangle|^2 \\ &= (1-\beta^2)(1-\lambda|\langle v_1\,,U_1v_1\rangle|^2-(1-\lambda)|\langle v_2\,,U_2v_2\rangle|^2)\,.\end{aligned} \quad (4.5.6)$$

For the second equality we used that

$$\|\mathcal{P}_\parallel u\|^2 \,=\, |\langle v_1\,,u_1\rangle|^2+|\langle v_2\,,u_2\rangle|^2, \qquad u \,=\, (u_1,u_2)\in\mathbb{C}^{p+n}.$$

With the choice of variables

$$\xi := |\langle v_1 , U_1 v_1\rangle|^2 , \qquad \eta := |\langle v_2 , U_2 v_2\rangle|^2 ,$$

we are minimizing the last line in (4.5.6) under the restrictions that are satisfied in this regime,

$$\min\{1 - \lambda\xi - (1-\lambda)\eta : \ \xi, \eta \in [0,1] , \ 2\xi + 2\eta \le 4 - \kappa\} \ \ge \ \frac{1}{2}\kappa \min\{1-\lambda, \lambda\} .$$

We use the resulting estimate in (4.5.5) and in this way we arrive at

$$\|(\mathcal{U} + \mathcal{A})w\| \ \ge \ \frac{1}{\sqrt{2}}\kappa^{1/2}\sqrt{1-\beta^2}\,\min\{1-\lambda, \lambda\}^{1/2} - 2\beta \ \ge \ \frac{\kappa^{1/2}}{100} \ \ge \ \frac{\kappa}{200} .$$

In the second to last inequality we used $\beta \le 1/5$ which was already established in the consideration of Regime 2 and in the last inequality we used $\kappa \le 2$.

Regime 5: In this regime we project onto the span of a_+ and a_-,

$$\begin{aligned} \|(\mathcal{U} + \mathcal{A})w\| \ &= \ \|(1 + \mathcal{U}^*\mathcal{A})w\| \\ &\ge \ \|\mathcal{P}_{\|}(1 + \mathcal{U}^*\mathcal{A})(\alpha_+ a_+ + \alpha_- a_-)\| - \beta \|\mathcal{P}_{\|}(1 + \mathcal{U}^*\mathcal{A})b\| \\ &= \ \sqrt{|\alpha_+|^2 + |\alpha_-|^2}\,\kappa - \beta\,\|\mathcal{P}_{\|}\mathcal{U}^*\mathcal{A}b\| . \end{aligned} \tag{4.5.7}$$

The second term in the last line is estimated by using

$$\|\mathcal{P}_{\|}\mathcal{U}^*\mathcal{A}b\|^2 \ \le \ \|\mathcal{A}b\| \sup_{h \| a_\pm} \sup_{u \perp a_\pm} |\langle h , \mathcal{U}^* u\rangle|^2 ,$$

where the suprema are taken over normalized vectors h and u in the 2-dimensional subspace spanned by $a_\pm$ and its orthogonal complement, respectively. First we perform the supremum over h and get

$$\begin{aligned} \|\mathcal{P}_{\|}\mathcal{U}^*\mathcal{A}b\|^2 \ &\le \ \sup_{u \perp a_\pm} \left(|\langle v_1 , U_1^* u_1\rangle|^2 + |\langle v_2 , U_2^* u_2\rangle|^2\right) \\ &\le \ \sup_{u_1 \perp v_1} |\langle v_1 , U_1^* u_1\rangle|^2 + \sup_{u_2 \perp v_2} |\langle v_2 , U_2^* u_2\rangle|^2, \end{aligned} \tag{4.5.8}$$

where the vectors $u_1 \in \mathbb{C}^p$ and $u_2 \in \mathbb{C}^n$ are normalized. Computing

$$\sup_{u_1 \perp v_1} |\langle v_1 , U_1^* u_1\rangle|^2 \ = \ 1 - |\langle v_1 , U_1 v_1\rangle|^2 , \qquad \sup_{u_2 \perp v_2} |\langle v_2 , U_2^* u_2\rangle|^2 \ = \ 1 - |\langle v_2 , U_2 v_2\rangle|^2 ,$$

we get

$$\|\mathcal{P}_{\|}\mathcal{U}^*\mathcal{A}b\|^2 \,\le\, 2-|\langle v_1\,,U_1v_1\rangle|^2-|\langle v_2\,,U_2v_2\rangle|^2 \,\le\, \kappa/2\,,$$

where we used the choice of Regime 5 in the last step. Plugging this bound into (4.5.7) and using $\beta \le \kappa^{1/2}/10$ as well as $\beta \le 1/5$ yields

$$\|(\mathcal{U}+\mathcal{A})w\| \,\ge\, \sqrt{1-\beta^2}\,\kappa-\beta\,\kappa^{1/2} \,\ge\, \kappa/2\,.$$

This finishes the proof of (4.5.4). In order to show (4.3.26), and thus the lemma, we notice that

$$\kappa \,\ge\, \inf_{u\|a_\pm}\|\mathcal{P}_{\|}(1+\mathcal{U}^*\mathcal{A})u\|\,,$$

where the infimum is taken over normalized vectors u in the span of a_+ and a_-. Thus, it suffices to estimate the norm of the inverse of $\mathcal{P}_{\|}(1+\mathcal{U}^*\mathcal{A})\mathcal{P}_{\|}$, restricted to the 2-dimensional subspace with orthonormal basis $(v_1,0)$ and $(0,v_2)$. In this basis this linear operator takes the form of the simple 2×2-matrix,

$$\begin{pmatrix} 1 & \rho\langle v_1\,,U_1v_1\rangle \\ \rho\langle v_2\,,U_2v_2\rangle & 1 \end{pmatrix}.$$

Its inverse is bounded by the right-hand side of (4.3.26), up to the factor $\mathrm{Gap}(AA^*)$ that we encountered in (4.5.4), and the lemma is proven. □

CHAPTER 5

Singularities of the density of states of random Gram matrices

In this chapter, we present the results from [11]. For large random matrices X with independent, centered entries but not necessarily identical variances, the eigenvalue density of XX^* is well approximated by a deterministic measure on $\mathbb{R}$. We show that the density of this measure has only square and cubic-root singularities away from zero. We also extend the bulk local law in Chapter 4 (cf. [14]) to the vicinity of these singularities.

5.1. Introduction

The empirical eigenvalue density or *density of states* of many large random matrices is well approximated by a deterministic probability measure, the *self-consistent density of states*. If X is a $p \times n$ random matrix with independent, centered entries of identical variances then the limit of the eigenvalue density of the *sample covariance matrix* XX^* for large p and n with p/n converging to a constant has been identified by Marchenko and Pastur in [112]. However, some applications in wireless communication require understanding the spectrum of XX^* without the assumption of identical variances of the entries of $X = (x_{kq})_{k,q}$ [52, 92, 150]. In this case, the matrix XX^* is a *random Gram matrix*.

For constant variances, the self-consistent density of states is obtained by solving a scalar equation for its Stieltjes transform, the *scalar Dyson equation*. In case the variances $s_{kq} := \mathbb{E}|x_{kq}|^2$ depend nontrivially on k and q, the self-consistent density of states is obtained from the solution $m(\zeta) = (m_1(\zeta), \ldots, m_p(\zeta)) \in \mathbb{H}^p$ of the *vector Dyson equation* [82]

$$-\frac{1}{m_k(\zeta)} = \zeta - \sum_{q=1}^{n} s_{kq}\Big(1 + \sum_{l=1}^{p} s_{lq} m_l(\zeta)\Big)^{-1} \qquad \text{for all } k \in [p], \tag{5.1.1}$$

for all $\zeta \in \mathbb{H}$. Here, we introduced $\mathbb{H} := \{\zeta \in \mathbb{C} \colon \operatorname{Im}\zeta > 0\}$ and $[p] := \{1, \ldots, p\}$. Indeed, the average $\langle m(\zeta)\rangle_1 := p^{-1}\sum_{k=1}^{p} m_k(\zeta)$ is the Stieltjes transform of the self-consistent density of states denoted by $\langle \nu \rangle_1$. If the limit of $\langle \nu \rangle_1$ as $p, n \to \infty$ exists then it can be studied via an infinite-dimensional version of (5.1.1) (see (5.2.3) below).

For Wigner-type matrices, i.e., Hermitian random matrices with independent (up to the Hermiticity constraint), centered entries, the analogue of (5.1.1) is a quadratic vector equation (QVE) in the language of [4, 5]. In these papers, finite and infinite-dimensional versions of the QVE have been extensively studied to analyze the self-consistent density of states whose Stieltjes transform is the average of the solution to the QVE. The authors show that the self-consistent density of states has a 1/3-Hölder continuous density. Except for finitely many square-root and cubic-root singularities this density is real-analytic. The square-root behaviour emerges solely at the edges of the connected components of the support of the self-consistent density of states, whereas the cubic-root singularities lie inside these components. The detailed stability analyis in [4] is then used in [7] to obtain the local law for Wigner-type matrices. A *local law* typically refers to a statement about the convergence of the eigenvalue density to a deterministic measure on a scale slightly above the typical local eigenvalue spacing.

For the Dyson equation for random Gram matrices, we obtain away from $\zeta = 0$ the same results as mentioned above in the QVE setup. Furthermore, we extend our local law for random Gram matrices in Chapter 4 (cf. [14]) to the vicinity of the singularities of the self-consistent density of states. This can be seen as another instance of the universality phenomenon in random matrix theory. Despite the different structure of Gram and Wigner-type matrices, the densities of states of these Hermitian random matrices have the same types of singularities. We refer to Chapter 4 and the references therein for related results about random Gram matrices.

There is a close connection between Gram and Wigner-type matrices. The Dyson equation, (5.1.1), can be transformed into a QVE in the sense of [4] and the spectrum of XX^* is closely related to that of a Wigner-type matrix in the sense of [7]. This is easiest explained on the random matrix level through a special case of the linearization tricks: If X has independent and centered entries then the random matrix

$$\boldsymbol{H} = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix} \tag{5.1.2}$$

is a Wigner-type matrix and the spectra of $\boldsymbol{H}^2$ and XX^* agree away from zero. Therefore, instead of trying to analyze (5.1.1) and XX^* directly, it is more efficient to study the corresponding QVE and Wigner-type matrix as in Chapter 4. However, owing to the large zero blocks in $\boldsymbol{H}$, its variance matrix is not uniformly primitive (see **A3** in [4]), a key assumption for the analysis in [4]. Indeed, the stability operator of the QVE possesses an additional unstable direction $\boldsymbol{f}_-$, which has to be treated separately. In Chapter 4, this study has been conducted in the bulk spectrum and away from the support of $\langle \nu \rangle_1$, where $\boldsymbol{f}_-$ did not play an important role at least away from zero.

In this note, we present a new argument needed in the analysis of the cubic equation (see (5.3.18) below) describing the stability of the QVE close to its singularities in order to incorporate the additional unstable direction. In fact, the analysis of the cubic equation in [4] heavily relies on the uniform primitivity of the variance matrix. Adapting this argument to the current setup cannot exclude that the coefficients of the cubic and the quadratic term in the cubic equation vanish at the same time due to the presence of $\boldsymbol{f}_-$. A nonvanishing cubic or quadratic coefficient is however absolutely crucial for the cubic stability analysis in [4]. Otherwise not only square-root or cubic-root but also higher order singularities would emerge. Our main novel ingredient, a very detailed analysis of these coefficients, actually excludes this scenario. With this essential new input, the regularity and the singularity structure of (5.1.1) as well as the local law for XX^* follow by correctly combining the arguments in [4, 7] and Chapter 4.

Acknowledgement. The author is very grateful to László Erdős for many fruitful discussions and many valuable suggestions. The author would also like to thank Torben Krüger for several helpful conversations.

5.2. Main results

5.2.1. Structure of the solution to the Dyson equation. Let $(\mathfrak{X}_1, \mathcal{S}_1, \pi_1)$ and $(\mathfrak{X}_2, \mathcal{S}_2, \pi_2)$ be two finite measure spaces such that $\pi_1(\mathfrak{X}_1)$ and $\pi_2(\mathfrak{X}_2)$ are strictly positive. Moreover, we denote the spaces of bounded and measurable functions on $\mathfrak{X}_1$ and $\mathfrak{X}_2$ by

$$\mathscr{B}_i := \left\{ u \colon \mathfrak{X}_i \to \mathbb{C} : \|u\|_\infty := \sup_{x \in \mathfrak{X}_i} |u(x)| < \infty \right\}$$

for $i = 1, 2$. We consider $\mathscr{B}_1$ and $\mathscr{B}_2$ equipped with the supremum norm $\|\cdot\|_\infty$. We denote the induced operator norms by $\|\cdot\|_{\mathscr{B}_1 \to \mathscr{B}_2}$ and $\|\cdot\|_{\mathscr{B}_2 \to \mathscr{B}_1}$. For $u \in \mathscr{B}_1$, we write $u_k = u(k)$ for $k \in \mathfrak{X}_1$. We use the same notation for $v \in \mathscr{B}_2$.

Let $s\colon \mathfrak{X}_1 \times \mathfrak{X}_2 \to \mathbb{R}_0^+, s(k,q) = s_{kq}$ be a measurable nonnegative function such that

$$\sup_{k \in \mathfrak{X}_1} \int_{\mathfrak{X}_2} s_{kq} \pi_2(\mathrm{d}q) < \infty, \qquad \sup_{q \in \mathfrak{X}_2} \int_{\mathfrak{X}_1} s_{kq} \pi_1(\mathrm{d}k) < \infty. \tag{5.2.1}$$

We define the bounded linear operators $S\colon \mathscr{B}_2 \to \mathscr{B}_1$ and $S^t\colon \mathscr{B}_1 \to \mathscr{B}_2$ through

$$(Sv)_k = \int_{\mathfrak{X}_2} s_{kr} v_r \pi_2(\mathrm{d}r), \quad k \in \mathfrak{X}_1,\ v \in \mathscr{B}_2, \quad (S^t u)_q = \int_{\mathfrak{X}_1} s_{lq} u_l \pi_1(\mathrm{d}l), \quad q \in \mathfrak{X}_2,\ u \in \mathscr{B}_1. \tag{5.2.2}$$

We are interested in the solution $m\colon \mathbb{H} \to \mathscr{B}_1$ of the *Dyson equation*

$$-\frac{1}{m(\zeta)} = \zeta - S\frac{1}{1 + S^t m(\zeta)}, \tag{5.2.3}$$

for $\zeta \in \mathbb{H}$, which satisfies $\operatorname{Im} m(\zeta) > 0$ for all $\zeta \in \mathbb{H}$.

Proposition 5.2.1 (Existence and Uniqueness)**.** *If* (5.2.1) *holds true then there is a unique function* $m\colon \mathbb{H} \to \mathscr{B}_1$ *satisfying* (5.2.3) *and* $\operatorname{Im} m(\zeta) > 0$ *for all* $\zeta \in \mathbb{H}$*. Moreover,* $m\colon \mathbb{H} \to \mathscr{B}_1$ *is analytic. For each* $k \in \mathfrak{X}_1$*, there is a unique probability measure* ν_k *on* $\mathbb{R}$ *such that* m_k *is the Stieltjes transform of* ν_k*, i.e.,*

$$m_k(\zeta) = \int_0^\infty \frac{1}{E - \zeta} \nu_k(\mathrm{d}E) \tag{5.2.4}$$

for all $\zeta \in \mathbb{H}$*. The support of* ν_k *is contained in* $[0, \Sigma]$ *for each* $k \in \mathfrak{X}_1$*, where*

$$\Sigma := 4 \max\left\{ \|S\|_{\mathscr{B}_2 \to \mathscr{B}_1}, \|S^t\|_{\mathscr{B}_1 \to \mathscr{B}_2} \right\}. \tag{5.2.5}$$

Further assumptions on π_1, π_2 and S will yield a more detailed understanding of the measures ν_k. To formulate these assumptions, we introduce the averages of $u \in \mathscr{B}_1$ and $v \in \mathscr{B}_2$ through

$$\langle u \rangle_1 = \frac{1}{\pi_1(\mathfrak{X}_1)} \int_{\mathfrak{X}_1} u_k \pi_1(\mathrm{d}k), \qquad \langle v \rangle_2 = \frac{1}{\pi_2(\mathfrak{X}_2)} \int_{\mathfrak{X}_2} v_q \pi_2(\mathrm{d}q).$$

Additionally, we set $\|u\|_t := \langle |u|^t \rangle_1^{1/t}$ and $\|v\|_t := \langle |v|^t \rangle_2^{1/t}$ for $u \in \mathscr{B}_1$, $v \in \mathscr{B}_2$ and $t \geq 1$. Moreover, for $k \in \mathfrak{X}_1$ and $q \in \mathfrak{X}_2$, we define the functions $S_k\colon \mathfrak{X}_2 \to \mathbb{R}_0^+,\ S_k(r) = s_{kr}$

and $(S^t)_q\colon \mathfrak{X}_1 \to \mathbb{R}_0^+$, $(S^t)_q(l) = s_{lq}$. We call S_k and $(S^t)_q$ the *rows* and *columns* of S, respectively.

Assumptions 5.2.2. **(A1)** The total measures $\pi_1(\mathfrak{X}_1)$ and $\pi_2(\mathfrak{X}_2)$ are comparable, i.e., there are constants $0 < \pi_* < \pi^*$ such that

$$\pi_* \le \frac{\pi_1(\mathfrak{X}_1)}{\pi_2(\mathfrak{X}_2)} \le \pi^*.$$

(A2) The operators S and S^t are irreducible in the sense that there are L_1, $L_2 \in \mathbb{N}$ and $\kappa_1, \kappa_2 > 0$ such that

$$\Big((SS^t)^{L_1} u\Big)_k \ge \kappa_1 \langle u\rangle_1, \quad \Big((S^tS)^{L_2} v\Big)_q \ge \kappa_2 \langle v\rangle_2,$$

for all $u \in \mathscr{B}_1$, $v \in \mathscr{B}_2$ satisfying $u \ge 0$ and $v \ge 0$ and for all $k \in \mathfrak{X}_1$, $q \in \mathfrak{X}_2$.

(A3) The rows and columns of S are sufficiently close to each other in the sense that there is a continuous strictly monotonically decreasing function $\gamma\colon (0,1] \to \mathbb{R}_0^+$ such that $\lim_{\varepsilon\downarrow 0}\gamma(\varepsilon) = \infty$ and for all $\varepsilon \in (0,1]$, we have

$$\gamma(\varepsilon) \le \min\Big\{ \inf_{k\in\mathfrak{X}_1} \frac{1}{\pi_1(\mathfrak{X}_1)} \int_{\mathfrak{X}_1} \frac{\pi_1(\mathrm{d}l)}{\varepsilon + \|S_k - S_l\|_2^2}, \inf_{q\in\mathfrak{X}_2} \frac{1}{\pi_2(\mathfrak{X}_2)} \int_{\mathfrak{X}_2} \frac{\pi_2(\mathrm{d}r)}{\varepsilon + \|(S^t)_q - (S^t)_r\|_2^2}\Big\}.$$

(A4) The operators S and S^t map square-integrable functions continuously to bounded functions, i.e., there are constants $\Psi_1, \Psi_2 > 0$ such that

$$\|S\|_{L^2(\pi_2/\pi_2(\mathfrak{X}_2))\to\mathscr{B}_1} \le \Psi_1, \qquad \|S^t\|_{L^2(\pi_1/\pi_1(\mathfrak{X}_1))\to\mathscr{B}_2} \le \Psi_2.$$

Our estimates will be uniform in all models that satisfy Assumptions 5.2.2 with the same constants. Therefore, the constants π_*, π^* from **(A1)**, L_1, L_2, κ_1, κ_2 from **(A2)**, the function γ from **(A3)** and Ψ_1, Ψ_2 from **(A4)** are called *model parameters*. We refer to Remark 5.2.4 below for an easily checkable sufficient condition for **(A3)**. We now state our main result about the regularity and the possible singularities of ν_k defined in (5.2.4).

Theorem 5.2.3. *If we assume **(A1)** – **(A4)** then the following statements hold true:*

(i) (Regularity of ν) There are $\nu^0 \in \mathscr{B}_1$ and $\nu^d\colon \mathfrak{X}_1 \times (0,\infty) \to [0,\infty)$, $(k,E) \mapsto \nu_k^d(E)$ such that

$$\nu_k(\mathrm{d}E) = \nu_k^0 \delta_0(\mathrm{d}E) + \nu_k^d(E)\mathrm{d}E \tag{5.2.6}$$

for all $k \in \mathfrak{X}_1$*. For all* $\delta > 0$*, the function* ν^d *is uniformly* $1/3$*-Hölder continuous on* $[\delta, \infty)$*, i.e.,*

$$\sup_{k \in \mathfrak{X}_1} \sup_{E_1 \neq E_2,\, E_1, E_2 \geq \delta} \frac{|\nu_k^d(E_1) - \nu_k^d(E_2)|}{|E_1 - E_2|^{1/3}} < \infty.$$

For all $k \in \mathfrak{X}_1$*, we have*

$$\{E \in (0,\infty)\colon \langle \nu^d(E)\rangle > 0\} = \{E \in (0,\infty)\colon \nu_k^d(E) > 0\}.$$

We set $\mathfrak{P} := \{E \in (0,\infty)\colon \langle \nu^d(E)\rangle > 0\}$*. For each* $\delta > 0$*, the set* $\mathfrak{P} \cap (\delta, \infty)$ *is a finite union of open intervals. The map* $\nu^d\colon (0,\infty) \setminus \partial\mathfrak{P} \to \mathscr{B}_1$ *is real-analytic. There is* $\rho_* > 0$ *depending only on the model parameters and* δ *such that the Lebesgue measure of each connected component of* $\mathfrak{P} \cap (\delta, \infty)$ *is at least* $2\rho_*$*.*

(ii) (Singularities of ν^d*) Fix* $\delta > 0$*. For any* $E_0 \in (\partial\mathfrak{P}) \cap (\delta, \infty)$*, there are two cases*

CUSP: *The point* E_0 *is the intersection of the closures of two connected components of* $\mathfrak{P} \cap (\delta, \infty)$ *and* ν^d *has a cubic root singularity at* E_0*, i.e., there is* $c \in \mathscr{B}_1$ *satisfying* $\inf_{k \in \mathfrak{X}_1} c_k > 0$ *such that*

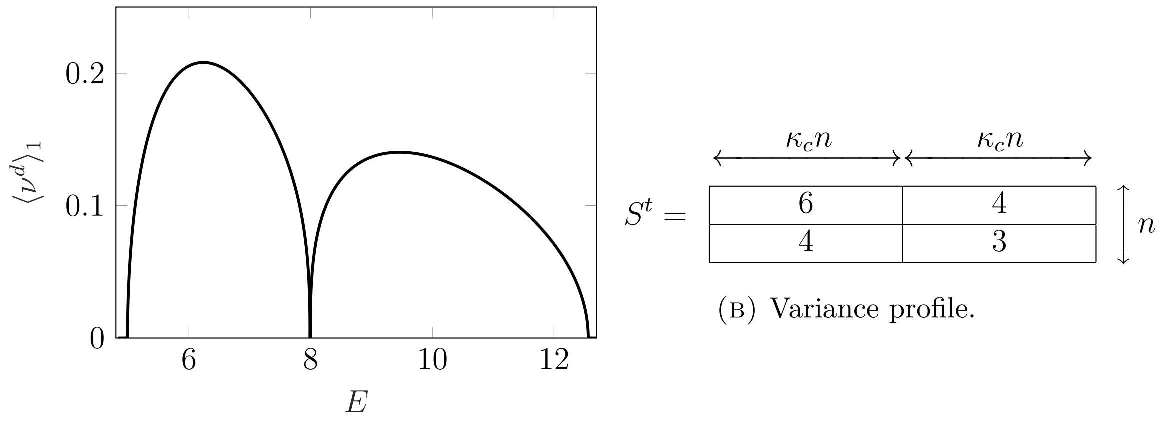
$$\nu_k^d(E_0 + \lambda) = c_k |\lambda|^{1/3} + \mathcal{O}(|\lambda|^{2/3}), \qquad \lambda \to 0.$$

EDGE: *The point* E_0 *is the left or right endpoint of a connected component of* $\overline{\mathfrak{P}} \cap (\delta, \infty)$ *and* ν^d *has a square root singularity at* E_0*, i.e., there is* $c \in \mathscr{B}_1$ *satisfying* $\inf_{k \in \mathfrak{X}_1} c_k > 0$ *such that*

$$\nu_k^d(E_0 + \theta\lambda) = c_k \lambda^{1/2} + \mathcal{O}(\lambda), \qquad \lambda \downarrow 0,$$

where $\theta = +1$ *if* E_0 *is a left endpoint of* $\mathfrak{P}$ *and* $\theta = -1$ *if* E_0 *is a right endpoint.*

In Figure 5.1, we present an example of a self-consistent density of states $\langle \nu^d \rangle_1$ for $\mathfrak{X}_1 = [\kappa_c n]$ and $\mathfrak{X}_2 = [n]$ with $\kappa_c > 0$ and $n \in \mathbb{N}$. If π_1 and π_2 are the (unnormalized) counting measures on $\mathfrak{X}_1$ and $\mathfrak{X}_2$, respectively, and κ_c is chosen suitably then we obtain Figure 5.1 with a cubic cusp at $E \approx 8$.



(A) Self-consistent density of states $\langle \nu^d \rangle_1$.

(B) Variance profile.

FIGURE 5.1. Example of a self-consistent density of states with variance profile S. It has square-root edges at the left and right endpoint of its support and a cubic cusp at $E \approx 8$.

Remark 5.2.4 (Piecewise Hölder-continuous rows and columns of S imply **(A3)**)**.** Let $\mathfrak{X}_1$ and $\mathfrak{X}_2$ be two nontrivial compact intervals in $\mathbb{R}$ and π_1 and π_2 the Lebesgue measures. In this case, **(A3)** holds true if the maps $k \mapsto S_k$ and $r \mapsto (S^t)_r$ are piecewise 1/2-Hölder continuous in the sense that there are two finite partitions $(I_\alpha)_{\alpha \in A}$ and $(J_\beta)_{\beta \in B}$ of $\mathfrak{X}_1$ and $\mathfrak{X}_2$, respectively, such that, for all $\alpha \in A$ and $\beta \in B$, we have

$$\|S_k - S_l\|_2 \leq C_\alpha |k - l|^{1/2}, \qquad \|(S^t)_q - (S^t)_r\|_2 \leq D_\beta |q - r|^{1/2}$$

for all $k, l \in I_\alpha$ and for all $q, r \in J_\beta$. There is a similar condition for **(A3)** if $\mathfrak{X}_1 = [p]$ and $\mathfrak{X}_2 = [n]$ for some $p, n \in \mathbb{N}$ and the measures π_1 and π_2 are the (unnormalized) counting measures on $[p]$ and $[n]$, respectively.

5.2.2. Local law for random Gram matrices. In this subsection, we state our results on random Gram matrices. We now set $\mathfrak{X}_1 = [p]$, $\mathfrak{X}_2 = [n]$ as well as π_1 and π_2 the (unnormalized) counting measures on $[p]$ and $[n]$, respectively. In particular, $\pi_1(\mathfrak{X}_1) = p$ and $\pi_2(\mathfrak{X}_2) = n$.

Assumptions 5.2.5. Let $X = (x_{kq})_{k,q}$ be a $p \times n$ random matrix with independent, centered entries and variance matrix $S = (s_{kq})_{k,q}$, i.e., $\mathbb{E}\, x_{kq} = 0$ and $s_{kq} := \mathbb{E}|x_{kq}|^2$ for $k \in [p]$, $q \in [n]$. Moreover, we assume that **(A1)**, **(A2)** and **(A3)** in Assumptions 5.2.2 and the following conditions are satisfied.

(B1) The variances are bounded in the sense that there exists $s^* > 0$ such that

$$s_{kq} \leq \frac{s^*}{p+n} \qquad \text{for } k \in [p],\ q \in [n].$$

(B2) All entries of X have bounded moments in the sense that there are $\mu_m > 0$ for $m \geq 3$ such that

$$\mathbb{E}|x_{kq}|^m \leq \mu_m s_{kq}^{m/2} \qquad \text{for all } k \in [p],\ q \in [n].$$

The sequence $(\mu_m)_{m\geq 3}$ in **(B2)** is also considered a model parameter.

Since **(B1)** implies **(A4)**, we can apply Theorem 5.2.3. By its first part, for every $\delta > 0$, there are $\alpha_1, \ldots, \alpha_K$, $\beta_1, \ldots, \beta_K \in [\delta, \infty)$ for some $K \in \mathbb{N}$ such that

$$\operatorname{supp} \left\langle \nu^d|_{[\delta,\infty)} \right\rangle_1 = \bigcup_{i=1}^K [\alpha_i, \beta_i], \qquad \alpha_j < \beta_j < \alpha_{j+1}$$

and $\rho_* > 0$ depending only on the model parameters and δ such that $\beta_i - \alpha_i \geq 2\rho_*$ for all $i \in [K]$. For $\rho \in [0, \rho_*)$, we introduce the *local gap size* Δ_ρ via

$$\Delta_\rho(E) := \begin{cases} \alpha_{i+1} - \beta_i, & \text{if } \beta_i - \rho \leq E \leq \alpha_{i+1} + \rho \text{ for some } i \in [K], \\ 1, & \text{if } E \leq \alpha_1 + \rho \text{ or } E \geq \beta_K - \rho, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2.7}$$

For $\delta, \gamma > 0$, we define the spectral domain $\mathbb{D}_{\delta,\gamma} := \{\zeta \in \mathbb{H} \colon |\zeta| \geq \delta,\ \operatorname{Im} \zeta \geq p^{-1+\gamma}\}$. We introduce the resolvent $R(\zeta) := (XX^* - \zeta)^{-1}$ of XX^* at $\zeta \in \mathbb{H}$ and denote its entries by $R_{kl}(\zeta)$ for $k, l \in [p]$.

Theorem 5.2.6 (Local law for Gram matrices)**.** *Let Assumptions 5.2.5 hold true. Fix $\delta > 0$ and $\gamma \in (0, 1)$. Then there is $\rho \in (0, \rho_*)$ depending only on the model parameters and δ such that if we define $\kappa = \kappa^{(p)} \colon \mathbb{H} \to (0, \infty]$ through*

$$\kappa(\zeta) = \left(\Delta_\rho(\operatorname{Re} \zeta)^{1/3} + \langle \operatorname{Im} m(\zeta) \rangle\right)^{-1}$$

then, for each $\varepsilon > 0$ *and* $D > 0$*, there is a constant* $C_{\varepsilon,D} > 0$ *such that*

$$\mathbb{P}\left(\sup_{\substack{\zeta \in \mathbb{D}_{\delta,\gamma} \\ k,l \in [p]}} p^{-\varepsilon} \Big| R_{kl}(\zeta) - m_k(\zeta)\delta_{kl} \Big| \leq \sqrt{\frac{\langle \mathrm{Im}\, m(\zeta) \rangle}{p \mathrm{Im}\, \zeta}} + \min\left\{ \frac{1}{\sqrt{p \mathrm{Im}\, \zeta}}, \frac{\kappa(\zeta)}{p \mathrm{Im}\, \zeta} \right\} \right) \geq 1 - \frac{C_{\varepsilon,D}}{p^D}. \tag{5.2.8a}$$

Furthermore, for any $\varepsilon > 0$ *and* $D > 0$*, there is a constant* $C_{\varepsilon,D} > 0$ *such that, for any deterministic vector* $w \in \mathbb{C}^p$ *satisfying* $\max_{k \in [p]} |w_k| \leq 1$*, we have*

$$\mathbb{P}\left(\sup_{\zeta \in \mathbb{D}_{\delta,\gamma}} \left| \frac{1}{p} \sum_{k=1}^{p} w_k \Big(R_{kk}(\zeta) - m_k(\zeta) \Big) \right| \leq p^{\varepsilon} \min\left\{ \frac{1}{\sqrt{p \mathrm{Im}\, \zeta}}, \frac{\kappa(\zeta)}{p \mathrm{Im}\, \zeta} \right\} \right) \geq 1 - \frac{C_{\varepsilon,D}}{p^D}. \tag{5.2.8b}$$

The constant $C_{\varepsilon,D}$ *in* (5.2.8) *depends only on the model parameters as well as* δ *and* γ *in addition to* ε *and* D*.*

Remark 5.2.7. (i) (Corollaries of the local law) In the same way as in [7] and in Chapter 4, the standard corollaries of a local law – convergence of cumulative distribution function, rigidity of eigenvalues, anisotropic law and delocalization of eigenvectors – may be proven.

(ii) (Local law in the bulk and away from $\mathrm{supp}\,\nu$) In the bulk, Theorem 5.2.6 has already been proven in Chapter 4. Away from $\mathrm{supp}\,\nu$, the convergence rate in (5.2.8a) and (5.2.8b) can be improved and thus the condition $\mathrm{Im}\,\zeta \geq p^{-1+\gamma}$ can be removed there. See Chapter 4 for Gram matrices and Chapter 7 for Kronecker matrices.

(iii) (Local law close to zero) Strengthening the assumption **(A2)**, we have proven the local law close to zero in the cases, $n = p$ and $|p - n| \geq cn$, in Chapter 4.

5.3. Quadratic vector equation

In this section, we translate (5.2.3) into a quadratic vector equation of [4] (see (5.3.2) below) and show that Proposition 5.2.1 trivially follows from [4]. However, the singularity analysis in [4] has to be changed essentially due to the violation of the uniform primitivity condition, **A3** in [4], on $\boldsymbol{S}$ (cf. (5.3.1) below) in our setup.

Let $\mathfrak{X} := \mathfrak{X}_1 \sqcup \mathfrak{X}_2$ be the disjoint union of $\mathfrak{X}_1$ and $\mathfrak{X}_2$ and π the probability measure defined through

$$\pi(A \sqcup B) = (\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))^{-1}(\pi_1(A) + \pi_2(B)), \qquad \text{for } A \subset \mathfrak{X}_1,\ B \subset \mathfrak{X}_2.$$

Moreover, we denote the set of bounded measurable functions $\mathfrak{X} \to \mathbb{C}$ by $\mathscr{B} := \{\boldsymbol{w} \colon \mathfrak{X} \to \mathbb{C} \colon \|\boldsymbol{w}\|_\infty := \sup_{x\in\mathfrak{X}} |\boldsymbol{w}(x)| < \infty\}$ with the supremum norm $\|\cdot\|_\infty$. Finally, on $\mathscr{B} = \mathscr{B}_1 \oplus \mathscr{B}_2$, we define the linear operator $\boldsymbol{S} \colon \mathscr{B} \to \mathscr{B}$ through

$$\boldsymbol{S} := \begin{pmatrix} 0 & S \\ S^t & 0 \end{pmatrix}, \qquad \text{i.e.,} \quad \boldsymbol{S}\boldsymbol{w} = S(\boldsymbol{w}|_{\mathfrak{X}_2}) + S^t(\boldsymbol{w}|_{\mathfrak{X}_1}) \quad \text{for } \boldsymbol{w} \in \mathscr{B}. \tag{5.3.1}$$

Here, we consider $S(\boldsymbol{w}|_{\mathfrak{X}_2})$ and $S^t(\boldsymbol{w}|_{\mathfrak{X}_1})$ as functions $\mathfrak{X} \to \mathbb{C}$, extended by zero outside of $\mathfrak{X}_1$ and $\mathfrak{X}_2$, respectively. Instead of (5.2.3), we study the quadratic vector equation (QVE)

$$-\frac{1}{\boldsymbol{m}} = z + \boldsymbol{S}\boldsymbol{m} \tag{5.3.2}$$

for $z \in \mathbb{H}$. Here, we used the change of variables $z^2 = \zeta$. We now explain how $\boldsymbol{m}$ and m are related. If $\boldsymbol{m}$ is a solution of (5.3.2) then $m_1 := \boldsymbol{m}|_{\mathfrak{X}_1}$ and $m_2 := \boldsymbol{m}|_{\mathfrak{X}_2}$ satisfy $-m_1^{-1} = z + Sm_2$ and $-m_2^{-1} = z + S^t m_1$. Solving the second equation for m_2, plugging the result into the first relation and choosing $z = \sqrt{\zeta} \in \mathbb{H}$, we see that m defined through

$$m(\zeta) = \frac{m_1(\sqrt{\zeta})}{\sqrt{\zeta}} \tag{5.3.3}$$

for $\zeta \in \mathbb{H}$ is a solution of (5.2.3). If $\boldsymbol{m}$ has positive imaginary part then m as well.

For $\boldsymbol{u} \in \mathscr{B}$, we write $\boldsymbol{u}_x := \boldsymbol{u}(x)$ with $x \in \mathfrak{X}$. For $\boldsymbol{u}, \boldsymbol{w} \in \mathscr{B}$, we denote the scalar product of $\boldsymbol{u}$ and $\boldsymbol{w}$ and the average of $\boldsymbol{u}$ by

$$\langle \boldsymbol{u}, \boldsymbol{w} \rangle := \int_{\mathfrak{X}} \overline{\boldsymbol{u}_x}\, \boldsymbol{w}_x \pi(\mathrm{d}x), \quad \langle \boldsymbol{u} \rangle := \langle 1, \boldsymbol{u} \rangle = \int_{\mathfrak{X}} \boldsymbol{u}_x \pi(\mathrm{d}x). \tag{5.3.4}$$

We also introduce the Hilbert space $L^2(\pi) := \{\boldsymbol{u} \colon \mathfrak{X} \to \mathbb{C} \colon \langle \boldsymbol{u}, \boldsymbol{u} \rangle < \infty\}$. The operator $\boldsymbol{S}$ is symmetric on $\mathscr{B}$ with respect to $\langle \cdot\,, \cdot \rangle$ and positivity preserving, as $s_{kr} \geq 0$ for all $k \in \mathfrak{X}_1$ and $r \in \mathfrak{X}_2$. Therefore, by Theorem 2.1 in [4], there exists $\boldsymbol{m} \colon \mathbb{H} \to \mathscr{B}$ which satisfies (5.3.2) for all $z \in \mathbb{H}$. This function is unique if we require that the solution of (5.3.2) satisfies $\operatorname{Im} \boldsymbol{m}(z) > 0$ for $z \in \mathbb{H}$. Moreover, $\boldsymbol{m} \colon \mathbb{H} \to \mathscr{B}$ is analytic and, for all

$z \in \mathbb{H}$, we have

$$\|\boldsymbol{m}(z)\|_2 \le 2|z|^{-1}.$$

Furthermore, for all $x \in \mathfrak{X}$, there are symmetric probability measures $\boldsymbol{\rho}_x$ on $\mathbb{R}$ such that

$$\boldsymbol{m}_x(z) = \int_{\mathbb{R}} \frac{1}{\tau - z} \boldsymbol{\rho}_x(\mathrm{d}\tau) \tag{5.3.5}$$

for all $z \in \mathbb{H}$ [4]. That means that $\boldsymbol{m}_x$ is the Stieltjes transform of $\boldsymbol{\rho}_x$. By (2.7) in [4], the definition of Σ in (5.2.5) and $\|\boldsymbol{S}\| = \|\boldsymbol{S}\|_{\mathscr{B}\to\mathscr{B}} = \max\{\|S\|_{\mathscr{B}_2\to\mathscr{B}_1}, \|S^t\|_{\mathscr{B}_1\to\mathscr{B}_2}\}$, the support of $\boldsymbol{\rho}_x$ is contained in $[-\Sigma^{1/2}, \Sigma^{1/2}]$.

Proof of Proposition 5.2.1. The existence of m directly follows from the transform in (5.3.3) and the existence of $\boldsymbol{m}$. The uniqueness of m and the existence of ν_k, $k \in \mathfrak{X}_1$, are obtained as in the proof of Theorem 4.2.2 in Chapter 4. □

The special structure of $\boldsymbol{S}$ (cf. (5.3.1)) implies an important symmetry of the solution $\boldsymbol{m}$. We multiply (5.3.2) by $\boldsymbol{m}$ and take the scalar product of the result with $\boldsymbol{e}_- \in \mathscr{B}$ defined through $\boldsymbol{e}_-(k) = 1$ if $k \in \mathfrak{X}_1$ and $\boldsymbol{e}_-(q) = -1$ if $q \in \mathfrak{X}_2$. As $\langle \boldsymbol{e}_-, \boldsymbol{m}(\boldsymbol{S}\boldsymbol{m})\rangle = 0$, we have

$$z\langle \boldsymbol{e}_-, \boldsymbol{m}\rangle = -\langle \boldsymbol{e}_-\rangle = -\frac{\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)}{\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2)}, \tag{5.3.6}$$

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

Assumptions 5.3.1. In the remainder of this section, we assume that **(A1)**, **(A2)**, **(A4)** and the following condition hold true:

(C2) There are $\tilde{\delta} > 0$ and $\Phi > 0$ such that for all $z \in \mathbb{H}$ satisfying $|z| \ge \tilde{\delta}$, we have

$$\|\boldsymbol{m}(z)\|_\infty \le \Phi.$$

Remark 5.3.2 (Relation between **(A3)** and **(C2)**)**.** By slightly adapting the proofs of Theorem 6.1 (ii) and Proposition 6.6 in [4], we see that, by **(A3)**, for each $\tilde{\delta} > 0$, there is $\Phi_{\tilde{\delta}} > 0$ such that **(C2)** is satisfied with a constant $\Phi \equiv \Phi_{\tilde{\delta}}$.

Since our estimates in this section will be uniform in all models that satisfy **(A1)**, **(A2)**, **(A4)** and **(C2)** with the same constants, we introduce the following notion.

Convention 5.3.3 (Comparison relation)**.** *For nonnegative scalars or vectors f and g, we will use the notation $f \lesssim g$ if there is a constant $c > 0$, depending only on π_*, π^* in* ***(A1)****, $L_1, L_2, \kappa_1, \kappa_2$ in* ***(A2)****, Ψ_1, Ψ_2 in* ***(A4)*** *as well as $\tilde{\delta}$ and Φ in* ***(C2)****, such that $f \le cg$. Moreover, we write $f \sim g$ if both, $f \lesssim g$ and $f \gtrsim g$, hold true.*

5.3.1. Hölder continuity and analyticity. We recall Σ from (5.2.5) and introduce the set $\mathbb{H}_{\tilde{\delta}}^{\Sigma} := \{z \in \mathbb{H} \colon 2\tilde{\delta} \le |z| \le 10\Sigma^{1/2}\}$ and its closure $\overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$.

Proposition 5.3.4 (Regularity of $\boldsymbol{m}$)**.** *Assume* ***(A1)****,* ***(A2)****,* ***(A4)*** *and* ***(C2)****.*

(i) The restriction $\boldsymbol{m} \colon \mathbb{H}_{\tilde{\delta}}^{\Sigma} \to \mathscr{B}$ is uniformly 1/3-Hölder continuous, i.e.,

$$\|\boldsymbol{m}(z) - \boldsymbol{m}(z')\|_\infty \lesssim |z - z'|^{1/3} \tag{5.3.7}$$

for all $z, z' \in \mathbb{H}_{\tilde{\delta}}^{\Sigma}$. In particular, $\boldsymbol{m}$ can be uniquely extended to a uniformly 1/3-Hölder continuous function $\overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma} \to \mathscr{B}$, which we also denote by $\boldsymbol{m}$.

(ii) The measure $\boldsymbol{\rho}$ from (5.3.5) is absolutely continuous, i.e., there is a function $\boldsymbol{\rho}^d \colon \mathfrak{X} \times \mathbb{R} \setminus (-2\tilde{\delta}, 2\tilde{\delta}) \to [0, \infty)$, $(x, \tau) \mapsto \boldsymbol{\rho}_x^d(\tau)$ such that

$$\left(\boldsymbol{\rho}_x|_{\mathbb{R}\setminus(-2\tilde{\delta},2\tilde{\delta})}\right)(\mathrm{d}\tau) = \boldsymbol{\rho}_x^d(\tau)\mathrm{d}\tau, \qquad \text{for all } x \in \mathfrak{X}. \tag{5.3.8}$$

The components $\boldsymbol{\rho}_x^d$ are comparable with each other, i.e., $\boldsymbol{\rho}_x^d(\tau) \sim \boldsymbol{\rho}_y^d(\tau)$ for all $x, y \in \mathfrak{X}$ and $\tau \in \mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}]$. Moreover, the function $\boldsymbol{\rho}^d \colon \mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}] \to \mathscr{B}$ is uniformly 1/3-Hölder continuous, symmetric in τ, $\boldsymbol{\rho}^d(\tau) = \boldsymbol{\rho}^d(-\tau)$, and real-analytic around any $\tau \in \mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}]$ apart from points $\tau \in \operatorname{supp}\langle \boldsymbol{\rho}^d \rangle$, where $\boldsymbol{\rho}^d(\tau) = 0$.

A similar result has been obtained in Theorem 2.4 in [4] essentially relying on the uniform primitivity assumption **A3** in [4]. For discrete $\mathfrak{X}_1$ and $\mathfrak{X}_2$ without assuming **(C2)**, Lemma 4.3.8 in Chapter 4 shows Hölder continuity of $\langle \boldsymbol{m} \rangle$ instead of $\boldsymbol{m}$ with a smaller exponent than 1/3. Both conditions, **A3** in [4] and the discreteness of $\mathfrak{X}_1$ and $\mathfrak{X}_2$, are violated in our setup. However, based on the proof of Theorem 2.4 in [4], we now explain how to extend the arguments of [4] and Chapter 4 to show Proposition 5.3.4.

Lemma 5.3.5. *Uniformly for all $z \in \mathbb{H}_{\tilde{\delta}}^{\Sigma}$, we have*

$$|\boldsymbol{m}(z)| \sim 1, \tag{5.3.9}$$

$$\operatorname{Im} \boldsymbol{m}(z) \sim \langle \operatorname{Im} \boldsymbol{m}(z) \rangle. \tag{5.3.10}$$

Using the arguments in the proof of Lemma 5.4 in [4], Lemma 5.3.5 follows immediately from **(A2)**, **(C2)** and (5.3.2). Here, as in the proof of Lemma 4.3.1 in Chapter 4, the uniform primitivity assumption **A3** of [4] has to be replaced by (B') in Chapter 4, which is a direct consequence of **(A2)**.

The Hölder continuity and the analyticity of $\boldsymbol{m}$ and hence $\boldsymbol{\rho}^d$ will be consequences of analyzing the perturbed QVE

$$-\frac{1}{\boldsymbol{g}} = z + \boldsymbol{S}\boldsymbol{g} + \boldsymbol{d} \tag{5.3.11}$$

for $z \in \mathbb{H}$ and $\boldsymbol{d} = z - z'$ as well as the stability operator $\boldsymbol{B}$ defined through

$$\boldsymbol{B}(z)\boldsymbol{u} = \frac{|\boldsymbol{m}(z)|^2}{\boldsymbol{m}(z)^2}\boldsymbol{u} - \boldsymbol{F}(z)\boldsymbol{u}, \tag{5.3.12}$$

where $\boldsymbol{F}(z)\colon \mathscr{B} \to \mathscr{B}$ is defined through $\boldsymbol{F}(z)\boldsymbol{u} = |\boldsymbol{m}(z)|\boldsymbol{S}\left(|\boldsymbol{m}(z)|\boldsymbol{u}\right)$ for any $\boldsymbol{u} \in \mathscr{B}$ (cf. [4] and Chapter 4). Correspondingly, we introduce $F(z)\colon \mathscr{B}_2 \to \mathscr{B}_1$ via

$$F(z)w = |m_1(z)|S(|m_2(z)|w)$$

for $w \in \mathscr{B}_2$ and $F^t(z)\colon \mathscr{B}_1 \to \mathscr{B}_2$ via $F^t(z)u = |m_2(z)|S^t(|m_1(z)|u)$ for $u \in \mathscr{B}_1$.

To formulate the key properties of $\boldsymbol{F}$ and $\boldsymbol{B}$, we now introduce some notation. The operator norms for operators on $\mathscr{B}$ and $L^2(\pi)$ are denoted by $\|\cdot\|_\infty$ and $\|\cdot\|_2$, respectively. If $T\colon L^2 \to L^2$ is a compact self-adjoint operator then the *spectral gap* $\operatorname{Gap}(T)$ is the difference between the two largest eigenvalues of $|T|$. We remark that S and hence FF^t are compact operators due to **(A4)**.

Lemma 5.3.6 (Properties of $\boldsymbol{F}$)**.** *The eigenspace of $\boldsymbol{F}$ associated to $\|\boldsymbol{F}\|_2$ is one-dimensional and spanned by a unique $L^2(\pi)$-normalized positive $\boldsymbol{f}_+ \in \mathscr{B}$. The eigenspace associated to $-\|\boldsymbol{F}\|_2$ is one-dimensional and spanned by $\boldsymbol{f}_- := \boldsymbol{f}_+\boldsymbol{e}_- \in \mathscr{B}$. We have*

$$\boldsymbol{f}_+ \sim 1 \tag{5.3.13}$$

uniformly for $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}}$. *There is* $\varepsilon \sim 1$ *such that*

$$\|\boldsymbol{F}\boldsymbol{u}\|_2 \leq (\|\boldsymbol{F}\|_2 - \varepsilon)\|\boldsymbol{u}\|_2 \tag{5.3.14}$$

uniformly for $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}}$ *and for all* $\boldsymbol{u} \in \mathscr{B}$ *satisfying* $\langle \boldsymbol{f}_+, \boldsymbol{u}\rangle = 0$ *and* $\langle \boldsymbol{f}_-, \boldsymbol{u}\rangle = 0$. *Furthermore, we have* $\|\boldsymbol{F}\|_2 \leq 1$, $\mathrm{Gap}(F(z)F^t(z)) \sim 1$ *uniformly for* $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}}$.

Lemma 5.3.6 is a consequence of the proof of Lemma 4.3.3 in Chapter 4 with $r = |\boldsymbol{m}|$ and (5.3.9).

Lemma 5.3.7. *Uniformly for* $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}}$, *we have*

$$\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim \frac{1}{\langle \mathrm{Im}\, \boldsymbol{m}(z)\rangle^2}. \tag{5.3.15}$$

Proof. We describe the modifications in the proof of Lemma 4.3.5 in Chapter 4 necessary to obtain (5.3.15). We remark that (4.3.11) in Chapter 4 holds true due to **(A4)**.

Let $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}}$. Taking the real part in (5.3.2), using (5.3.9) and Lemma 5.3.6, we obtain the bound $\|\mathrm{Re}\, \boldsymbol{m}|\boldsymbol{m}|^{-1}\|_2 \geq |\mathrm{Re}\, z|\|\boldsymbol{m}\|_2/2 \gtrsim |\mathrm{Re}\, z|$. Therefore, using $\langle(\mathrm{Im}\, \boldsymbol{m})^2\rangle \geq \langle \mathrm{Im}\, \boldsymbol{m}\rangle^2$ by Jensen's inequality, we obtain (4.3.28) in Chapter 4 with $\kappa = 2$. Employing

$$\mathrm{Gap}\left(F(z)F^t(z)\right) \sim 1,$$

we get $\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim (\mathrm{Re}\, z)^{-2}\langle \mathrm{Im}\, \boldsymbol{m}(z)\rangle^{-2}$. As $\|\boldsymbol{B}^{-1}(z)\|_2 \leq (1 - \|\boldsymbol{F}(z)\|_2)^{-1} \lesssim (\mathrm{Im}\, z)^{-1}$ by (4.3.22) in Chapter 4 we conclude from $\mathrm{Im}\, \boldsymbol{m} \lesssim \min\{1, (\mathrm{Im}\, z)^{-1}\}$ that

$$\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim |z|^{-2}\langle \mathrm{Im}\, \boldsymbol{m}(z)\rangle^{-2}.$$

This concludes the proof of (5.3.15) since $|z| \geq 2\tilde{\delta}$. □

Note that if $\boldsymbol{\rho}$ has a density $\boldsymbol{\rho}^d$ around a point τ_0 then, uniformly for τ in a neighbourhood of τ_0, we have

$$\boldsymbol{\rho}^d(\tau) = \pi^{-1} \lim_{\eta\downarrow 0} \mathrm{Im}\, \boldsymbol{m}(\tau + \mathrm{i}\eta). \tag{5.3.16}$$

Proof of Proposition 5.3.4. Following the proof of Proposition 7.1 in [4] yields the uniform 1/3-Hölder continuity of $\boldsymbol{m}$ and $\boldsymbol{\rho}^d$. In this proof, the estimate (5.40b) has to be replaced by (5.3.15). Furthermore, (5.3.10) substitutes Proposition 5.3 (ii) in [4], in particular, $\boldsymbol{\rho}^d_x(\tau) \sim \boldsymbol{\rho}^d_y(\tau)$. We remark that now the same proofs extend Lemma 5.3.5, Lemma 5.3.6 and Lemma 5.3.7 to all $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$. Hence, the proof of Corollary 7.6 in [4] yields the analyticity using (5.3.16) for $\tau \in \mathbb{R} \cap \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$. □

5.3.2. Singularities of $\boldsymbol{\rho}^d$ and the cubic equation. We now study the behaviour of $\boldsymbol{\rho}^d$ near points $\tau \in \mathbb{R}$, where $\boldsymbol{\rho}^d$ is not analytic. Theorem 2.6 in [4] describes the density near the edges and the cusps as well as the transition between the bulk and the singularity regimes in a quantitative manner. The same results hold for $\boldsymbol{\rho}^d$ as well:

Proposition 5.3.8. *We assume **(A1)**, **(A2)**, **(A4)** and **(C2)**. Then all statements of Theorem 2.6 in [4] hold true on* $\mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}]$.

For the proof of Proposition 5.3.8 we follow Chapter 8 and 9 in [4] which contain the proof of the analogue of Proposition 5.3.8, Theorem 2.6 in [4], and describe the necessary changes as well as the main philosophy.

The shape of the singularities of $\boldsymbol{m}$ as well as the stability of the QVE (cf. Chapter 10 in [4]) will be a consequence of the stability of a cubic equation. We note that similar as in Lemma 8.1 of [4], the following properties of the stability operator $\boldsymbol{B} = \boldsymbol{B}(z)$ defined in (5.3.12) can be proven. There is $\varepsilon_* \sim 1$ such that for $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ satisfying $\langle \operatorname{Im} \boldsymbol{m}(z) \rangle \leq \varepsilon_*$, $\boldsymbol{B}$ has a unique eigenvalue $\beta = \beta(z)$ of smallest modulus and $|\beta'| - |\beta| \gtrsim 1$ for all $\beta' \in \operatorname{Spec}(\boldsymbol{B}) \setminus \{\beta\}$. The eigenspace associated to β is one-dimensional and there is a unique vector $\boldsymbol{b} = \boldsymbol{b}(z) \in \mathscr{B}$ in this eigenspace such that $\langle \boldsymbol{b}(z), \boldsymbol{f}_+ \rangle = 1$.

Let $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ such that $\langle \operatorname{Im} \boldsymbol{m}(z) \rangle \leq \varepsilon_*$ and $\boldsymbol{g} \in \mathscr{B}$ satisfy the perturbed QVE, (5.3.11), at z. We define

$$\Theta(z) := \left\langle \frac{\bar{\boldsymbol{b}}(z)}{\langle \boldsymbol{b}(z)^2 \rangle}, \frac{\boldsymbol{g} - \boldsymbol{m}(z)}{|\boldsymbol{m}(z)|} \right\rangle. \tag{5.3.17}$$

By possibly shrinking $\varepsilon_* \sim 1$, we obtain that if $\|\boldsymbol{g} - \boldsymbol{m}(z)\|_\infty \leq \varepsilon_*$ then it can be shown as in Proposition 8.2 in [4] that Θ satisfies

$$\mu_3 \Theta^3 + \mu_2 \Theta^2 + \mu_1 \Theta + \langle |\boldsymbol{m}| \bar{\boldsymbol{b}}, \boldsymbol{d} \rangle = \kappa\left((\boldsymbol{g} - \boldsymbol{m})/|\boldsymbol{m}|, \boldsymbol{d} \right), \tag{5.3.18}$$

where μ_1, μ_2 and μ_3, which depend only on S and z, as well as κ are given in [4].

The main ingredient that needs to be changed in our setup is the estimate in (8.13) of [4]. It gives a lower bound on the nonnegative quadratic form

$$\mathcal{D}(\boldsymbol{w}) := \Big\langle \boldsymbol{Q}_+\boldsymbol{w}\,,\,(\|\boldsymbol{F}\|_2 + \boldsymbol{F})\,(1-\boldsymbol{F})^{-1}\,\boldsymbol{Q}_+\boldsymbol{w}\Big\rangle \tag{5.3.19}$$

for $\boldsymbol{w} \in \mathscr{B}$, where the projection $\boldsymbol{Q}_+$ is defined through $\boldsymbol{Q}_+\boldsymbol{w} := \boldsymbol{w} - \langle \boldsymbol{f}_+, \boldsymbol{w}\rangle \boldsymbol{f}_+$. For some $c(z) > 0$ and all $\boldsymbol{w} \in \mathscr{B}$, this lower bounds reads as follows

$$\mathcal{D}(\boldsymbol{w}) \geq c(z)\|\boldsymbol{Q}_+\boldsymbol{w}\|_2^2. \tag{5.3.20}$$

However, in our setup, owing to the second unstable direction $\boldsymbol{f}_- \perp \boldsymbol{f}_+$, $\boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2\boldsymbol{f}_-$, we have $\mathcal{D}(\boldsymbol{f}_-) = 0$ which contradicts (5.3.20). In [4], the estimate (5.3.20) is only used to obtain

$$|\mu_3(z)| + |\mu_2(z)| \gtrsim 1 \tag{5.3.21}$$

(cf. (8.34) in [4]) for all $z \in \overline{\mathbb{H}}_\delta^\Sigma$ satisfying $\langle \operatorname{Im} \boldsymbol{m}(z)\rangle \leq \varepsilon_*$ and $\|\boldsymbol{g} - \boldsymbol{m}(z)\|_\infty \leq \varepsilon_*$ for $\varepsilon_* \sim 1$ small enough. In fact, it is shown above (8.50) in [4] that

$$|\mu_3| \gtrsim \psi + \mathcal{O}(\alpha) \qquad |\mu_2| \gtrsim |\sigma| + \mathcal{O}(\alpha). \tag{5.3.22}$$

Here, we introduced the notations $\psi := \mathcal{D}(\boldsymbol{p}\boldsymbol{f}_+^2)$ with $\boldsymbol{p} := \operatorname{sign}(\operatorname{Re} \boldsymbol{m})$ as well as $\alpha := \langle \boldsymbol{f}_+ \operatorname{Im} \boldsymbol{m}/|\boldsymbol{m}|\rangle$ and $\sigma := \langle \boldsymbol{f}_+, \boldsymbol{p}\boldsymbol{f}_+^2\rangle$. The proof used in [4] to show (5.3.22) works in our setup as well. Since $\alpha = \langle \boldsymbol{f}_+ \operatorname{Im} \boldsymbol{m}/|\boldsymbol{m}|\rangle \sim \langle \operatorname{Im} \boldsymbol{m}\rangle \leq \varepsilon_*$ by (5.3.9) and (5.3.13), we conclude that $|\mu_3| + |\mu_2| \gtrsim \psi + |\sigma|$ for $\varepsilon_* \sim 1$ small enough. Hence, (5.3.21) is a consequence of

Lemma 5.3.9 (Stability of the cubic equation)**.** *There exists $\varepsilon_* \sim 1$ such that*

$$\psi(z) + \sigma^2(z) \sim 1 \tag{5.3.23}$$

uniformly for all $z \in \overline{\mathbb{H}}_\delta^\Sigma$ satisfying $\langle \operatorname{Im} \boldsymbol{m}(z)\rangle \leq \varepsilon_*$.

Proof. We first remark that due to (5.3.9), (5.3.10) and possibly shrinking $\varepsilon_* \sim 1$ we can assume

$$|\operatorname{Re} \boldsymbol{m}(z)| \sim 1 \tag{5.3.24}$$

for $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ satisfying $\langle \operatorname{Im} \boldsymbol{m}(z)\rangle \leq \varepsilon_*$. Second, owing to (5.3.14), for all $\boldsymbol{w} \in \mathscr{B}$, we have the following analogue of (5.3.20)

$$\mathcal{D}(\boldsymbol{w}) \gtrsim \|\boldsymbol{Q}_{\pm}\boldsymbol{w}\|_2^2, \tag{5.3.25}$$

where $\boldsymbol{Q}_{\pm}$ is the projection onto the orthogonal complement of $\boldsymbol{f}_+$ and $\boldsymbol{f}_-$, i.e, $\boldsymbol{Q}_{\pm}\boldsymbol{w} = \boldsymbol{w} - \langle \boldsymbol{f}_+\,, \boldsymbol{w}\rangle \boldsymbol{f}_+ - \langle \boldsymbol{f}_-\,, \boldsymbol{w}\rangle \boldsymbol{f}_-$. Note that (5.3.14) also yields the upper bound $\mathcal{D}(\boldsymbol{w}) \lesssim \|\boldsymbol{Q}_+\boldsymbol{w}\|_2^2$ and hence the upper bound in (5.3.23) by (5.3.13). Therefore, it suffices to prove the lower bound in (5.3.23). A straightforward computation starting from (5.3.25) and using $\boldsymbol{f}_- = \boldsymbol{e}_-\boldsymbol{f}_+$ yields

$$\psi + \sigma^2 = \mathcal{D}(\boldsymbol{p}\boldsymbol{f}_+^2) + \langle \boldsymbol{p}\boldsymbol{f}_+^3\rangle^2 \gtrsim \|\boldsymbol{p}\boldsymbol{f}_+^2 - \langle \boldsymbol{f}_-\,, \boldsymbol{p}\boldsymbol{f}_+^2\rangle \boldsymbol{f}_-\|_2^2 = \left\langle \boldsymbol{f}_+^2 \left(\boldsymbol{p}\boldsymbol{f}_+ - \langle \boldsymbol{p}\boldsymbol{e}_-\boldsymbol{f}_+^3\rangle \boldsymbol{e}_-\right)^2\right\rangle. \tag{5.3.26}$$

Using (5.3.13), (5.3.24) and $|\operatorname{Re} \boldsymbol{m}| = \boldsymbol{p}\operatorname{Re} \boldsymbol{m}$, we conclude

$$\begin{aligned}\psi + \sigma^2 &\gtrsim \left\langle (\operatorname{Re} \boldsymbol{m})^2 \left(\boldsymbol{p}\boldsymbol{f}_+ - \langle \boldsymbol{p}\boldsymbol{e}_-\boldsymbol{f}_+^3\rangle \boldsymbol{e}_-\right)^2\right\rangle \\ &\geq \langle \boldsymbol{f}_+ |\operatorname{Re} \boldsymbol{m}|\rangle \left(\langle \boldsymbol{f}_+ |\operatorname{Re} \boldsymbol{m}|\rangle + 2\langle \boldsymbol{p}\boldsymbol{e}_-\boldsymbol{f}_+^3\rangle\langle \boldsymbol{e}_-\rangle \operatorname{Re} \frac{1}{z}\right)\end{aligned} \tag{5.3.27}$$

Here, we employed Jensen's inequality and (5.3.6) in the second step. Since $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ and $\langle \boldsymbol{e}_-\rangle = 0$ for $\pi_1(\mathfrak{X}_1) = \pi_2(\mathfrak{X}_2)$, there exists $\iota_* \sim 1$ such that the last factor on the right-hand side of (5.3.27) is bounded from below by $\langle \boldsymbol{f}_+|\operatorname{Re} \boldsymbol{m}|\rangle/2$ for all $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ and $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| \leq \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$. Since $\langle \boldsymbol{f}_+|\operatorname{Re} \boldsymbol{m}|\rangle^2 \gtrsim 1$ by (5.3.13) and (5.3.24), this finishes the proof of (5.3.23) for $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| \leq \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$. For the proof of (5.3.23) in the remaining regime, $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| > \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$, we introduce $\boldsymbol{y} := \boldsymbol{e}_-\boldsymbol{p}\boldsymbol{f}_+$ and use $\boldsymbol{y}^2 = \boldsymbol{f}_+^2 \sim 1$ and $\left(\boldsymbol{y} + \langle \boldsymbol{y}^3\rangle\right)^2 \lesssim 1$ by (5.3.13) to obtain from (5.3.26) the bound

$$\psi + \sigma^2 \gtrsim \left\langle \left(\boldsymbol{y} - \langle \boldsymbol{y}^3\rangle\right)^2 \left(\boldsymbol{y} + \langle \boldsymbol{y}^3\rangle\right)^2\right\rangle = \left\langle \left((\boldsymbol{y}^2 - 1) + (1 - \langle \boldsymbol{y}^3\rangle^2)\right)^2\right\rangle \geq \left\langle \left(\boldsymbol{y}^2 - 1\right)^2\right\rangle. \tag{5.3.28}$$

Here, we used $\langle \boldsymbol{y}^2\rangle = \langle \boldsymbol{f}_+^2\rangle = 1$ and $\left(1 - \langle \boldsymbol{y}^3\rangle^2\right)^2 \geq 0$. Since $0 = \langle \boldsymbol{f}_-\,, \boldsymbol{f}_+\rangle = \langle \boldsymbol{e}_-\boldsymbol{y}^2\rangle$, using (5.3.28), we conclude

$$\langle \boldsymbol{e}_-\rangle^2 = \langle \boldsymbol{e}_-(1 - \boldsymbol{y}^2)\rangle^2 \leq \langle (1 - \boldsymbol{y}^2)^2\rangle \lesssim \psi + \sigma^2. \tag{5.3.29}$$

This implies (5.3.23) for $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| > \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$ as $\langle \boldsymbol{e}_- \rangle^2 \geq \iota_*^2 \sim 1$. This completes the proof of Lemma 5.3.9. □

Following the remaining arguments of Chapter 8 and 9 in [4] yields Proposition 5.3.8.

5.4. Proofs of Theorem 5.2.3 and Theorem 5.2.6

Proof of Theorem 5.2.3. By Remark 5.3.2, we can apply Proposition 5.3.4 for each $\tilde{\delta} > 0$. Hence, there are $\boldsymbol{\rho}^0 \in \mathscr{B}$ and $\boldsymbol{\rho}^d \colon \mathfrak{X} \times \mathbb{R} \setminus \{0\} \to [0, \infty)$ such that

$$\boldsymbol{\rho}_x(\mathrm{d}\tau) = \boldsymbol{\rho}_x^0 \delta_0(\mathrm{d}\tau) + \boldsymbol{\rho}_x^d(\tau)\mathrm{d}\tau$$

for all $x \in \mathfrak{X}$. For $k \in \mathfrak{X}_1$, we set $\nu_k^0 := \boldsymbol{\rho}_k^0$ and

$$\nu_k^d(E) := E^{-1/2} \boldsymbol{\rho}_k^d(E^{1/2}) \mathbf{1}(E > 0) \tag{5.4.1}$$

with $E \in \mathbb{R}$. Therefore, using (5.3.3), we obtain (5.2.6) (cf. the proof of Theorem 4.2.2 in Chapter 4). The 1/3-Hölder continuity of $\boldsymbol{\rho}^d$ implies the 1/3-Hölder continuity of ν^d. Similarly, the analyticity of ν^d is obtained from the analyticity of $\boldsymbol{\rho}^d$. From Proposition 5.3.8 with $\tilde{\delta} = \sqrt{\delta}/2$, we conclude that $\mathfrak{P} \cap (\delta, \infty)$ is a finite union of open intervals and its connected components have a Lebesgue measure of at least $2\rho_*$ for some ρ_* depending only on the model parameters and δ. This completes the proof (i).

For the proof of (ii), we follow the proof of Theorem 2.6 in [5]. We replace the estimates (4.1), (4.2), (5.3) and (6.7) as well as their proofs in [5] by (5.3.9), (5.3.10), (5.3.15) and (5.3.23) as well as their proofs in this note, respectively. This proves a result corresponding to Theorem 2.6 in [5] for $\boldsymbol{\rho}^d$ and $\tau_0 \in (\partial\mathfrak{P}) \cap (0, \infty)$ in our setup. Using the transform (5.4.1) completes the proof of Theorem 5.2.3. □

Proof of Theorem 5.2.6. Note that **(B1)** implies **(A4)**. By Remark 5.3.2, **(A3)** implies **(C2)**. Using (5.3.21) to replace (8.34) in [4], we obtain an analogue of Proposition 10.1 in [4] in our setup on $\overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$. Therefore, we have proven in our setup analogues of all the ingredients provided in [4] and used in [7] to prove a local law for Wigner-type random matrices with a uniform primitive variance matrix. Thus, following the arguments in [7], we obtain a local law for the resolvent of $\boldsymbol{H}$ defined in (5.1.2) and spectral parameters $z \in \mathbb{H}_{\tilde{\delta}}^{\Sigma} \cap \{w \in \mathbb{H} \colon \operatorname{Im} w \geq (p+n)^{-1+\gamma}\}$, where $\tilde{\delta} = \sqrt{\delta}/2$ and $\gamma \in (0, 1)$. Proceeding as in the proof of Theorem 4.2.3 in Chapter 4 yields Theorem 5.2.6. □

CHAPTER 6

Local inhomogeneous circular law

This section is devoted to the article [13] which is joint work with László Erdős and Torben Krüger. We consider large random matrices X with centered, independent entries which have comparable but not necessarily identical variances. Girko's circular law asserts that the spectrum is supported in a disk and in case of identical variances, the limiting density is uniform. In this special case, the *local circular law* by Bourgade *et al.* [44, 45] shows that the empirical density converges even locally on scales slightly above the typical eigenvalue spacing. In the general case, the limiting density is typically inhomogeneous and it is obtained via solving a system of deterministic equations. Our main result is the local *inhomogeneous* circular law in the bulk spectrum on the optimal scale for a general variance profile of the entries of X.

6.1. Introduction

The density of eigenvalues of large random matrices typically converges to a deterministic limit as the dimension n of the matrix tends to infinity. In the Hermitian case, the best known examples are the Wigner semicircle law for Wigner ensembles and the Marchenko-Pastur law for sample covariance matrices. In both cases the spectrum is real, and these laws state that the empirical eigenvalue distribution converges to an explicit density on the real line.

The spectra of non-Hermitian random matrices concentrate on a domain of the complex plane. The most prominent case is the *circular law*, asserting that for an $n \times n$ matrix X with independent, identically distributed entries, satisfying $\mathbb{E}\, x_{ij} = 0$, $\mathbb{E}|x_{ij}|^2 = n^{-1}$, the empirical density converges to the uniform distribution on the unit disk $\{z : |z| < 1\} \subset \mathbb{C}$. Despite the apparent similarity in the statements, it is considerably harder to analyze non-Hermitian random matrices than their Hermitian counterparts since eigenvalues of non-Hermitian matrices may respond very drastically to small perturbations. This instability is one reason why the universality of local eigenvalue statistics in the bulk spectrum, exactly on the scale of the eigenvalue spacing, is not yet established for X with independent (even for i.i.d.) entries, while the corresponding statement for Hermitian Wigner matrices, known as the Wigner-Dyson-Mehta universality conjecture, has been proven recently, see [69] for an overview.

The circular law for i.i.d. entries has a long history, we refer to the extensive review [40]. The complex Gaussian case (Ginibre ensemble) has been settled in the sixties by Mehta using explicit computations. Girko in [81] found a key formula to relate linear statistics of eigenvalues of X to eigenvalues of the family of Hermitian matrices $(X - z\mathbb{1})^*(X - z\mathbb{1})$, where $z \in \mathbb{C}$ is a complex parameter and $\mathbb{1}$ is the identity matrix in $\mathbb{C}^{n\times n}$. Technical difficulties still remained until Bai [22] presented a complete proof under two additional assumptions requiring higher moments and bounded density for the single entry distribution. After a series of further partial results [83, 116, 142] the circular law for i.i.d. entries under the optimal condition, assuming only the existence of the second moment, was established by Tao and Vu [143].

Another line of research focused on the local version of the circular law with constant variances, $\mathbb{E}|x_{ij}|^2 = n^{-1}$, which asserts that the local density of eigenvalues is still uniform on scales $n^{-1/2+\epsilon}$, i.e., slightly above the typical spacing between neighboring eigenvalues. The optimal result was achieved in Bourgade, Yau and Yin [44, 45] and Yin [162] both inside the unit disk ("bulk regime") and at the edge $|z| = 1$. If the first three moments match those of a standard complex Gaussian, then a similar result has also been obtained by Tao and Vu in [146]. In [146], this result was used to prove the universality of local eigenvalue statistics under the assumption that the first four moments match those of a complex Gaussian. While there is no proof of universality for general distributions without moment matching conditions yet, similar to the development in the Hermitian case, the local law is expected to be one of the key ingredients of such a proof in the future.

In this paper we study non-Hermitian matrices X with a general matrix of variances $S = (s_{ij})_{i,j=1}^n$, i.e., we assume that x_{ij} are centered, independent, but $s_{ij} := \mathbb{E}|x_{ij}|^2$ may depend nontrivially on the indices i, j. We show that the eigenvalue density is close to a deterministic density σ on the smallest possible scale. As a direct application, our local law implies that the spectral radius $\rho(X)$ of X is arbitrarily close to $\sqrt{\rho(S)}$, where $\rho(S)$ is the spectral radius of S. More precisely, we prove that for every $\varepsilon > 0$

$$\sqrt{\rho(S)} - \varepsilon \leq \rho(X) \leq \sqrt{\rho(S)} + \varepsilon$$

with a very high probability as n tends to infinity. The fact that the spectral radius of X becomes essentially deterministic is the key mathematical mechanism behind the sharp "transition to chaos" in a commonly studied mean field model of dynamical neural networks [135]. This transition is described by the stability/instability of the system of ordinary differential equations

$$\dot{q}_i(t) = q_i(t) - \lambda \sum_{j=1}^{n} x_{ij} q_j(t)$$

for $i = 1, \ldots, n$ as λ varies. Moreover, the number of unstable modes close to the critical value of the parameter λ is determined by the behaviour of σ at the spectral edge which we also analyze. Such systems have originally been studied under the assumption that the coefficients x_{ij} are independent and identically distributed [113]. More recently, however, it was argued [9, 10] that for more realistic applications in neuroscience one should allow x_{ij} to have varying distributions with an arbitrary variance profile S.

After Girko's Hermitization, understanding the spectrum of X reduces to analyzing the spectrum of the family

$$\boldsymbol{H}^z := \begin{pmatrix} 0 & X - z\mathbb{1} \\ X^* - \bar{z}\mathbb{1} & 0 \end{pmatrix} \tag{6.1.1}$$

of Hermitian matrices of double dimension, where $z \in \mathbb{C}$. The Stieltjes transform of the spectral density of $\boldsymbol{H}^z$ at any spectral parameter ζ in the upper half plane $\mathbb{H} := \{\zeta \in \mathbb{C} : \operatorname{Im} \zeta > 0\}$ is approximated via the solution of a system of $2n$ nonlinear equations, written concisely as

$$
\begin{aligned}
-\frac{1}{m_1} &= \zeta + Sm_2 - \frac{|z|^2}{\zeta + S^t m_1}, \\
-\frac{1}{m_2} &= \zeta + S^t m_1 - \frac{|z|^2}{\zeta + Sm_2},
\end{aligned}
\tag{6.1.2}
$$

where $m_a = m_a^z(\zeta) \in \mathbb{H}^n$, $a = 1, 2$ are n-vectors with each component in the upper half plane. The normalized trace of the resolvent, $\frac{1}{2n}\text{trace}(\boldsymbol{H}^z - \zeta\mathbb{1})^{-1}$, is approximately equal to $\frac{1}{n}\sum_j [m_1^z(\zeta)]_j = \frac{1}{n}\sum_j [m_2^z(\zeta)]_j$ in the $n \to \infty$ limit. The spectral density of $\boldsymbol{H}^z$ at any $E \in \mathbb{R}$ is then given by setting $\zeta = E + \mathrm{i}\eta$ and taking the limit $\eta \to 0+$ for the imaginary part of these averages. In fact, for Girko's formula it is sufficient to study the resolvent only along the positive imaginary axis $\zeta \in \mathrm{i}\mathbb{R}_+$. Heuristically, the equations in (6.1.2) arise from second order perturbation theory and in physics they are commonly called *Dyson equations*. Their analogues for general Hermitian ensembles with independent or weakly dependent entries play an essential role in random matrix theory. They have been systematically studied by Girko, for example, (6.1.2) in the current random matrix context appears as the *canonical equation of type* K_{25} in Theorem 25.1 in [82]. In particular, under the condition that all s_{ij} variances are comparable, i.e., $c/n \le s_{ij} \le C/n$ with some positive constants c, C, Girko identifies the limiting density. From his formulas it is clear that this density is rotationally symmetric. He also presents a proof for the weak convergence of the empirical eigenvalue distribution but the argument was considered incomplete. This deficiency can be resolved in a similar manner as for the circular law assuming a bounded density of the single entry distribution using the argument from Section 4.4 of [40]. In a recent preprint [51] Cook *et al.* substantially relax the condition on the uniform bound $s_{ij} \ge c/n$ by replacing it with a concept of *robust irreducibility*. Moreover, relying on the bound by Cook [50] on the smallest singular value of X, they also remove any condition on the regularity of the single entry distribution and prove weak convergence on the global scale.

The matrix $\boldsymbol{H}^z$ may be viewed as the sum of a *Wigner-type matrix* [7] with centered, independent (up to Hermitian symmetry) entries and a deterministic matrix whose two off-diagonal blocks are $-z\mathbb{1}$ and $-\bar{z}\mathbb{1}$, respectively. Disregarding these z terms for the moment, (6.1.2) has the structure of the *Quadratic Vector Equations* that were extensively studied in [4, 5]. Including the z-terms, $\boldsymbol{H}^z$ at first sight seems to be a special case of the random matrix ensembles with nonzero expectations analyzed in [6] and (6.1.2) is the diagonal part of the corresponding *Matrix Dyson Equation (MDE).* In [6] an optimal local law was proven for such ensembles. However, the large zero blocks in the diagonal prevent us from applying these results to $\boldsymbol{H}^z$ or even to $\boldsymbol{H}^{z=0}$. In fact, the flatness condition **A1** in [6] (see (6.3.1) later) prohibit such large zero diagonal blocks. These conditions are essential for the proofs in [6] since they ensure the stability of the corresponding Dyson equation against *any* small perturbation. In this case, there is only one potentially unstable direction, that is associated to a certain Perron-Frobenius eigenvector, and this direction is regularized by the positivity of the density of states at least in the bulk regime of the spectrum.

If the flatness condition **A1** is not satisfied, then the MDE can possess further unstable directions. In particular, in our setup, the MDE is not stable in the previously described strong sense; there is at least one additional unstable direction which cannot be regularized by the positivity of the density of states. Owing to the specific structure of $\boldsymbol{H}^z$, the *matrix* Dyson equation decouples and its diagonal parts satisfy a closed system of *vector* equations (6.1.2). Compared to the MDE, the reduced vector equations (6.1.2) are rather cubic than quadratic in nature. For this reduced system, however, we can show that there is only one further unstable direction, at least when S is entrywise bounded from below by some c/n. The system is not stable against an arbitrary perturbation, but for the perturbation arising in the random matrix problem we reveal a key cancellation in the leading contribution to the unstable direction. Armed with this new insight we will perform a detailed stability analysis of (6.1.2).

This delicate stability analysis is the key ingredient for the proof of our main result, the optimal local law for X with an optimal speed of convergence as $n \to \infty$. In this paper we consider the bulk regime, i.e., spectral parameter z inside the disk with boundary $|z|^2 = \rho(S)$, where $\rho(S)$ is the spectral radius of S. We defer the analysis of the edge of the spectrum of X to later works.

In the special case $z = 0$, we thoroughly studied the system of equations (6.1.2) even for the case when S is a rectangular matrix in Chapter 4 (cf. [14]); the main motivation was to prove the local law for random *Gram matrices*, i.e., matrices of the form XX^*. Note that in Chapter 4 we needed to tackle a much simpler quadratic system since taking $z = 0$ in (6.1.2) removes the most complicated nonlinearity.

Finally, we list two related recent results. Local circular law on the optimal scale in the bulk has been proven in [161] for ensembles of the form TX, where T is a deterministic $N \times M$ matrix and X is a random $M \times N$ matrix with independent, centered entries whose variances are constant and have vanishing third moments. The structure of the product matrix TX is very different from our matrices that could be viewed as the Hadamard (entrywise) product of the matrix $(s_{ij}^{1/2})$ and a random matrix with identical variances. The approach of [161] is also very different from ours: it relies on first assuming that X is Gaussian and using its invariance to reduce the problem to the case when T^*T is diagonal. Then the corresponding Dyson equations are much simpler, in fact they consist of only two scalar equations and they are characterized by a vector of parameters (of the singular values of T) instead of an entire matrix of parameters S. The vanishing third moment condition in [161] is necessary to compare the general distribution with the Gaussian case via a moment matching argument. We also mention the recent proof of the local *single ring theorem* on optimal scale in the bulk [27]. This concerns another prominent non-Hermitian random matrix ensemble that consists of matrices of the form $U\Sigma V$, where U, V are two independent Haar distributed unitaries and Σ is deterministic (may be assumed to be diagonal). The spectrum lies in a ring about the origin and the limiting density can be computed via free convolution [85].

Acknowledgement. We are grateful to David Renfrew for discussing some applications of our results with us and to Dominik Schröder for helping us visualizing our results.

Notation. For vectors $v, w \in \mathbb{C}^l$, we write their componentwise product as $vw = (v_i w_i)_{i=1}^l$. If $f\colon U \to \mathbb{C}$ is a function on $U \subset \mathbb{C}$, then we define $f(v) \in \mathbb{C}^l$ for $v \in U^l$ to be the vector with components $f(v)_i = f(v_i)$ for $i = 1, \ldots, l$. We will in particular apply this notation with $f(z) = 1/z$ for $z \in \mathbb{C} \setminus \{0\}$. We say that a vector $v \in \mathbb{C}^l$ is positive, $v > 0$, if $v_i > 0$ for all $i = 1, \ldots, l$. Similarly, the notation $v \leq w$ means $v_i \leq w_i$ for all $i = 1, \ldots, l$. For vectors $v, w \in \mathbb{C}^l$, we define $\langle w \rangle = l^{-1} \sum_{i=1}^l w_i$, $\langle v, w \rangle =$

$l^{-1}\sum_{i=1}^{l}\overline{v_i}w_i$, $\|w\|_2^2 = l^{-1}\sum_{i=1}^{l}|w_i|^2$ and $\|w\|_\infty = \max_{i=1,\dots,l}|w_i|$, $\|v\|_1 := \langle |v|\rangle$. Note that $\langle w\rangle = \langle 1, w\rangle$, where we used the convention that 1 also denotes the vector $(1,\dots,1) \in \mathbb{C}^l$. In general, we use the notation that if a scalar α appears in a vector-valued relation, then it denotes the constant vector $(\alpha,\dots,\alpha)$. In most cases we will work in n or $2n$ dimensional spaces. Vectors in $\mathbb{C}^{2n}$ will usually be denoted by boldface symbols like $\boldsymbol{v}$, $\boldsymbol{u}$ or $\boldsymbol{y}$. Correspondingly, capitalized boldface symbols denote matrices in $\mathbb{C}^{2n\times 2n}$, for example $\boldsymbol{R}$. We use the symbol $\mathbb{1}$ for the identity matrix in $\mathbb{C}^{l\times l}$, where the dimension $l = n$ or $l = 2n$ is understood from the context. For a matrix $A \in \mathbb{C}^{l\times l}$, we use the short notation $\|A\|_\infty := \|A\|_{\infty\to\infty}$ and $\|A\|_2 := \|A\|_{2\to 2}$ if the domain and the target are equipped with the same norm whereas we use $\|A\|_{2\to\infty}$ to denote the matrix norm of A when it is understood as a map $(\mathbb{C}^l, \|\cdot\|_2) \to (\mathbb{C}^l, \|\cdot\|_\infty)$. We define the normalized trace of an $l\times l$ matrix $B = (b_{ij})_{i,j=1}^l \in \mathbb{C}^{l\times l}$ as

$$\operatorname{tr} B := \frac{1}{l}\sum_{j=1}^{l} b_{jj}. \tag{6.1.3}$$

For a vector $y \in \mathbb{C}^l$, we write $\operatorname{diag} y$ or $\operatorname{diag}(y)$ for the diagonal $l\times l$ matrix with y on its diagonal, i.e., this matrix acts on any vector $x \in \mathbb{C}^l$ as

$$\operatorname{diag}(y)x = yx. \tag{6.1.4}$$

We write $\mathrm{d}^2 z$ for indicating integration with respect to the Lebesgue measure on $\mathbb{C}$. For $a \in \mathbb{C}$ and $\varepsilon > 0$, the open disk in the complex plane centered at a with radius ε is denoted by $D(a,\varepsilon) := \{b \in \mathbb{C} \mid |a-b| < \varepsilon\}$. Furthermore, we denote the characteristic function of some event A by $\mathbf{1}(A)$, the positive real numbers by $\mathbb{R}_+ := (0,\infty)$ and the nonnegative real numbers by $\mathbb{R}_0^+ := [0,\infty)$.

6.2. Main results

Let X be a random $n\times n$ matrix with centered entries, $\mathbb{E}\, x_{ij} = 0$, and $s_{ij} := \mathbb{E}|x_{ij}|^2$ the corresponding variances. We introduce its variance matrix $S := (s_{ij})_{i,j=1}^n$.

Assumptions 6.2.1. (A) The variance matrix S is *flat*, i.e., there are $0 < s_* < s^*$ such that

$$\frac{s_*}{n} \le s_{ij} \le \frac{s^*}{n} \tag{6.2.1}$$

for all $i, j = 1, \ldots, n$.

(B) All entries of X have bounded moments in the sense that there are $\mu_m > 0$ for $m \in \mathbb{N}$ such that

$$\mathbb{E}|x_{ij}|^m \leq \mu_m n^{-m/2} \tag{6.2.2}$$

for all $i, j = 1, \ldots, n$.

(C) Each entry of $\sqrt{n}\, X$ has a density, i.e., there are probability densities $f_{ij} \colon \mathbb{C} \to [0, \infty)$ such that

$$\mathbb{P}\left(\sqrt{n}\, x_{ij} \in B\right) = \int_B f_{ij}(z) \mathrm{d}^2 z$$

for all $i, j = 1, \ldots, n$ and $B \subset \mathbb{C}$ a Borel set. There are $\alpha, \beta > 0$ such that $f_{ij} \in L^{1+\alpha}(\mathbb{C})$ and

$$\|f_{ij}\|_{1+\alpha} \leq n^{\beta} \tag{6.2.3}$$

for all $i, j = 1, \ldots, n$.

In the following, we will assume that s_*, s^*, α, β and the sequence $(\mu_m)_m$ are fixed constants which we will call *model parameters*. The constants in all our estimates will depend on the model parameters without further notice.

Remark 6.2.2. The Assumption (C) is used in our proof solely for controlling the smallest singular value of $X - z\mathbb{1}$ with very high probability uniformly for $z \in D(0, \tau^*)$ with some fixed $\tau^* > 0$ in Proposition 6.5.9. All our other results do not make use of Assumption (C). Provided a version of Proposition 6.5.9 that tracks the z-dependence can effectively be obtained without (C), our main result, the local inhomogeneous circular law in Theorem 6.2.6, will hold true solely assuming (A) and (B). For example a very high probability estimate uniform in z in a statement similar to Corollary 1.22 of [50] would be sufficient.

The density of states of X will be expressed in terms of v_1^τ and v_2^τ which are the positive solutions of the following two coupled vector equations

$$\frac{1}{v_1^\tau} = \eta + Sv_2^\tau + \frac{\tau}{\eta + S^t v_1^\tau}, \tag{6.2.4a}$$

$$\frac{1}{v_2^\tau} = \eta + S^t v_1^\tau + \frac{\tau}{\eta + S v_2^\tau}, \tag{6.2.4b}$$

for all $\eta \in \mathbb{R}_+$ and $\tau \in \mathbb{R}_0^+$. Here, $v_1^\tau, v_2^\tau \in \mathbb{R}_+^n$ and recall that the algebraic operations are understood componentwise, e.g., $(1/v)_i = 1/v_i$ for the i-th component of the vector $1/v$. The system (6.2.4) is a special case of (6.1.2) with $w = \mathrm{i}\eta$, $\tau = |z|^2$ and $v_a = \operatorname{Im} m_a$ for $a = 1, 2$. The existence and uniqueness of solutions to equations of the type (6.2.4) are considered standard knowledge in the literature [82]. The equations can be viewed as a special case of the matrix Dyson equation for which existence and uniqueness was proven in [96]. We explain this connection in more detail in Section 6.6 below, where we give the proof of Lemma 6.2.3 for the convenience of the reader.

Lemma 6.2.3 (Existence and uniqueness). *For every $\tau \in \mathbb{R}_0^+$, there exist two uniquely determined functions $v_1^\tau \colon \mathbb{R}_+ \to \mathbb{R}_+^n$, $v_2^\tau \colon \mathbb{R}_+ \to \mathbb{R}_+^n$ which satisfy* (6.2.4).

We denote the spectral radius of S by $\rho(S)$, i.e.,

$$\rho(S) := \max|\mathrm{Spec}(S)|.$$

Now, we define the density of states of X through the solution to (6.2.4).

Definition 6.2.4 (Density of states of X)**.** Let v_1^τ and v_2^τ be the unique positive solutions of (6.2.4). The *density of states* $\sigma \colon \mathbb{C} \to \mathbb{R}$ of X is defined through

$$\sigma(z) := -\frac{1}{2\pi} \int_0^\infty \Delta_z \left\langle v_1^\tau(\eta)\Big|_{\tau=|z|^2} \right\rangle \mathrm{d}\eta \tag{6.2.5}$$

for $|z|^2 < \rho(S)$ and $\sigma(z) := 0$ for $|z|^2 \geq \rho(S)$. The right-hand side of (6.2.5) is well-defined by part (i) of the following proposition.

In the following proposition, we present some key properties of the density of states σ of X. Some of them have previously been known [51, 82]. For an alternative representation of σ, see (6.4.8) later.

Proposition 6.2.5 (Properties of σ)**.** *Let v_1^τ and v_2^τ be the unique positive solutions of* (6.2.4). *Then*

(i) The function $\mathbb{R}_+ \times \mathbb{C} \to \mathbb{R}_+^{2n}, (\eta, z) \mapsto (v_1^\tau(\eta), v_2^\tau(\eta))\,|_{\tau=|z|^2}$ is infinitely often differentiable and $\eta \mapsto \Delta_z \left\langle v_1^\tau(\eta)\big|_{\tau=|z|^2}\right\rangle$ is integrable on $\mathbb{R}_+$ for each $z \in D(0, \sqrt{\rho(S)})$.

(ii) The function σ, defined in (6.2.5)*, is a rotationally symmetric probability density on $\mathbb{C}$.*

(iii) The restriction $\sigma|_{D(0,\sqrt{\rho(S)})}$ is infinitely often differentiable such that for every $\varepsilon > 0$ each derivative is bounded uniformly in n on $D(0, \sqrt{\rho(S)} - \varepsilon)$. Moreover, there exist constants $c_1 > c_2 > 0$, which depend only on s_* and s^*, such that

$$c_1 \geq \sigma(z) \geq c_2 \tag{6.2.6}$$

for all $z \in D(0, \sqrt{\rho(S)})$. In particular, the support of σ is the closed disk of radius $\sqrt{\rho(S)}$ around zero. In fact, the jump height $\lim \sigma(z)$ as $|z| \uparrow \sqrt{\rho(S)}$ can be computed explicitly (see Remark 6.4.2 below).

The next theorem, the main result of the present article, states that the eigenvalue distribution of X, with a very high probability, can be approximated by σ on the mesoscopic scales n^{-a} for any $a \in (0, 1/2)$. Note that $n^{-1/2}$ is the typical eigenvalue spacing so our result holds down to the optimal local scale. To study the local scale, we shift and rescale the test functions as follows. Let $f \in C_0^2(\mathbb{C})$. For $w \in \mathbb{C}$ and $a > 0$, we define

$$f_{w,a} \colon \mathbb{C} \to \mathbb{C}, \quad f_{w,a}(z) := n^{2a} f(n^a(z - w)).$$

We denote the eigenvalues of X by $z_1, \dots, z_n$.

Theorem 6.2.6 (Local inhomogeneous circular law)**.** *Let X be a random matrix which has independent centered entries and satisfies (A), (B) and (C). Furthermore, let $a \in (0, 1/2)$, $\varphi > 0$, $\tau_* > 0$ and σ defined as in* (6.2.5)*.*

(i) (Bulk spectrum) For every $\varepsilon > 0$, $D > 0$, there is a positive constant $C_{\varepsilon,D}$ such that

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n f_{w,a}(z_i) - \int_{\mathbb{C}} f_{w,a}(z) \sigma(z) \mathrm{d}^2 z \right| \geq n^{-1+2a+\varepsilon} \|\Delta f\|_{L^1} \right) \leq \frac{C_{\varepsilon,D}}{n^D} \tag{6.2.7}$$

holds true for all $n \in \mathbb{N}$*, for every* $w \in \mathbb{C}$ *satisfying* $|w|^2 \leq \rho(S) - \tau_*$ *and for every* $f \in C_0^2(\mathbb{C})$ *satisfying* $\operatorname{supp} f \subset D(0, \varphi)$*. The point* w *and the function* f *may depend on* n*.*

(ii) (Away from the spectrum) For every $D > 0$*, there exists a positive constant* C_D *such that*

$$\mathbb{P}\left(\exists\, i \in \{1, \ldots, n\} \;\middle|\; |z_i|^2 \geq \rho(S) + \tau_*\right) \leq \frac{C_D}{n^D} \tag{6.2.8}$$

holds true for all $n \in \mathbb{N}$*.*

In addition to the model parameters, the constant $C_{\varepsilon,D}$ *in* (6.2.7) *depends only on* a*,* φ *and* τ_* *(apart from* ε *and* D*) and the constant* C_D *in* (6.2.8) *only on* τ_* *(apart from* D*).*

The key technical input for the proof of Theorem 6.2.6 is the local law for $\boldsymbol{H}^z$ (see Theorem 6.5.2). In Figure 6.1 below, we illustrate how the empirical spectral measure of X converges to σ for an example with a nontrivial variance profile S. We now state a simple corollary of the local law for $\boldsymbol{H}^z$ on the complete delocalization of the bulk eigenvectors of X.

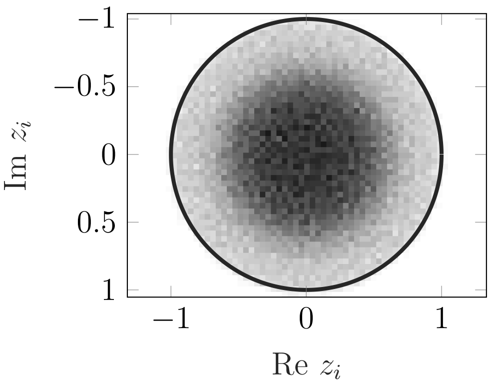
Corollary 6.2.7 (Eigenvector delocalization)**.** *Let* $\tau_* > 0$*. For all* $\varepsilon > 0$ *and* $D > 0$*, there is a positive constant* $C_{\varepsilon,D}$ *such that*

$$\mathbb{P}\left(\|y\|_\infty \geq n^{-1/2+\varepsilon}\right) \leq \frac{C_{\varepsilon,D}}{n^D} \tag{6.2.9}$$

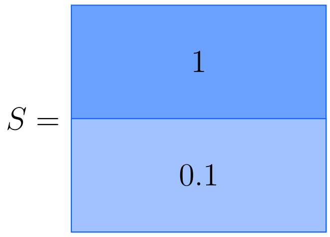
holds true for all $n \in \mathbb{N}$ *and for all eigenvectors* $y \in \mathbb{C}^n$ *of* X*, normalized as* $\sum_{i=1}^n |y_i|^2 = 1$*, corresponding to an eigenvalue* $z \in \operatorname{Spec} X$ *with* $|z|^2 \leq \rho(S) - \tau_*$*. The constant* $C_{\varepsilon,D}$ *in* (6.2.9) *depends only on* τ_* *and the model parameters (in addition to* ε *and* D*).*

The proof of Corollary 6.2.7 will be given after the statement of Theorem 6.5.2. We remark that eigenvector delocalization for random matrices with independent entries was first proven by Rudelson and Vershynin in [124].

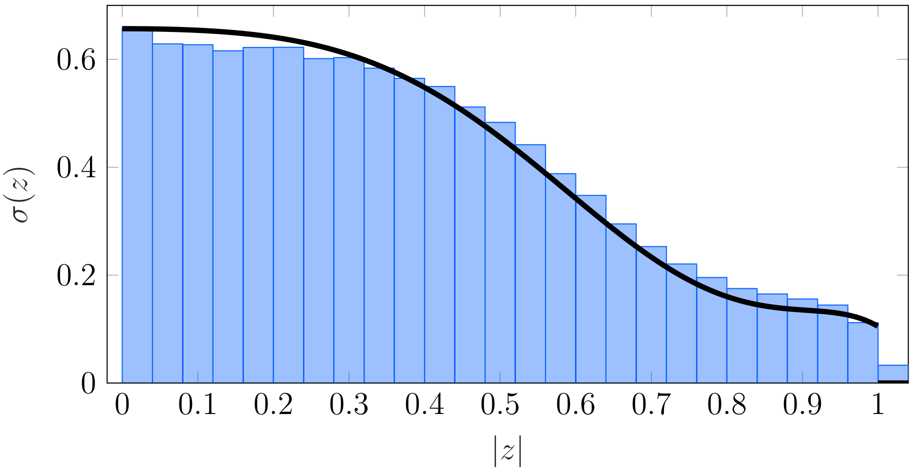
6.2.1. Short outline of the proof. We start with the Hermitization trick due to Girko which expresses $\sum_{i=1}^n f_{w,a}(z_i)$ in terms of an integral of the log-determinant of $X - z\mathbb{1}$ for any $z \in \mathbb{C}$. Furthermore, the log-determinant of $X - z\mathbb{1}$ can be rewritten as the log-determinant of a Hermitian matrix $\boldsymbol{H}^z$.



(A) Eigenvalue locations



(B) Variance profile S



(C) Eigenvalue histogram and density of states σ

FIGURE 6.1. These figures were obtained by sampling 200 matrices of size 2000×2000 with centered complex Gaussian entries and the variance profile S. Figure (a) shows the eigenvalue density for the variance profile S given in Figure (b) (We rescaled S such that $\rho(S) = 1$). The eigenvalue density is rotationally invariant and almost all eigenvalues are contained in the disk of radius 1 around zero. Moreover, the eigenvalue density is considerably higher around 0. Figure (c) compares the histogram of the eigenvalue with the density of states σ obtained from (6.2.4) and (6.2.5).

Using the log-transform of the empirical spectral measure of X, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} f_{w,a}(z_i) = \frac{1}{2\pi n}\int_{\mathbb{C}} \Delta f_{w,a}(z) \log|\det(X - z\mathbb{1})| \mathrm{d}^2 z. \tag{6.2.10}$$

To express the log-determinant of $X - z\mathbb{1}$ in terms of a Hermitian matrix, we introduce the $2n \times 2n$ matrix

$$\boldsymbol{H}^z := \begin{pmatrix} 0 & X - z\mathbb{1} \\ X^* - \bar{z}\mathbb{1} & 0 \end{pmatrix} \tag{6.2.11}$$

for all $z \in \mathbb{C}$. Note that the eigenvalues of $\boldsymbol{H}^z$ come in opposite pairs and we denote them by $\lambda_{2n} \leq \ldots \leq \lambda_{n+1} \leq 0 \leq \lambda_n \leq \ldots \leq \lambda_1$ with $\lambda_i = -\lambda_{2n+1-i}$ for $i = 1, \ldots, 2n$. We remark that the moduli of these real numbers are the singular values of $X - z\mathbb{1}$. The Stieltjes transform of its empirical spectral measure is denoted by m^z, i.e.,

$$m^z(\zeta) = \frac{1}{2n} \sum_{i=1}^{2n} \frac{1}{\lambda_i(z) - \zeta} \tag{6.2.12}$$

for $\zeta \in \mathbb{C}$ satisfying $\operatorname{Im} \zeta > 0$. It will turn out that on the imaginary axis $\operatorname{Im} m^z(\mathrm{i}\eta)$ is very well approximated by $\langle v_1^\tau(\eta) \rangle = \langle v_2^\tau(\eta) \rangle$, where $\tau = |z|^2$ and (v_1^τ, v_2^τ) is the solution of (6.2.4). This fact is commonly called a *local law* for $\boldsymbol{H}^z$. With this notation, we have the following relation between the determinant of $X - z\mathbb{1}$ and the determinant of $\boldsymbol{H}^z$

$$\log|\det(X - z\mathbb{1})| = \frac{1}{2} \log|\det \boldsymbol{H}^z|. \tag{6.2.13}$$

We write the log-determinant in terms of the Stieltjes transform (this formula was used by Tao and Vu [146] in a similar context)

$$\log|\det \boldsymbol{H}^z| = \log|\det(\boldsymbol{H}^z - \mathrm{i}T\mathbb{1})| - 2n \int_0^T \operatorname{Im} m^z(\mathrm{i}\eta) \mathrm{d}\eta, \tag{6.2.14}$$

for any $T > 0$. Combining (6.2.5), (6.2.10), (6.2.13) and (6.2.14) as well as subtracting $1/(1+\eta)$ freely and using integration by parts, we obtain

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f_{w,a}(z_i) - \int_{\mathbb{C}} f_{w,a}(z) \sigma(z) \mathrm{d}^2 z = & \\ & \frac{1}{4\pi n} \int_{\mathbb{C}} \Delta f_{w,a}(z) \log|\det(\boldsymbol{H}^z - \mathrm{i}T\mathbb{1})| \mathrm{d}^2 z \\ & - \frac{1}{2\pi} \int_{\mathbb{C}} \Delta f_{w,a}(z) \int_0^T \Big[\operatorname{Im} m^z(\mathrm{i}\eta) - \Big\langle v_1^\tau(\eta) \big|_{\tau = |z|^2} \Big\rangle \Big] \mathrm{d}\eta \, \mathrm{d}^2 z \\ & + \frac{1}{2\pi} \int_{\mathbb{C}} \Delta f_{w,a}(z) \int_T^\infty \left(\Big\langle v_1^\tau(\eta) \big|_{\tau = |z|^2} \Big\rangle - \frac{1}{\eta + 1} \right) \mathrm{d}\eta \, \mathrm{d}^2 z. \end{aligned} \tag{6.2.15}$$

The task is then to prove that each of the terms on the right-hand side of (6.2.15) is dominated by $n^{-1+2a}\|\Delta f\|_1$ with very high probability. The parameter T will be chosen to be a large power of n, so that the first and the third term will easily satisfy this bound. Estimating the second term on the right-hand side of (6.2.15) is much more involved and we focus only on this term in this outline.

We split its $\mathrm{d}\eta$ - integral into two parts. For $\eta \le n^{-1+\varepsilon}$, $\varepsilon \in (0, 1/2)$, the integral is controlled by an estimate on the smallest singular value of $X - z\mathbb{1}$. This is the only step in our proof which uses Assumption (C), i.e., that the entries of X have bounded densities in the sense of (6.2.3).

For $\eta \ge n^{-1+\varepsilon}$, we use a local law for $\boldsymbol{H}^z$, i.e., an optimal pointwise estimate (up to negligible n^ε-factors) on

$$\operatorname{Im} m^z(\mathrm{i}\eta) - \left\langle v_1^\tau(\eta)\Big|_{\tau=|z|^2}\right\rangle, \tag{6.2.16}$$

uniformly in η and z (see Theorem 6.5.2 for the precise formulation). Note that a local law for $\boldsymbol{H}^z$ is needed only at spectral parameters on the imaginary axis. This will simplify the proof of the local law we need in this paper.

The proof of the local law is based on a stability estimate of (6.2.4). To write these equations in a more concise form, we introduce the $2n \times 2n$ matrices

$$\boldsymbol{S}_o = \begin{pmatrix} 0 & S \\ S^t & 0 \end{pmatrix}, \quad \boldsymbol{S}_d = \begin{pmatrix} S^t & 0 \\ 0 & S \end{pmatrix}. \tag{6.2.17}$$

We remark that $\boldsymbol{S}_o$ is denoted by $\boldsymbol{S}$ in Chapter 4 and Chapter 5. Moreover, $\boldsymbol{H}$ in these chapters agrees with $\boldsymbol{H}^{z=0}$ from (6.2.11) at $z = 0$. With the notation from (6.2.17), the system of equations (6.2.4) can be written as

$$\mathrm{i}\boldsymbol{v} + \left(\mathrm{i}\eta + \boldsymbol{S}_o\mathrm{i}\boldsymbol{v} - \frac{\tau}{\mathrm{i}\eta + \boldsymbol{S}_d\mathrm{i}\boldsymbol{v}}\right)^{-1} = 0, \tag{6.2.18}$$

where we introduced $\boldsymbol{v} := (v_1, v_2) \in \mathbb{R}^{2n}$.

Let $\boldsymbol{G}^z(\eta) := (\boldsymbol{H}^z - \mathrm{i}\eta\mathbb{1})^{-1}$, $\eta > 0$, be the resolvent of $\boldsymbol{H}^z$ at spectral parameter $\mathrm{i}\eta$. We will prove that its diagonal $\boldsymbol{g}(\eta) = (\langle \boldsymbol{e}_i, \boldsymbol{G}^z(\eta)\boldsymbol{e}_i\rangle)_{i=1}^{2n}$, where $\boldsymbol{e}_i$ denotes the i-th standard basis vector in $\mathbb{C}^{2n}$, satisfies a perturbed version of (6.2.18),

$$\boldsymbol{g} + \left(\mathrm{i}\eta + \boldsymbol{S}_o \boldsymbol{g} - \frac{\tau}{\mathrm{i}\eta + \boldsymbol{S}_d \boldsymbol{g}} \right)^{-1} = \boldsymbol{d}, \tag{6.2.19}$$

with $\tau = |z|^2$ and a small random error term $\boldsymbol{d}$. As $m^z(\mathrm{i}\eta) = \langle \boldsymbol{g}(\eta) \rangle$ (cf. (6.2.12)) obtaining a local law, i.e., an optimal pointwise estimate on (6.2.16), reduces to a stability problem for the *Dyson equation*, (6.2.18).

Computing the difference of (6.2.19) and (6.2.18), we obtain

$$\boldsymbol{L} \left(\boldsymbol{g} - \mathrm{i}\boldsymbol{v} \right) = \boldsymbol{r} \tag{6.2.20}$$

for some error vector $\boldsymbol{r} = \mathcal{O}(\|\boldsymbol{d}\|)$ (for the precise definition we refer to (6.3.24) below) and with the matrix $\boldsymbol{L}$ defined through its action on $\boldsymbol{y} \in \mathbb{C}^{2n}$ via

$$\boldsymbol{L}\boldsymbol{y} := \boldsymbol{y} + \boldsymbol{v}^2 (\boldsymbol{S}_o \boldsymbol{y}) - \tau \frac{\boldsymbol{v}^2}{(\eta + \boldsymbol{S}_d \boldsymbol{v})^2} (\boldsymbol{S}_d \boldsymbol{y}). \tag{6.2.21}$$

Therefore, a bound on $\boldsymbol{g} - \mathrm{i}\boldsymbol{v}$ uniformly for $\eta \geq n^{-1+\varepsilon}$ requires a uniform bound on the inverse of $\boldsymbol{L}$ down to this local spectral scale.

In fact, the mere invertibility of $\boldsymbol{L}$ even for η bounded away from zero is a nontrivial fact that is not easily seen from (6.2.21). In Section 6.3 we will factorize $\boldsymbol{L}$ into the form

$$\boldsymbol{L} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{V}$$

for some invertible matrix $\boldsymbol{V}$ and self-adjoint matrices $\boldsymbol{T}$ and $\boldsymbol{F}$ with the properties $\|\boldsymbol{T}\|_2 = 1$ and $\|\boldsymbol{F}\|_2 \leq 1 - c\eta$ for some $c > 0$. In particular, this representation shows the a priori bound $\|\boldsymbol{L}^{-1}\|_2 \leq C\eta^{-1}$ for some $C > 0$. The blow-up in the norm of $\boldsymbol{L}^{-1}$ is potentially caused by the two extremal eigendirections $\boldsymbol{f}_+$ and $\boldsymbol{f}_-$ of $\boldsymbol{F}$, which satisfy

$$\boldsymbol{F}\boldsymbol{f}_\pm \;=\; \pm \|\boldsymbol{F}\|_2 \boldsymbol{f}_\pm \,.$$

However, it turns out that the positivity of the solutions v_1, v_2 of (6.2.4) implies that $\|\boldsymbol{T}\boldsymbol{f}_+\|_2$ is strictly smaller than 1, so that $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{f}_+\|_2 \geq c\|\boldsymbol{f}_+\|_2$ for some constant $c > 0$. In this sense the solution of the Dyson equation regularizes the potentially unstable direction $\boldsymbol{f}_+$.

In contrast, the other instability caused by $\boldsymbol{f}_-$ persists since we will find that $(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{f}_- = \mathcal{O}(\eta)$. This problem can only be resolved by exploiting an extra cancellation that originates from the special structure of the random matrix $\boldsymbol{H}^z$. The leading contribution of the random error $\boldsymbol{r} = \mathcal{O}(\|\boldsymbol{d}\|)$ from (6.2.20) pointing in the unstable direction happens to vanish with a remaining subleading term of order $\eta\|\boldsymbol{d}\|$. The extra η-factor cancels the η^{-1}-divergence of $\|\boldsymbol{L}^{-1}\|_2$ and allows us to invert the stability operator $\boldsymbol{L}$ in (6.2.20).

From this analysis, we conclude $\|\boldsymbol{g} - \mathrm{i}\boldsymbol{v}\| \le C\|\boldsymbol{d}\|$. This result allows us to follow the general arguments developed in [6] for verifying the optimal local law for $\boldsymbol{H}^z$. These steps are presented only briefly in Section 6.5.

6.3. Dyson equation for the inhomogeneous circular law

As explained in Section 6.2.1 a main ingredient in the proof of Theorem 6.2.6 is the local law for the self-adjoint random matrix $\boldsymbol{H}^z$ with noncentered independent entries above the diagonal. In [6] such a local law was proven for a large class of self-adjoint random matrices with noncentered entries and general short range correlations. For any fixed $z \in \mathbb{C}$, the matrix $\boldsymbol{H}^z$ satisfies the assumptions made for the class of random matrices covered in [6] with one crucial exception: $\boldsymbol{H}^z$ is not *flat* (cf. (2.28) in [6]), i.e., for any constant $c > 0$, the inequality

$$\frac{1}{n}\mathbb{E}\,|\langle \boldsymbol{a}, (\boldsymbol{H} - \mathbb{E}\,\boldsymbol{H})\boldsymbol{b}\rangle|^2 \ \ge\ c\|\boldsymbol{a}\|_2^2\|\boldsymbol{b}\|_2^2, \tag{6.3.1}$$

is not satisfied for $\boldsymbol{H} = \boldsymbol{H}^z$ and vectors $\boldsymbol{a}, \boldsymbol{b}$ that both have support either in $\{1, \dots, n\}$ or $\{n+1, \dots, 2n\}$. Nevertheless we will show that the conclusion from Theorem 2.9 of [6] remains true for spectral parameters $\mathrm{i}\eta$ on the imaginary axis, namely that the resolvent $\boldsymbol{G}^z(\eta) := (\boldsymbol{H}^z - \mathrm{i}\eta\mathbb{1})^{-1}$ approaches the solution $\boldsymbol{M}^z(\eta)$ of the *Matrix Dyson Equation (MDE)*

$$-\boldsymbol{M}^z(\eta)^{-1} \ =\ \mathrm{i}\eta\mathbb{1} - \boldsymbol{A}^z + \mathcal{S}[\boldsymbol{M}^z(\eta)]\,, \qquad \eta > 0\,, \tag{6.3.2}$$

as $n \to \infty$. In fact, the solution of (6.3.2) is unique under the constraint that the imaginary part $\operatorname{Im}\boldsymbol{M} := (\boldsymbol{M} - \boldsymbol{M}^*)/(2\mathrm{i})$ is positive definite [96]. The data $\boldsymbol{A}^z \in \mathbb{C}^{2n\times 2n}$

and $\mathcal{S}\colon \mathbb{C}^{2n\times 2n} \to \mathbb{C}^{2n\times 2n}$ determining (6.3.2) are given in terms of the first and second moments of the entries of $\boldsymbol{H}^z$,

$$\boldsymbol{A}^z := \mathbb{E}\,\boldsymbol{H}^z = \begin{pmatrix} 0 & -z \\ -\overline{z} & 0 \end{pmatrix}, \qquad \mathcal{S}[\boldsymbol{W}] := \begin{pmatrix} \operatorname{diag}(Sw_2) & 0 \\ 0 & \operatorname{diag}(S^t w_1) \end{pmatrix}, \tag{6.3.3}$$

for an arbitrary $2n \times 2n$ matrix

$$\boldsymbol{W} = (w_{ij})_{i,j=1}^{2n} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}, \qquad w_1 := (w_{ii})_{i=1}^{n}, \qquad w_2 := (w_{ii})_{i=n+1}^{2n}. \tag{6.3.4}$$

In the following, we will not keep the z-dependence in our notation and just write $\boldsymbol{M}$, $\boldsymbol{A}$ and $\boldsymbol{G}$ instead of $\boldsymbol{M}^z$, $\boldsymbol{A}^z$ and $\boldsymbol{G}^z$. A simple calculation (cf. the proof of Lemma 6.2.3 in Section 6.6 below) shows that $\boldsymbol{M}\colon \mathbb{R}_+ \to \mathbb{C}^{2n\times 2n}$ is given by

$$\boldsymbol{M}^z(\eta) := \begin{pmatrix} \mathrm{i}\operatorname{diag}\left(v_1^\tau(\eta)\right) & -z\operatorname{diag}\left(u^\tau(\eta)\right) \\ -\bar{z}\operatorname{diag}\left(u^\tau(\eta)\right) & \mathrm{i}\operatorname{diag}\left(v_2^\tau(\eta)\right) \end{pmatrix}, \tag{6.3.5}$$

where $z \in \mathbb{C}$, $\tau = |z|^2$, (v_1^τ, v_2^τ) is the solution of (6.2.4) and $u^\tau := v_1^\tau/(\eta + S^t v_1^\tau)$. In this section we will therefore analyze the solution and the stability of (6.2.4).

6.3.1. Analysis of the Dyson equation (6.2.4)**.** Combining the equations in (6.2.4), recalling $\boldsymbol{v} = (v_1, v_2)$ and the definitions of $\boldsymbol{S}_o$ and $\boldsymbol{S}_d$ in (6.2.17), we obtain

$$\frac{1}{\boldsymbol{v}} = \eta + \boldsymbol{S}_o\boldsymbol{v} + \frac{\tau}{\eta + \boldsymbol{S}_d\boldsymbol{v}} \tag{6.3.6}$$

for $\eta > 0$ and $\tau \in \mathbb{R}_0^+$, where $\boldsymbol{v}\colon \mathbb{R}_+ \to \mathbb{R}_+^{2n}$. This equation is equivalent to (6.2.18). The τ-dependence of $\boldsymbol{v}$, v_1 and v_2 will mostly be suppressed but sometimes we view $\boldsymbol{v} = \boldsymbol{v}^\tau(\eta)$ as a function of both parameters.

Equation (6.3.6) has an obvious scaling invariance when S is rescaled to λS for $\lambda > 0$. If $\boldsymbol{v}^\tau(\eta)$ is the positive solution of (6.3.6), then $\boldsymbol{v}_\lambda^\tau(\eta) := \lambda^{-1/2}\boldsymbol{v}^{\tau\lambda^{-1}}(\eta\lambda^{-1/2})$ is the positive solution of

$$\frac{1}{\boldsymbol{v}_\lambda} = \eta + \lambda\boldsymbol{S}_o\boldsymbol{v}_\lambda + \frac{\tau}{\eta + \lambda\boldsymbol{S}_d\boldsymbol{v}_\lambda}. \tag{6.3.7}$$

Therefore, without loss of generality, we may assume that the spectral radius of S is one,

$$\rho(S) = 1,$$

in the remainder of the paper.

The following proposition, the first main result of this section, collects some basic estimates on the solution $\boldsymbol{v}$ of (6.3.6). For the whole section, we fix $\tau_* > 0$ and $\tau^* > \tau_* + 1$ and except for Proposition 6.3.2, we exclude the small interval $[1-\tau_*, 1+\tau_*]$ from our analysis of $\boldsymbol{v}^\tau$. Because of the definition of σ in (6.2.5) – recall $\tau = |z|^2$ in the definition – we will talk about inside and outside regimes for $\tau \in [0, 1-\tau_*]$ and $\tau \in [1+\tau_*, \tau^*]$, respectively.

Recalling s_* and s^* from (6.2.1) we make the following convention in order to suppress irrelevant constants from the notation.

Convention 6.3.1. *For nonnegative scalars or vectors f and g, we will use the notation $f \lesssim g$ if there is a constant $c > 0$, depending only on τ_*, τ^*, s_* and s^* such that $f \leq cg$ and $f \sim g$ if $f \lesssim g$ and $f \gtrsim g$ both hold true. If f, g and h are scalars or vectors and $h \geq 0$ such that $|f - g| \lesssim h$, then we write $f = g + \mathcal{O}(h)$. Moreover, we define*

$$\mathcal{P} := \{\tau_*, \tau^*, s_*, s^*\}$$

because many constants in the following will depend only on $\mathcal{P}$.

Proposition 6.3.2. *The solution $\boldsymbol{v}^\tau$ of (6.3.6) satisfies*

$$\langle v_1^\tau(\eta) \rangle = \langle v_2^\tau(\eta) \rangle. \tag{6.3.8}$$

for all $\eta > 0$ and $\tau \in \mathbb{R}_0^+$ as well as the following estimates:

(i) (Large η) Uniformly for $\eta \geq 1$ and $\tau \in [0, \tau^*]$, we have

$$\boldsymbol{v}^\tau(\eta) \sim \eta^{-1}. \tag{6.3.9}$$

(ii) (Inside regime) Uniformly for $\eta \leq 1$ and $\tau \in [0, 1]$, we have

$$\boldsymbol{v}^\tau(\eta) \sim \eta^{1/3} + (1-\tau)^{1/2}. \tag{6.3.10}$$

(iii) (Outside regime) Uniformly for $\eta \leq 1$ and $\tau \in [1, \tau^*]$, we have

$$\boldsymbol{v}^\tau(\eta) \sim \frac{\eta}{\tau - 1 + \eta^{2/3}}. \tag{6.3.11}$$

Proof of Proposition 6.3.2. We start with proving (6.3.8). By multiplying (6.2.4a) by $(\eta + S^t v_1)$ and (6.2.4b) by $(\eta + S v_2)$ and realizing that both right-hand sides agree, we obtain

$$\frac{v_1}{\eta + S^t v_1} = \frac{v_2}{\eta + S v_2}. \tag{6.3.12}$$

From (6.3.12), we also get

$$0 = \eta(v_1 - v_2) + v_1 S v_2 - v_2 S^t v_1.$$

We take the average on both sides, use $\langle v_1 S v_2 \rangle = \langle v_1 \, , S v_2 \rangle = \langle v_2 S^t v_1 \rangle$ and divide by $\eta > 0$ to infer (6.3.8).

From (6.2.1), we immediately deduce the following auxiliary bounds

$$\langle v_1 \rangle \lesssim S^t v_1 \lesssim \langle v_1 \rangle, \quad \langle v_2 \rangle \lesssim S v_2 \lesssim \langle v_2 \rangle. \tag{6.3.13}$$

We start with establishing $\boldsymbol{v} \sim \langle \boldsymbol{v} \rangle$. Since the entries of S are strictly positive and $\rho(S) = 1$ there is a unique vector $p \in \mathbb{R}_+^n$ which has strictly positive entries such that

$$Sp = p, \quad \langle p \rangle = 1, \quad p \sim 1 \tag{6.3.14}$$

by the Perron-Frobenius Theorem and (6.2.1). We multiply (6.2.4a) by v_1 as well as $\eta + S^t v_1$ and obtain $\eta + S^t v_1 = v_1(\eta + S v_2)(\eta + S^t v_1) + \tau v_1$. Taking the scalar product with p and using $\langle p \rangle = 1$ and $\rho(S) = 1$ yield

$$\eta + \langle p v_1 \rangle = \Big\langle p v_1 (\eta + S^t v_1)(\eta + S v_2) \Big\rangle + \tau \langle p v_1 \rangle. \tag{6.3.15}$$

Therefore, (6.3.13), $\langle v_1 \rangle = \langle v_2 \rangle = \langle \boldsymbol{v} \rangle$ by (6.3.8) and (6.3.14) imply

$$\eta + \langle \boldsymbol{v} \rangle \sim \Big[(\eta + \langle \boldsymbol{v} \rangle)^2 + \tau \Big] \langle \boldsymbol{v} \rangle. \tag{6.3.16}$$

We use (6.3.13) in (6.2.4a) and (6.2.4b) to conclude

$$\boldsymbol{v} \sim \frac{1}{\eta + \langle \boldsymbol{v} \rangle + \frac{\tau}{\eta + \langle \boldsymbol{v} \rangle}} = \frac{\eta + \langle \boldsymbol{v} \rangle}{(\eta + \langle \boldsymbol{v} \rangle)^2 + \tau} \sim \langle \boldsymbol{v} \rangle, \tag{6.3.17}$$

where we applied (6.3.16) in the last step. Hence, it suffices to prove all estimates (6.3.9), (6.3.10) and (6.3.11) for $\boldsymbol{v}$ replaced by $\langle \boldsymbol{v} \rangle$ only.

We start with an auxiliary upper bound on $\langle \boldsymbol{v} \rangle$. By multiplying (6.3.6) with $\boldsymbol{v}$, we get $1 = \eta \boldsymbol{v} + \boldsymbol{v} \boldsymbol{S}_o \boldsymbol{v} + \tau \boldsymbol{v}/(\eta + \boldsymbol{S}_d \boldsymbol{v}) \geq \boldsymbol{v} \boldsymbol{S}_o \boldsymbol{v}$. Hence, $1 \geq \langle v_1 S v_2 \rangle \gtrsim \langle v_1 \rangle \langle v_2 \rangle = \langle \boldsymbol{v} \rangle^2$, where we used (6.3.13) in the second step and (6.3.8) in the last step.

Next, we show (6.3.9). Clearly, (6.3.6) implies $\boldsymbol{v} \leq \eta^{-1}$. Moreover, as $\tau \leq \tau^*$ and $\eta \geq 1 \gtrsim \langle \boldsymbol{v} \rangle$ we find $\eta \lesssim \eta^2 \langle \boldsymbol{v} \rangle$ from (6.3.16). This gives the lower bound on $\boldsymbol{v}$ in (6.3.9) when combined with (6.3.17).

We note that (6.3.16) immediately implies $\langle \boldsymbol{v} \rangle \gtrsim \eta$ for $\eta \leq 1$. Now, we show (6.3.10). For $\tau \in [0, 1]$, we bring the term $\tau \langle p v_1 \rangle$ to the left-hand side in (6.3.15) and use $v_1 \sim v_2 \sim \langle \boldsymbol{v} \rangle$ and (6.3.13) as well as $\langle \boldsymbol{v} \rangle \gtrsim \eta$ to obtain

$$\eta + (1 - \tau) \langle \boldsymbol{v} \rangle \sim \langle \boldsymbol{v} \rangle^3. \tag{6.3.18}$$

From (6.3.18), it is an elementary exercise to conclude (6.3.10) for $\eta \leq 1$.

Similarly, for $1 \leq \tau \leq \tau^*$, we bring $\langle p v_1 \rangle$ to the right-hand side of (6.3.15), use $\langle \boldsymbol{v} \rangle \gtrsim \eta$ for $\eta \leq 1$ and conclude

$$\eta \sim \langle \boldsymbol{v} \rangle^3 + (\tau - 1) \langle \boldsymbol{v} \rangle. \tag{6.3.19}$$

As before it is easy to conclude (6.3.11) from (6.3.19). We leave this to the reader. This completes the proof of Proposition 6.3.2. ☐

Our next goal is a stability result for (6.3.6) in the regime $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$. In the following proposition, the second main result of this section, we prove that $\mathrm{i}\boldsymbol{v}(\eta)$ well approximates $\boldsymbol{g}(\eta)$ for all $\eta > 0$ if $\boldsymbol{g}$ satisfies (6.2.19) and as long as $\boldsymbol{d}$ is small. However, we will need an additional assumption on $\boldsymbol{g} = (g_1, g_2)$, namely that $\langle g_1 \rangle = \langle g_2 \rangle$ (see (6.3.20) below). Note that this is imposed on the solution $\boldsymbol{g}$ of (6.2.19) and not directly on the perturbation $\boldsymbol{d}$. Nevertheless, in our applications, the constraint (6.3.20) will be automatically satisfied owing to the specific block structure of the matrix $\boldsymbol{H}^z$ from (6.2.11).

Proposition 6.3.3 (Stability). *Suppose that some functions $\boldsymbol{d} \colon \mathbb{R}_+ \to \mathbb{C}^{2n}$ and $\boldsymbol{g} = (g_1, g_2) \colon \mathbb{R}_+ \to \mathbb{H}^{2n}$ satisfy* (6.2.19) *and*

$$\langle g_1(\eta) \rangle = \langle g_2(\eta) \rangle \tag{6.3.20}$$

for all $\eta > 0$. There is a number $\lambda_* \gtrsim 1$, depending only on $\mathcal{P}$, such that

$$\|\boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\|_\infty \cdot \mathbf{1}\Big(\|\boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\|_\infty \le \lambda_*\Big) \lesssim \|\boldsymbol{d}(w)\|_\infty \tag{6.3.21}$$

uniformly for $\eta > 0$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$.

Moreover, there is a matrix-valued function $\boldsymbol{R}\colon \mathbb{R}_+ \to \mathbb{C}^{2n\times 2n}$, depending only on τ and S and satisfying $\|\boldsymbol{R}(\eta)\|_\infty \lesssim 1$, such that

$$|\langle \boldsymbol{y}, \boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\rangle| \cdot \mathbf{1}\Big(\|\boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\|_\infty \le \lambda_*\Big) \lesssim \|\boldsymbol{y}\|_\infty \|\boldsymbol{d}(\eta)\|_\infty^2 + |\langle \boldsymbol{R}(\eta)\boldsymbol{y}, \boldsymbol{d}(\eta)\rangle| \tag{6.3.22}$$

uniformly for all $\boldsymbol{y} \in \mathbb{C}^{2n}$, $\eta > 0$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$.

The proof of this result is based on deriving a quadratic equation for the difference $\boldsymbol{h} := \boldsymbol{g} - \mathrm{i}\boldsymbol{v}$ and establishing a quantitative estimate on $\boldsymbol{h}$ in terms of the perturbation $\boldsymbol{d}$. Computing the difference of (6.2.19) and (6.2.18), we obtain an equation for $\boldsymbol{g} - \mathrm{i}\boldsymbol{v}$. A straightforward calculation yields

$$\boldsymbol{L}\boldsymbol{h} = \boldsymbol{r}, \quad \text{for } \boldsymbol{h} = \boldsymbol{g} - \mathrm{i}\boldsymbol{v}, \tag{6.3.23}$$

where we used $\boldsymbol{L}$ defined in (6.2.21) and introduced the vector $\boldsymbol{r}$ through

$$\boldsymbol{r} := \boldsymbol{d} + \mathrm{i}\boldsymbol{v}(\boldsymbol{h} - \boldsymbol{d})\boldsymbol{S}_o\boldsymbol{h} - \tau\boldsymbol{u}\left[\frac{\boldsymbol{d} - \boldsymbol{g}}{\mathrm{i}\eta + \boldsymbol{S}_d\boldsymbol{g}} + \boldsymbol{u}\right]\boldsymbol{S}_d\boldsymbol{h}. \tag{6.3.24}$$

The vector $\boldsymbol{u}$ in (6.3.24) is defined through

$$u := \frac{v_1}{\eta + S^t v_1} = \frac{v_2}{\eta + S v_2}, \qquad \boldsymbol{u} := (u, u) = \frac{\boldsymbol{v}}{\eta + \boldsymbol{S}_d\boldsymbol{v}} \tag{6.3.25}$$

which is consistent by (6.3.12).

Notice that all terms on the right-hand side of (6.3.24) are either second order in $\boldsymbol{h}$ or they are of order $\boldsymbol{d}$, so (6.3.23) is the linearization of (6.2.19) around (6.2.18).

In the following estimates, we need a bound on $\boldsymbol{u}$ as well. Indeed, Proposition 6.3.2 yields

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\eta + \boldsymbol{S}_d\boldsymbol{v}} \sim \frac{1}{1+\eta^2} \tag{6.3.26}$$

uniformly for $\eta > 0$ and $\tau \in [0, \tau^*]$.

To shorten the upcoming relations, we introduce the vector

$$\widetilde{\boldsymbol{v}} := (v_2, v_1)$$

and the matrices $\boldsymbol{T}$, $\boldsymbol{F}$ and $\boldsymbol{V}$ defined by their action on a vector $\boldsymbol{y} = (y_1, y_2)$, $y_1, y_2 \in \mathbb{C}^n$ as follows

$$\boldsymbol{T}\boldsymbol{y} := \frac{1}{\boldsymbol{u}}\begin{pmatrix} -v_1 v_2 y_1 + \tau u^2 y_2 \\ \tau u^2 y_1 - v_1 v_2 y_2 \end{pmatrix}, \tag{6.3.27a}$$

$$\boldsymbol{F}\boldsymbol{y} := \sqrt{\frac{\boldsymbol{v}\boldsymbol{u}}{\widetilde{\boldsymbol{v}}}}\boldsymbol{S}_o\left(\sqrt{\frac{\boldsymbol{v}\boldsymbol{u}}{\widetilde{\boldsymbol{v}}}}\,\boldsymbol{y}\right), \tag{6.3.27b}$$

$$\boldsymbol{V}\boldsymbol{y} := \sqrt{\frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}\boldsymbol{v}}}\,\boldsymbol{y}. \tag{6.3.27c}$$

All these matrices are functions of η and τ. They provide a crucial factorization of the stability operator $\boldsymbol{L}$; indeed, a simple calculation shows that

$$\boldsymbol{L} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{V}. \tag{6.3.28}$$

This factorization reveals many properties of $\boldsymbol{L}$ which are difficult to observe directly. Owing to (6.3.23), the stability analysis of (6.3.6) requires a control on the invertibility of the matrix $\boldsymbol{L}$. The matrices $\boldsymbol{V}$ and $\boldsymbol{V}^{-1}$ are harmless. A good understanding of the spectral decompositions of the simpler matrices $\boldsymbol{F}$ and $\boldsymbol{T}$ will then yield that $\boldsymbol{L}$ has only one direction, in which its inverse is not bounded. We remark that the factorization (6.3.28) is the diagonal part of the one used in the stability analysis of the matrix Dyson equation in [6].

Because of (6.3.28), we can study the stability of

$$(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})(\boldsymbol{V}\boldsymbol{h}) = \boldsymbol{V}\boldsymbol{r} \tag{6.3.29}$$

instead of (6.3.23). From Proposition 6.3.2 and (6.3.26), we conclude that

$$\|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1 \tag{6.3.30}$$

uniformly for all $\eta > 0$ and $\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$. Hence, it suffices to control the invertibility of $\mathbb{1} - \boldsymbol{T}\boldsymbol{F}$.

For later usage, we derive two relations for $\boldsymbol{u}$. From (6.3.25), recalling $\widetilde{\boldsymbol{v}} = (v_2, v_1)$, we immediately get

$$\frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} = \eta + \boldsymbol{S}_o\boldsymbol{v}. \tag{6.3.31}$$

We multiply (6.3.6) by $\boldsymbol{v}\boldsymbol{u}$ and use (6.3.31) to obtain

$$\boldsymbol{u} = \boldsymbol{v}\widetilde{\boldsymbol{v}} + \tau\boldsymbol{u}^2, \qquad 1 = \frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} + \tau\boldsymbol{u}. \tag{6.3.32}$$

The next lemma collects some properties of $\boldsymbol{F}$. For this formulation, we introduce

$$\boldsymbol{e}_- := (1, -1) \in \mathbb{C}^{2n}.$$

Lemma 6.3.4 (Spectral properties of $\boldsymbol{F}$)**.** *The eigenspace of $\boldsymbol{F}$ corresponding to its largest eigenvalue $\|\boldsymbol{F}\|_2$ is one dimensional. It is spanned by a unique positive normalized eigenvector $\boldsymbol{f}_+$, i.e., $\boldsymbol{F}\boldsymbol{f}_+ = \|\boldsymbol{F}\|_2\boldsymbol{f}_+$ and $\|\boldsymbol{f}_+\|_2 = 1$. For every $\eta > 0$, the norm of $\boldsymbol{F}$ is given by*

$$\|\boldsymbol{F}\|_2 = 1 - \eta\frac{\left\langle \boldsymbol{f}_+\sqrt{\boldsymbol{v}/(\eta + \boldsymbol{S}_o\boldsymbol{v})}\right\rangle}{\left\langle \boldsymbol{f}_+\sqrt{\boldsymbol{v}(\eta + \boldsymbol{S}_o\boldsymbol{v})}\right\rangle}. \tag{6.3.33}$$

Defining $\boldsymbol{f}_- := \boldsymbol{f}_+\boldsymbol{e}_-$, we have

$$\boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2\boldsymbol{f}_-. \tag{6.3.34}$$

(i) (Inside regime) The following estimates hold true uniformly for $\tau \in [0, 1 - \tau_*]$. We have

$$1 - \|\boldsymbol{F}\|_2 \sim \eta. \tag{6.3.35}$$

uniformly for $\eta \in (0, 1]$. Furthermore, uniformly for $\eta \geq 1$, we have

$$1 - \|\boldsymbol{F}\|_2 \sim 1. \tag{6.3.36}$$

Moreover, uniformly for $\eta \in (0, 1]$, $\boldsymbol{f}_+$ satisfies

$$\boldsymbol{f}_+ \sim 1 \tag{6.3.37}$$

and there is $\varepsilon \sim 1$ such that

$$\|\boldsymbol{F}\boldsymbol{x}\|_2 \leq (1 - \varepsilon)\|\boldsymbol{x}\|_2 \tag{6.3.38}$$

for all $\boldsymbol{x} \in \mathbb{C}^{2n}$ *satisfying* $\boldsymbol{x} \perp \boldsymbol{f}_+$ *and* $\boldsymbol{x} \perp \boldsymbol{f}_-$*.*

(ii) (Outside regime) Uniformly for all $\eta > 0$ *and* $\tau \in [1+\tau_*, \tau^*]$*, we have*

$$1 - \|\boldsymbol{F}\|_2 \sim 1. \tag{6.3.39}$$

Proof. The statements about the eigenspace corresponding to $\|\boldsymbol{F}\|_2$ and $\boldsymbol{f}_+$ follow from Lemma 4.3.3 in Chapter 4.

For the proof of (6.3.33), we multiply (6.3.6) by $\boldsymbol{v}$ and take the scalar product of the resulting relation with $\boldsymbol{f}_+\sqrt{\boldsymbol{u}/(\boldsymbol{v}\widetilde{\boldsymbol{v}})}$. Using that

$$\begin{aligned}\left\langle \boldsymbol{f}_+\sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}\,,\, \boldsymbol{v}\boldsymbol{S}_o\boldsymbol{v}\right\rangle &= \left\langle \boldsymbol{f}_+\sqrt{\frac{\boldsymbol{v}\boldsymbol{u}}{\widetilde{\boldsymbol{v}}}}\,,\, \boldsymbol{S}_o\boldsymbol{v}\right\rangle = \left\langle \boldsymbol{S}_o\Big(\boldsymbol{f}_+\sqrt{\frac{\boldsymbol{v}\boldsymbol{u}}{\widetilde{\boldsymbol{v}}}}\Big)\,,\, \boldsymbol{v}\right\rangle \\ &= \left\langle \sqrt{\frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{v}\boldsymbol{u}}}\boldsymbol{F}\boldsymbol{f}_+\,,\, \boldsymbol{v}\right\rangle = \|\boldsymbol{F}\|_2\left\langle \boldsymbol{f}_+\,,\, \sqrt{\frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}}\right\rangle,\end{aligned}$$

this yields

$$\|\boldsymbol{F}\|_2\left\langle \boldsymbol{f}_+\,,\, \sqrt{\frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}}\right\rangle = \left\langle \boldsymbol{f}_+\sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}\,,\, 1-\tau\boldsymbol{u}\right\rangle - \eta\left\langle \boldsymbol{f}_+\sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}\,,\, \boldsymbol{v}\right\rangle.$$

We conclude (6.3.33) from applying (6.3.32) and (6.3.31) to the last relation.

Since $\boldsymbol{F}$ from (6.3.27b) has the form

$$\boldsymbol{F} = \begin{pmatrix} 0 & F \\ F^t & 0 \end{pmatrix},$$

for some $F \in \mathbb{C}^{n\times n}$ we have $\boldsymbol{F}(\boldsymbol{e}_-\boldsymbol{y}) = -\boldsymbol{e}_-(\boldsymbol{F}\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{C}^{2n}$. Thus, we get (6.3.34) from $\boldsymbol{F}\boldsymbol{f}_+ = \|\boldsymbol{F}\|_2\boldsymbol{f}_+$.

In the regime $\tau \in [0, 1-\tau_*]$ and $\eta \in (0,1]$, we have uniform lower and upper bounds on $\boldsymbol{v}$ from Proposition 6.3.2. Therefore, the estimates in (6.3.37) and (6.3.38) follow from Lemma 4.3.3 in Chapter 4. Combining (6.3.37), (6.3.33) and Proposition 6.3.2 yields (6.3.35). In the large η regime, i.e., for $\eta \geq 1$, since $\boldsymbol{v} \sim \eta^{-1}$ by Proposition 6.3.2 we obtain

$$\frac{\boldsymbol{v}}{\eta + \boldsymbol{S}_o\boldsymbol{v}} \sim \eta^{-2}, \quad \boldsymbol{v}(\eta + \boldsymbol{S}_o\boldsymbol{v}) \sim 1. \tag{6.3.40}$$

Hence, as $\boldsymbol{f}_+ > 0$ we conclude

$$\frac{\left\langle \boldsymbol{f}_+\sqrt{\boldsymbol{v}/(\eta+\boldsymbol{S}_o\boldsymbol{v})}\right\rangle}{\left\langle \boldsymbol{f}_+\sqrt{\boldsymbol{v}(\eta+\boldsymbol{S}_o\boldsymbol{v})}\right\rangle} \sim \frac{\langle \boldsymbol{f}_+\rangle}{\langle \boldsymbol{f}_+\rangle}\frac{1}{\eta} = \frac{1}{\eta}, \tag{6.3.41}$$

uniformly for all $\eta \geq 1$. This shows that (6.3.36) holds true for all $\eta \geq 1$ and $\tau \in [0, 1-\tau_*]$.

We now turn to the proof of (ii). If $\tau \in [1+\tau_*, \tau^*]$, then $\boldsymbol{v} \sim \eta$ by (6.3.11) for $\eta \leq 1$ and therefore

$$\frac{\boldsymbol{v}}{\eta+\boldsymbol{S}_o\boldsymbol{v}} \sim 1, \quad \boldsymbol{v}(\eta+\boldsymbol{S}_o\boldsymbol{v}) \sim \eta^2.$$

As $\boldsymbol{f}_+ > 0$, we thus have

$$\eta\frac{\left\langle \boldsymbol{f}_+\sqrt{\boldsymbol{v}/(\eta+\boldsymbol{S}_o\boldsymbol{v})}\right\rangle}{\left\langle \boldsymbol{f}_+\sqrt{\boldsymbol{v}(\eta+\boldsymbol{S}_o\boldsymbol{v})}\right\rangle} \sim \frac{\langle \boldsymbol{f}_+\rangle}{\langle \boldsymbol{f}_+\rangle} = 1. \tag{6.3.42}$$

For $\eta \geq 1$, we argue as in (6.3.40) and (6.3.41) and arrive at the same conclusion (6.3.42). Thus, because of (6.3.33) the estimate (6.3.39) holds true for all $\eta > 0$ and $\tau \in [1+\tau_*, \tau^*]$. □

Next, we give an approximation for the eigenvector $\boldsymbol{f}_-$ belonging to the isolated single eigenvalue $-\|\boldsymbol{F}\|_2$ of $\boldsymbol{F}$ by constructing an approximate eigenvector. For $\eta \leq 1$ and $\tau \in [0, 1-\tau_*]$, we define

$$\boldsymbol{a} := \frac{\boldsymbol{e}_-(\boldsymbol{V}\boldsymbol{v})}{\|\boldsymbol{V}\boldsymbol{v}\|_2} \tag{6.3.43}$$

which is normalized as $\|\boldsymbol{e}_-(\boldsymbol{V}\boldsymbol{v})\|_2 = \|\boldsymbol{V}\boldsymbol{v}\|_2$. We compute

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{V}\boldsymbol{v}) &= \sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}\boldsymbol{v}\,(\boldsymbol{S}_o\boldsymbol{v}) = \sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}}\,(1-\eta\boldsymbol{v}-\tau\boldsymbol{u}) \\ &= \sqrt{\frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}} - \eta\boldsymbol{v}\sqrt{\frac{\boldsymbol{u}}{\boldsymbol{v}\widetilde{\boldsymbol{v}}}} = \|\boldsymbol{F}\|_2\boldsymbol{V}\boldsymbol{v} + \mathcal{O}(\eta). \end{aligned} \tag{6.3.44}$$

Here, we used $\boldsymbol{v}\boldsymbol{S}_o\boldsymbol{v} = -\eta\boldsymbol{v} + \boldsymbol{v}\widetilde{\boldsymbol{v}}/\boldsymbol{u}$ by (6.3.31). For estimating the $\mathcal{O}(\eta)$ term we applied (6.3.10), (6.3.26) and (6.3.35) since $\tau \in [0, 1-\tau_*]$ and $\eta \leq 1$. Using the block structure of $\boldsymbol{F}$ as in the proof of (6.3.34), we obtain

$$\boldsymbol{F}(\boldsymbol{e}_-(\boldsymbol{V}\boldsymbol{v})) = -\|\boldsymbol{F}\|_2\boldsymbol{e}_-(\boldsymbol{V}\boldsymbol{v}) + \mathcal{O}(\eta). \tag{6.3.45}$$

The following lemma states that $\boldsymbol{a}$ approximates the nondegenerate eigenvector $\boldsymbol{f}_-$.

Lemma 6.3.5. *The eigenvector $\boldsymbol{f}_-$ can be approximated by $\boldsymbol{a}$ in the ℓ^∞-norm, i.e.,*

$$\|\boldsymbol{f}_- - \boldsymbol{a}\|_\infty = \mathcal{O}(\eta) \tag{6.3.46}$$

uniformly for $\eta \leq 1$ and $\tau \in [0, 1-\tau_*]$.

Lemma 6.3.5 is proven at the end of Section 6.7 below. In the following lemma, we show some properties of $\boldsymbol{T}$.

Lemma 6.3.6 (Spectral properties of $\boldsymbol{T}$)**.** *The symmetric operator $\boldsymbol{T}$, defined in* (6.3.27a)*, satisfies*

(i) $\|\boldsymbol{T}\|_2 = 1$, $\|\boldsymbol{T}\|_\infty = 1$.

(ii) The spectrum of $\boldsymbol{T}$ is given by

$$\operatorname{Spec}(\boldsymbol{T}) = \{-1\} \cup \left\{ \tau \boldsymbol{u}_i - \frac{(\boldsymbol{v}\widetilde{\boldsymbol{v}})_i}{\boldsymbol{u}_i} \,\middle|\, i = 1, \ldots, n \right\}.$$

(iii) For all $\eta > 0$, we have $\boldsymbol{T}(\tau = 0) = -\mathbb{1}$ and if $\tau > 0$, then the eigenspace of $\boldsymbol{T}$ corresponding to the eigenvalue -1 is n-fold degenerate and given by

$$\operatorname{Eig}(-1, \boldsymbol{T}) = \{(y, -y) | y \in \mathbb{C}^n\}. \tag{6.3.47}$$

(iv) The spectrum of $\boldsymbol{T}$ is strictly away from one, i.e., there is $\varepsilon > 0$, depending only on $\mathcal{P}$, such that

$$\operatorname{Spec}(\boldsymbol{T}) \subset [-1, 1-\varepsilon] \tag{6.3.48}$$

uniformly for $\tau \in [0, 1-\tau_*]$ and $\eta \in (0, 1]$.

Proof. The second relation in (6.3.32) implies $\|\boldsymbol{T}\|_\infty = 1$ and $\boldsymbol{T}(\tau = 0) = -\mathbb{1}$. Moreover, it yields that all vectors of the form $(y, -y)$ for $y \in \mathbb{C}^n$ are contained in $\operatorname{Eig}(-1, \boldsymbol{T})$. We define the vector $\boldsymbol{y}^{(j)} \in \mathbb{C}^{2n}$ by $\boldsymbol{y}^{(j)} := (\delta_{i,j} + \delta_{i,j+n})_{i=1}^{2n}$ and observe that

$$\boldsymbol{T}\boldsymbol{y}^{(j)} = \left(\tau \boldsymbol{u}_j - \frac{(\boldsymbol{v}\widetilde{\boldsymbol{v}})_j}{\boldsymbol{u}_j} \right) \boldsymbol{y}^{(j)}$$

for $j = 1, \ldots, n$. Counting dimensions implies that we have found all eigenvalues, hence (ii) follows. For $\tau > 0$, we have $\tau \boldsymbol{u}_j - (\boldsymbol{v}\widetilde{\boldsymbol{v}})_j / \boldsymbol{u}_j = 2\tau \boldsymbol{u}_j - 1 > -1$ by (6.3.32) and $\boldsymbol{u}_j > 0$ for all $j = 1, \ldots, n$. This yields the missing inclusion in (6.3.47). Since $\boldsymbol{T}$ is a symmetric operator, $\|\boldsymbol{T}\|_2 = 1$ follows from (ii) and $|\tau \boldsymbol{u} - \boldsymbol{v}\widetilde{\boldsymbol{v}}/\boldsymbol{u}| \leq 1$ by (6.3.32).

For the proof of (iv), we remark that there is $\varepsilon > 0$, depending only on $\mathcal{P}$, such that $2\boldsymbol{v}\widetilde{\boldsymbol{v}}/\boldsymbol{u} \geq \varepsilon$ for all $\eta \in (0,1]$ and $\tau \in [0, 1-\tau_*]$ by (6.3.10) and (6.3.26). Thus,

$$\tau\boldsymbol{u} - \frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} = 1 - 2\frac{\boldsymbol{v}\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} \leq 1 - \varepsilon$$

by (6.3.32). This concludes the proof of the lemma. $\square$

Now we are ready to give a proof of Proposition 6.3.3 based on inverting $\mathbb{1} - \boldsymbol{TF}$.

PROOF OF PROPOSITION 6.3.3. We recall that $\boldsymbol{h} = \boldsymbol{g} - \mathrm{i}\boldsymbol{v}$. Throughout the proof we will omit arguments, but we keep in mind that $\boldsymbol{g}$, $\boldsymbol{d}$, $\boldsymbol{h}$ and $\boldsymbol{v}$ depend on η and τ. The proof will be given in three steps.

The first step is to control $\|\boldsymbol{r}\|_\infty$ from (6.3.24) in terms of $\|\boldsymbol{h}\|_\infty^2$ and $\|\boldsymbol{d}\|_\infty$, i.e., to show

$$\|\boldsymbol{r}\|_\infty \mathbf{1}(\|\boldsymbol{h}\|_\infty \leq 1) \lesssim \|\boldsymbol{h}\|_\infty^2 + \|\boldsymbol{d}\|_\infty. \tag{6.3.49}$$

Inverting $\boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{TF})\boldsymbol{V}$ in (6.3.29), controlling the norm of the inverse and choosing $\lambda_* \leq 1$ small enough, we will conclude Proposition 6.3.3 from (6.3.49). For any $\eta_* \in (0,1]$, depending only on $\mathcal{P}$, this argument will be done in the second step for $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$ and $\eta \geq \eta_*$ as well as for $\tau \in [1+\tau_*, \tau^*]$ and $\eta \in (0, \eta_*]$. In the third step, we consider the most interesting regime $\tau \in [0, 1-\tau_*]$ and $\eta \leq \eta_*$ for a sufficiently small η_*, depending on $\mathcal{P}$ only. In this regime, we will use an extra cancellation for the contribution of $\boldsymbol{r}$ in the unstable direction of $\boldsymbol{L}$.

Step 1: For all $\eta > 0$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$, (6.3.49) holds true.
From (6.2.19), we obtain

$$\tau\frac{\boldsymbol{g} - \boldsymbol{d}}{\mathrm{i}\eta + \boldsymbol{S}_d\boldsymbol{g}} = 1 + (\mathrm{i}\eta + \boldsymbol{S}_o\boldsymbol{g})(\boldsymbol{g} - \boldsymbol{d}).$$

We start from (6.3.24), use the previous relation, $\tau\boldsymbol{u} = 1 + \mathrm{i}\boldsymbol{v}(\mathrm{i}\eta + \boldsymbol{S}_o\mathrm{i}\boldsymbol{v})$ by (6.3.6) and $\widetilde{\boldsymbol{v}} = (v_2, v_1) = \boldsymbol{u}(\eta + \boldsymbol{S}_o\boldsymbol{v})$ by (6.3.32) and get

$$\begin{aligned}
\boldsymbol{r} &= \boldsymbol{d} + \mathrm{i}\boldsymbol{v}(\boldsymbol{h} - \boldsymbol{d})\boldsymbol{S}_o\boldsymbol{h} - \boldsymbol{u}\left[\mathrm{i}\boldsymbol{v}(\mathrm{i}\eta + \boldsymbol{S}_o\mathrm{i}\boldsymbol{v}) - (\boldsymbol{g} - \boldsymbol{d})(\mathrm{i}\eta + \boldsymbol{S}_o\boldsymbol{g})\right]\boldsymbol{S}_d\boldsymbol{h} \\
&= \boldsymbol{d} + \mathrm{i}\boldsymbol{v}(\boldsymbol{h} - \boldsymbol{d})\boldsymbol{S}_o\boldsymbol{h} + \boldsymbol{u}\left[\boldsymbol{h}(\mathrm{i}\eta + \boldsymbol{S}_o\mathrm{i}\boldsymbol{v}) + \boldsymbol{g}\boldsymbol{S}_o\boldsymbol{h}\right]\boldsymbol{S}_d\boldsymbol{h} - \boldsymbol{d}\boldsymbol{u}(\mathrm{i}\eta + \boldsymbol{S}_o\boldsymbol{g})\boldsymbol{S}_d\boldsymbol{h} \qquad (6.3.50) \\
&= \mathrm{i}\boldsymbol{v}\boldsymbol{h}\boldsymbol{S}_o\boldsymbol{h} + \mathrm{i}\widetilde{\boldsymbol{v}}\boldsymbol{h}\boldsymbol{S}_d\boldsymbol{h} + \boldsymbol{u}\boldsymbol{g}\boldsymbol{S}_o\boldsymbol{h}\boldsymbol{S}_d\boldsymbol{h} + \boldsymbol{d} - \mathrm{i}\boldsymbol{v}\boldsymbol{d}\boldsymbol{S}_o\boldsymbol{h} - \boldsymbol{d}\boldsymbol{u}(\mathrm{i}\eta + \boldsymbol{S}_o\boldsymbol{g})\boldsymbol{S}_d\boldsymbol{h}.
\end{aligned}$$

Notice that the first three terms are quadratic in $\boldsymbol{h}$ (the linear terms dropped out), while the last three terms are controlled by $\boldsymbol{d}$. Now, we show that all other factors are bounded and hence irrelevant whenever $\|\boldsymbol{g}-\mathrm{i}\boldsymbol{v}\|_\infty \leq \lambda_*$ for $\eta > 0$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$. In this case, we conclude $\|\boldsymbol{g}\|_\infty \lesssim 1$ uniformly for all $\eta > 0$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$ by (6.3.9) and (6.3.10) from Proposition 6.3.2. Therefore, starting from (6.3.50) and using $\|\boldsymbol{v}\|_\infty \lesssim 1$ by (6.3.9) and (6.3.10), and $\|\boldsymbol{u}\|_\infty \lesssim 1$ by (6.3.26), we obtain (6.3.49).

Step 2: For any $\eta_* \in (0, 1]$, there exists $\lambda_* \gtrsim 1$, depending only on $\mathcal{P}$ and η_*, such that (6.3.21) holds true for $\eta \geq \eta_*$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*]$ and $\tau \in [1+\tau_*, \tau^*]$.
Moreover, with this choice of λ_*, (6.3.22) holds true in these (η, τ) parameter regimes as well.

Within Step 2, we redefine the comparison relation to depend both on $\mathcal{P}$ and η_*. Later in Step 3 we will choose an appropriate η_* depending only on $\mathcal{P}$, so eventually the comparison relations for our choice will depend only on $\mathcal{P}$.

We are now working in the regime, where $\eta \geq \eta_*$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$ or $\eta \in (0, \eta_*]$ and $\tau \in [1+\tau_*, \tau^*]$. In this case, to prove (6.3.21), we invert $\boldsymbol{L} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{V}$ (cf. (6.2.21)) in $\boldsymbol{L}\boldsymbol{h} = \boldsymbol{r}$, bound $\|\boldsymbol{L}^{-1}\|_\infty \lesssim 1$, which is proven below, and conclude

$$\|\boldsymbol{h}\|_\infty \mathbf{1}(\|\boldsymbol{h}\|_\infty \leq 1) \lesssim \|\boldsymbol{h}\|_\infty^2 + \|\boldsymbol{d}\|_\infty$$

from (6.3.49) for $\eta \geq \eta_*$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1-\tau_*]$. This means that there are $\Psi_1, \Psi_2 \sim 1$ such that

$$\|\boldsymbol{h}\|_\infty \mathbf{1}(\|\boldsymbol{h}\|_\infty \leq 1) \leq \Psi_1 \|\boldsymbol{h}\|_\infty^2 + \Psi_2 \|\boldsymbol{d}\|_\infty.$$

Choosing $\lambda_* := \min\{1, (2\Psi_1)^{-1}\}$ this yields

$$\|\boldsymbol{h}\|_\infty \mathbf{1}(\|\boldsymbol{h}\|_\infty \leq \lambda_*) \leq 2\Psi_2 \|\boldsymbol{d}\|_\infty.$$

Thus, we are left with controlling $\|\boldsymbol{L}^{-1}\|_\infty$, i.e., proving $\|\boldsymbol{L}^{-1}\|_\infty \lesssim 1$.

In the regime $\eta \geq \eta_*$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$, we have $\boldsymbol{v} \sim 1/\eta$ by Proposition 6.3.2 and $\boldsymbol{u} \sim 1/\eta^2$ by (6.3.26). Hence, $\boldsymbol{V} \sim \eta$ and $\boldsymbol{V}^{-1} \sim 1/\eta$. Therefore, $\|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1$ and due to $\|\boldsymbol{L}^{-1}\|_\infty \lesssim \|\boldsymbol{V}^{-1}\|_\infty \|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\|_\infty \|\boldsymbol{V}\|_\infty$, it suffices to show $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\|_\infty \lesssim 1$. Basic facts on the operator $\mathbb{1} - \boldsymbol{TF}$ are collected in Lemma 6.7.1 in Section 6.7 below. In particular, because of (6.7.9), the ℓ^∞ bound follows from $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\|_2 \lesssim 1$. Using (6.3.35), (6.3.36) and (6.3.39), we get that $1 - \|\boldsymbol{F}\|_2 \sim 1$ for all $\eta \geq \eta_*$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$. Hence, $1 - \|\boldsymbol{TF}\|_2 \sim 1$ by Lemma 6.3.6 (i), so the bound $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\|_2 \lesssim 1$ immediately follows. This proves (6.3.21) for $\eta \geq \eta_*$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$.

For $\eta \leq \eta_*$ and $\tau \in [1+\tau_*, \tau^*]$, we have $\boldsymbol{v} \sim \eta$ by (6.3.11), $\boldsymbol{u} \sim 1$ by (6.3.26). Thus, $\boldsymbol{V} \sim 1$, $\boldsymbol{V}^{-1} \sim 1$ as well as $\|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1$. As above it is enough to show $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\|_2 \lesssim 1$. By Lemma 6.3.6 (i) and (6.3.39), $1 - \|\boldsymbol{TF}\|_2 \sim 1$ which again leads to $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\|_2 \lesssim 1$. We conclude (6.3.21) for $\eta \leq \eta_*$ and $\tau \in [1+\tau_*, \tau^*]$.

Next, we verify (6.3.22) in these two regimes. Using $\boldsymbol{h} \cdot \mathbf{1}(\|\boldsymbol{h}\|_\infty \leq \lambda_*) = \mathcal{O}(\|\boldsymbol{d}\|_\infty)$ by (6.3.21), $\boldsymbol{v} \lesssim 1$ and $\boldsymbol{u} \lesssim 1$, we see that with the exception of $\boldsymbol{d}$, all terms in (6.3.50) are second order in $\boldsymbol{d}$. Therefore,

$$\boldsymbol{r} \cdot \mathbf{1}(\|\boldsymbol{h}\|_\infty \leq \lambda_*) = \boldsymbol{d} \cdot \mathbf{1}(\|\boldsymbol{h}\|_\infty \leq \lambda_*) + \mathcal{O}\left(\|\boldsymbol{d}\|_\infty^2\right) \tag{6.3.51}$$

uniformly for $\eta \geq \eta_*$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*]$ and $\tau \in [1+\tau_*, \tau^*]$.

We start from $\boldsymbol{Lh} = \boldsymbol{r}$ and compute

$$\langle \boldsymbol{y}, \boldsymbol{h} \rangle = \langle (\boldsymbol{L}^{-1})^* \boldsymbol{y}, \boldsymbol{r} \rangle = \langle (\boldsymbol{L}^{-1})^* \boldsymbol{y}, \boldsymbol{d} \rangle + \langle (\boldsymbol{L}^{-1})^* \boldsymbol{y}, \boldsymbol{r} - \boldsymbol{d} \rangle = \langle \boldsymbol{Ry}, \boldsymbol{d} \rangle + \langle (\boldsymbol{L}^{-1})^* \boldsymbol{y}, \boldsymbol{r} - \boldsymbol{d} \rangle. \tag{6.3.52}$$

Here, we defined the operator $\boldsymbol{R} = \boldsymbol{R}(\eta)$ on $\mathbb{C}^{2n}$ in the last step through its action on any $\boldsymbol{x} \in \mathbb{C}^{2n}$ via

$$\boldsymbol{Rx} := \left(\boldsymbol{L}^{-1}\right)^* \boldsymbol{x} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{FT})^{-1} \boldsymbol{V} \boldsymbol{x}. \tag{6.3.53}$$

Now, we establish that $\|(\boldsymbol{L}^{-1})^*\|_\infty \lesssim 1$ in the two regimes considered in Step 2. From this, we conclude that $\|\boldsymbol{R}\|_\infty \lesssim 1$ and that the last term in (6.3.52) when multiplied by $\mathbf{1}(\|\boldsymbol{h}\|_\infty \leq \lambda_*)$ is bounded by $\lesssim \|\boldsymbol{y}\|_\infty \|\boldsymbol{d}\|_\infty^2$ because of (6.3.51). By Lemma 6.3.6 (i), (6.3.35), (6.3.36) and (6.3.39) we have $1 - \|\boldsymbol{FT}\|_2 \sim 1$. Thus, $\|(\mathbb{1} - \boldsymbol{FT})^{-1}\|_2 \lesssim 1$ and hence $\|(\mathbb{1} - \boldsymbol{FT})^{-1}\|_\infty \lesssim 1$ by Lemma 6.7.1 (ii). As $\|\boldsymbol{V}\|_\infty \|\boldsymbol{V}^{-1}\|_\infty \lesssim 1$ we get $\|(\boldsymbol{L}^{-1})^*\|_\infty \lesssim 1$. Therefore, we conclude that (6.3.22) holds true uniformly for $\eta \geq \eta_*$ and

$\tau \in [0, 1 - \tau_*] \cup [1 + \tau_*, \tau^*]$ as well as for $\eta \in (0, \eta_*]$ and $\tau \in [1 + \tau_*, \tau^*]$. Thus, we have proven the proposition for these combinations of η and τ.

Finally, we prove the proposition in the most interesting regime, $\tau \in [0, 1 - \tau_*]$ and for small η:

Step 3: There exists $\eta_* > 0$, depending only on $\mathcal{P}$, and $\lambda_* \gtrsim 1$ such that (6.3.21) holds true for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1 - \tau_*]$. Moreover, with this choice of λ_*, (6.3.22) holds true for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1 - \tau_*]$.

The crucial step for proving (6.3.21) and (6.3.22) was the order one bound on $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\|_2$. However, in the regime $\tau \in [0, 1 - \tau_*]$ and small η such bound is not available since $(\mathbb{1} - \boldsymbol{TF})\boldsymbol{f}_- = \mathcal{O}(\eta)$ which can be deduced from (6.3.62) below. The simple bound

$$\|(\mathbb{1} - \boldsymbol{TF})^{-1}\|_2 \lesssim \eta^{-1} \tag{6.3.54}$$

which is a consequence of (6.3.35) and $\|\boldsymbol{T}\|_2 = 1$ is not strong enough. In order to control $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\boldsymbol{Vr}\|_2$ we will need to use a special property of the vector $\boldsymbol{Vr}$, namely that it is almost orthogonal to $\boldsymbol{f}_-$. This mechanism is formulated in the following *Contraction-Inversion Lemma* which is proven in Section 6.7 below. It is closely related to the Rotation-Inversion lemmas – Lemma 5.8 in [5] and Lemma 4.3.6 in Chapter 4 – which control the invertibility of $\mathbb{1} - UF$, where U is a unitary operator and F is symmetric.

Lemma 6.3.7 (Contraction-Inversion Lemma)**.** *Let $\varepsilon, \eta, c_1, c_2, c_3 > 0$ satisfying $\eta \leq \varepsilon c_1/(2c_2^2)$ and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{2n\times 2n}$ be two Hermitian matrices such that*

$$\|\boldsymbol{A}\|_2 \leq 1, \quad \|\boldsymbol{B}\|_2 \leq 1 - c_1\eta. \tag{6.3.55}$$

Suppose that there are ℓ^2-normalized vectors $\boldsymbol{b}_\pm \in \mathbb{C}^{2n}$ satisfying

$$\boldsymbol{B}\boldsymbol{b}_+ = \|\boldsymbol{B}\|_2\boldsymbol{b}_+, \quad \boldsymbol{B}\boldsymbol{b}_- = -\|\boldsymbol{B}\|_2\boldsymbol{b}_-, \quad \|\boldsymbol{B}\boldsymbol{x}\|_2 \leq (1-\varepsilon)\|\boldsymbol{x}\|_2 \tag{6.3.56}$$

for all $\boldsymbol{x} \in \mathbb{C}^{2n}$ such that $\boldsymbol{x} \perp \operatorname{span}\{\boldsymbol{b}_+, \boldsymbol{b}_-\}$.

Furthermore, assume that

$$\langle \boldsymbol{b}_+ , \boldsymbol{A}\boldsymbol{b}_+\rangle \leq 1 - \varepsilon, \quad \|(\mathbb{1} + \boldsymbol{A})\boldsymbol{b}_-\|_2 \leq c_2\eta. \tag{6.3.57}$$

Then there is a constant $C > 0$, depending only on c_1, c_2, c_3 and ε, such that for each $\boldsymbol{p} \in \mathbb{C}^{2n}$ satisfying

$$|\langle \boldsymbol{b}_-, \boldsymbol{p}\rangle| \leq c_3\eta\|\boldsymbol{p}\|_2, \tag{6.3.58}$$

it holds true that

$$\|(\mathbb{1} - \boldsymbol{AB})^{-1}\boldsymbol{p}\|_2 \leq C\|\boldsymbol{p}\|_2. \tag{6.3.59}$$

We will apply this lemma with the choices $\boldsymbol{A} = \boldsymbol{T}$, $\boldsymbol{B} = \boldsymbol{F}$, $\boldsymbol{b}_\pm = \boldsymbol{f}_\pm$ and $\boldsymbol{p} = \boldsymbol{Vr}$. The resulting bound on $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\boldsymbol{Vr}\|_2$ will be lifted to a bound on $\|(\mathbb{1} - \boldsymbol{TF})^{-1}\boldsymbol{Vr}\|_\infty$ by (6.7.9). All estimates in the remainder of this proof will hold true uniformly for $\tau \in [0, 1 - \tau_*]$. However, we will not stress this fact for each estimate. Moreover, the estimates will be uniform for $\eta \in (0, \eta_*]$. The threshold $\eta_* \leq 1$ will be chosen later such that it depends on $\mathcal{P}$ only and the assumptions of Lemma 6.3.7 are fulfilled. We now start checking the assumptions of Lemma 6.3.7.

By Proposition 6.3.2, there is $\Phi_1 \sim 1$ such that

$$\Phi_1^{-1} \leq \boldsymbol{v} \leq \Phi_1 \tag{6.3.60}$$

for all $\eta \in (0, 1]$. We recall from (6.3.35) that there is a constant $c_1 \sim 1$ such that $\|\boldsymbol{F}\|_2 \leq 1 - c_1\eta$ for all $\eta \in (0, 1]$. Recalling the definition of $\boldsymbol{a}$ from (6.3.43), we conclude from (6.3.46) the existence of $\Phi_2 \sim 1$ such that

$$\|\boldsymbol{f}_- - \boldsymbol{a}\|_2 \leq \|\boldsymbol{f}_- - \boldsymbol{a}\|_\infty \leq \Phi_2\eta \tag{6.3.61}$$

for all $\eta \in (0, 1]$. Here, we used that $\|\boldsymbol{y}\|_2 \leq \|\boldsymbol{y}\|_\infty$ for all $\boldsymbol{y} \in \mathbb{C}^{2n}$ due to the normalization of the ℓ^2 norm.

Since the first and the second n-component of the vector $\boldsymbol{Vv}$ are the same we have $\boldsymbol{Ta} = -\boldsymbol{a}$ by (6.3.43) and Lemma 6.3.6 (iii). Hence,

$$\|\boldsymbol{f}_- + \boldsymbol{Tf}_-\|_2 \leq \|\boldsymbol{f}_- - \boldsymbol{a}\|_2 + \|\boldsymbol{T}\|_2\|\boldsymbol{f}_- - \boldsymbol{a}\|_2 \leq 2\Phi_2\eta \tag{6.3.62}$$

by $\|\boldsymbol{T}\|_2 = 1$ and (6.3.61).

Due to (6.3.38), there exists $\varepsilon \sim 1$ such that

$$\|\boldsymbol{Fx}\|_2 \leq (1 - \varepsilon)\|\boldsymbol{x}\|_2$$

for all $\boldsymbol{x} \in \mathbb{C}^{2n}$ such that $\boldsymbol{x} \perp \boldsymbol{f}_+$ and $\boldsymbol{x} \perp \boldsymbol{f}_-$ and for all $\eta \in (0,1]$. As $\boldsymbol{T}$ is Hermitian we can also assume by (6.3.48) that

$$\langle \boldsymbol{f}_+, \boldsymbol{T}\boldsymbol{f}_+ \rangle \leq 1 - \varepsilon$$

for all $\eta \in (0,1]$ by possibly reducing ε but keeping $\varepsilon \gtrsim 1$.

So far we checked the conditions (6.3.55)–(6.3.57), it remains to verify (6.3.58) with the choice $\boldsymbol{p} = \boldsymbol{V}\boldsymbol{r}$. Assuming that $\langle \boldsymbol{a}, \boldsymbol{V}\boldsymbol{r} \rangle = 0$, we deduce from (6.3.61) that

$$|\langle \boldsymbol{f}_-, \boldsymbol{V}\boldsymbol{r} \rangle| \leq |\langle \boldsymbol{a}, \boldsymbol{V}\boldsymbol{r} \rangle| + \|\boldsymbol{f}_- - \boldsymbol{a}\|_2 \|\boldsymbol{V}\boldsymbol{r}\|_2 \leq \Phi_2 \eta \|\boldsymbol{V}\boldsymbol{r}\|_2. \tag{6.3.63}$$

This is the estimate required in (6.3.58). Hence, it suffices to show that $\boldsymbol{V}\boldsymbol{r}$ is perpendicular to $\boldsymbol{a}$, i.e.,

$$\langle \boldsymbol{e}_-(\boldsymbol{V}\boldsymbol{v}), \boldsymbol{V}\boldsymbol{r} \rangle = \left\langle \boldsymbol{e}_- \left(\boldsymbol{V}^2 \boldsymbol{v} \right), \boldsymbol{L}\boldsymbol{h} \right\rangle = \left\langle \boldsymbol{L}^* \left(\boldsymbol{e}_- \frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} \right), \boldsymbol{h} \right\rangle = 0, \tag{6.3.64}$$

where we used the symmetry of $\boldsymbol{V}$, that $\boldsymbol{V}$ is diagonal and (6.3.23) in the first equality, and the notation $\widetilde{\boldsymbol{v}} = (v_2, v_1)$.

We compute

$$\begin{aligned} \boldsymbol{L}^* \left(\boldsymbol{e}_- \frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} \right) &= \boldsymbol{e}_- \frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} + \boldsymbol{S}_o \left(\boldsymbol{v}^2 \boldsymbol{e}_- \frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} \right) - \tau \boldsymbol{S}_d^t \left(\boldsymbol{u}^2 \boldsymbol{e}_- \frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}} \right) \\ &= \begin{pmatrix} \eta + S v_2 - S \left(v_2 \left(\frac{v_1 v_2}{u} + \tau u \right) \right) \\ -\eta - S^t v_1 + S^t \left(v_1 \left(\frac{v_1 v_2}{u} + \tau u \right) \right) \end{pmatrix} = \eta \boldsymbol{e}_-. \end{aligned} \tag{6.3.65}$$

Here, we used (6.3.31) in the second step and the n-component relations of the second identity in (6.3.32) in the last step. Since $\langle \boldsymbol{e}_- \boldsymbol{g} \rangle = \langle \boldsymbol{e}_- \boldsymbol{v} \rangle = 0$ by (6.3.20) and (6.3.8), respectively, this proves (6.3.64) and therefore (6.3.63) as well. Thus, we checked all conditions of Lemma 6.3.7.

By possibly reducing η_* but keeping $\eta_* \gtrsim 1$, we can assume that $\eta_* \leq \varepsilon c_1/(8\Phi_2^2)$. Now, we can apply Lemma 6.3.7 with ε, c_1, $c_2 = 2\Phi_2$, $c_3 = \Phi_2$ for any $\eta \in (0, \eta_*]$. Thus, applying (6.3.59) in Lemma 6.3.7 to (6.3.29), we obtain $\|\boldsymbol{V}\boldsymbol{h}\|_2 \lesssim \|\boldsymbol{V}\boldsymbol{r}\|_2$ and hence $\|\boldsymbol{V}\boldsymbol{h}\|_\infty \lesssim \|\boldsymbol{V}\boldsymbol{r}\|_\infty$ because of (6.7.9). Therefore, for any $\lambda_* > 0$, depending only on $\mathcal{P}$, we have

$$\|\boldsymbol{h}\|_\infty \mathbf{1}(\|\boldsymbol{h}\|_\infty \le \lambda_*) \lesssim \|\boldsymbol{V}^{-1}\|_\infty \|\boldsymbol{V}\boldsymbol{r}\|_\infty \mathbf{1}(\|\boldsymbol{h}\|_\infty \le \lambda_*) \lesssim \|\boldsymbol{h}\|_\infty^2 + \|\boldsymbol{d}\|_\infty$$

uniformly for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1-\tau_*]$. Here, we used (6.3.30) and (6.3.49) in the second step. Choosing $\lambda_* > 0$ small enough as before, we conclude (6.3.21) for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1-\tau_*]$. Since $\eta_* > 0$ depends only on $\mathcal{P}$, and η_* was arbitrary in the proof of Step 2 we proved (6.3.21) for all $\eta > 0$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$.

In order to prove (6.3.22), we remark that because of (6.3.21) and (6.3.50) the estimate (6.3.51) holds true for $\eta \in (0, \eta_*]$ and $\tau \in [0, 1-\tau_*]$ as well. Due to the instability (6.3.54) of $(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}$ and, correspondingly, of its adjoint, the definition of $\boldsymbol{R}$ in (6.3.53) will not yield an operator satisfying $\|\boldsymbol{R}\|_\infty \lesssim 1$ in this regime. Therefore, we again employ that the inverse of $\mathbb{1} - \boldsymbol{T}\boldsymbol{F}$ is bounded on the subspace orthogonal to $\boldsymbol{f}_-$ and the blow-up in the direction of $\boldsymbol{f}_-$ is compensated by the smallness of $\langle \boldsymbol{f}_-, \boldsymbol{V}\boldsymbol{r}\rangle$ following from $\langle \boldsymbol{a}, \boldsymbol{V}\boldsymbol{r}\rangle = 0$ and $\|\boldsymbol{f}_- - \boldsymbol{a}\|_\infty = \mathcal{O}(\eta)$ by (6.3.46).

Let $\boldsymbol{Q}$ be the orthogonal projection onto the subspace $\boldsymbol{f}_-^\perp$, i.e., $\boldsymbol{Q}\boldsymbol{x} \coloneqq \boldsymbol{x} - \langle \boldsymbol{f}_-, \boldsymbol{x}\rangle \boldsymbol{f}_-$ for all $\boldsymbol{x} \in \mathbb{C}^{2n}$. Recalling the definition of $\boldsymbol{a}$ in (6.3.43), we now define the operator $\boldsymbol{R} = \boldsymbol{R}(\eta)$ on $\mathbb{C}^{2n}$ as follows:

$$\boldsymbol{R}\boldsymbol{x} \coloneqq \boldsymbol{V}\left((\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q}\right)^* \boldsymbol{V}^{-1}\boldsymbol{x} - \langle \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{f}_-, \boldsymbol{x}\rangle \boldsymbol{V}(\boldsymbol{f}_- - \boldsymbol{a}) \tag{6.3.66}$$

for every $\boldsymbol{x} \in \mathbb{C}^{2n}$. Note that this $\boldsymbol{R}$ is different from the one given in (6.3.53) that is used in the other parameter regimes. Now, we estimate $\|\boldsymbol{R}\boldsymbol{x}\|_\infty$. For the first term, we use the bound (6.7.11) whose assumptions we check first. The first condition, $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q}\|_2 \lesssim 1$, in (6.7.10) follows from (6.3.59) as (6.3.58) with $\boldsymbol{p} = \boldsymbol{Q}\boldsymbol{x}$ is trivially satisfied and hence $\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q}\boldsymbol{x}\|_2 \lesssim \|\boldsymbol{Q}\boldsymbol{x}\|_2 \lesssim \|\boldsymbol{x}\|_2$. The second condition in (6.7.10) is met by (6.3.35) and the third condition is exactly (6.3.62). Using $\|\boldsymbol{f}_-\|_\infty \lesssim 1$ from (6.3.37), (6.7.11) and (6.3.30), we conclude that the first term in (6.3.66) is $\lesssim \|\boldsymbol{x}\|_\infty$. In the second term, we use the trivial bound

$$\left\|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\right\|_\infty \lesssim \eta^{-1} \tag{6.3.67}$$

which is a consequence of the corresponding bound on $\|(1 - \boldsymbol{T}\boldsymbol{F})^{-1}\|_2$ in (6.3.54) and (6.7.9).

The potential blow-up in (6.3.67) for small η is compensated by the estimate $\|\boldsymbol{f}_- - \boldsymbol{a}\|_\infty = \mathcal{O}(\eta)$ from (6.3.46). Altogether this yields $\|\boldsymbol{R}(\eta)\|_\infty \lesssim 1$ for all $\eta \in (0, \eta_*]$.

From the definition of $\boldsymbol{R}$, we obtain

$$\begin{aligned}\langle \boldsymbol{y}, \boldsymbol{h}\rangle &= \langle \boldsymbol{y}, \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{V}\boldsymbol{r}\rangle \\ &= \Big\langle \boldsymbol{V}^{-1}\boldsymbol{y}\,,\,(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q}\boldsymbol{V}(\boldsymbol{r} - \boldsymbol{d})\Big\rangle \\ &\quad + \Big\langle \boldsymbol{y}\,,\,\boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{f}_-\Big\rangle \Big\langle \boldsymbol{f}_- - \boldsymbol{a}\,,\,\boldsymbol{V}(\boldsymbol{r} - \boldsymbol{d})\Big\rangle + \langle \boldsymbol{R}\boldsymbol{y}, \boldsymbol{d}\rangle.\end{aligned} \tag{6.3.68}$$

Notice that we first inserted $\mathbb{1} = \boldsymbol{Q} + |\boldsymbol{f}_-\rangle\langle\boldsymbol{f}_-|$ before $\boldsymbol{V}\boldsymbol{r}$, then we inserted the vector $\boldsymbol{a}$ in the second term for free by using $\langle \boldsymbol{a}, \boldsymbol{V}\boldsymbol{r}\rangle = 0$ from (6.3.64). This brought in the factor $\boldsymbol{f}_- - \boldsymbol{a} \sim \mathcal{O}(\eta)$ that compensates the $(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}$ on the unstable subspace parallel to $\boldsymbol{f}_-$. Finally, we subtracted the term $\boldsymbol{d}$ to $\boldsymbol{r}$ freely and we defined the operator $\boldsymbol{R}$ exactly to compensate for it. The reason for this counter term $\boldsymbol{d}$ is the formula (6.3.51) showing that $\boldsymbol{r} - \boldsymbol{d}$ is one order better in $\boldsymbol{d}$ than $\boldsymbol{r}$. Thus, the first two terms in the right-hand side of (6.3.68) are bounded by $\|\boldsymbol{d}\|_\infty^2 \|\boldsymbol{y}\|_\infty$. The compensating term, $\langle \boldsymbol{R}\boldsymbol{y}, \boldsymbol{d}\rangle$ remains first order in $\boldsymbol{d}$ but only in weak sense, tested against the vector $\boldsymbol{R}\boldsymbol{y}$, and not in norm sense. This is the essential improvement of (6.3.22) over (6.3.21). Recalling now $\boldsymbol{h} = \boldsymbol{g} - \mathrm{i}\boldsymbol{v}$, the identity (6.3.68) together with the bounds we just explained concludes the proof of Proposition 6.3.3. □

6.4. Proof of Proposition 6.2.5

As in the previous section, we assume without loss of generality that $\rho(S) = 1$. See the remark about (6.3.7).

For $\tau_* > 0$ and $\tau^* > \tau_* + 1$, we define

$$\mathbb{D}_< := \{z \in \mathbb{C} \;|\; |z|^2 \le 1 - \tau_*\}, \quad \mathbb{D}_> := \{z \in \mathbb{C} \;|\; 1 + \tau_* \le |z|^2 \le \tau^*\}. \tag{6.4.1}$$

Via $\tau = |z|^2$, the sets $\mathbb{D}_<$ and $\mathbb{D}_>$ correspond to the regimes $[0, 1 - \tau_*]$ and $[1 + \tau_*, \tau^*]$, respectively, which are used in the previous section.

Proof of Proposition 6.2.5. Since the defining equations in (6.2.4) are smooth functions of η, τ and $(\boldsymbol{v}_i)_{i=1,\ldots,2n}$ and the operator $\boldsymbol{L}$ is invertible for $\eta > 0$ the implicit function theorem implies that the function $\boldsymbol{v}\colon \mathbb{R}_+ \times \mathbb{R}_0^+ \to \mathbb{R}_+^{2n}$ is smooth. Therefore, the function $\mathbb{R}_+ \times \mathbb{C} \to \mathbb{R}_+^{2n}$, $(\eta, z) \mapsto \boldsymbol{v}^\tau(\eta)|_{\tau=|z|^2}$ is also smooth.

For $\alpha = (\alpha_1, \alpha_2) \in \mathbb{N}^2$, we define

$$\partial^\alpha \boldsymbol{v} := \partial_\eta^{\alpha_1} \partial_\tau^{\alpha_2} \boldsymbol{v}.$$

For fixed $\tau_* > 0$ and $\tau^* > \tau_* + 1$, we first prove that for all $\alpha \in \mathbb{N}^2$, we have

$$\|\partial^\alpha \boldsymbol{v}\|_\infty \lesssim 1 \tag{6.4.2}$$

uniformly for all $\eta > 0$ and $\tau \in [0, 1-\tau_*] \cup [1+\tau_*, \tau^*]$.

Differentiating (6.2.4) with respect to η and τ, respectively, yields

$$\boldsymbol{L}(\partial_\eta \boldsymbol{v}) = -\boldsymbol{v}^2 + \tau \boldsymbol{u}^2, \quad \boldsymbol{L}(\partial_\tau \boldsymbol{v}) = -\boldsymbol{u}\boldsymbol{v}. \tag{6.4.3}$$

By further differentiating with respect to η and τ, we iteratively obtain that for any multi-index $\alpha \in \mathbb{N}^2$

$$\boldsymbol{L}\partial^\alpha \boldsymbol{v} = \boldsymbol{r}_\alpha, \tag{6.4.4}$$

where $\boldsymbol{r}_\alpha$ only depends on η, τ and $\partial^\beta \boldsymbol{v}$ for $\beta \in \mathbb{N}^2$, $|\beta| = \beta_1 + \beta_2 < |\alpha|$. In fact, for all $\alpha \in \mathbb{N}^2$, we have

$$\boldsymbol{L}(\partial^{\alpha+e_1}\boldsymbol{v}) = \partial^\alpha \left(-\boldsymbol{v}^2 + \tau \boldsymbol{u}^2\right) - \sum_{\nu \le \alpha, \nu \neq (0,0)} \begin{pmatrix} \alpha \\ \nu \end{pmatrix} (\partial^\nu \boldsymbol{L}) \left(\partial^{\alpha-\nu+e_1}\boldsymbol{v}\right), \tag{6.4.5a}$$

$$\boldsymbol{L}(\partial^{\alpha+e_2}\boldsymbol{v}) = \partial^\alpha \left(-\boldsymbol{v}\boldsymbol{u}\right) - \sum_{\nu \le \alpha, \nu \neq (0,0)} \begin{pmatrix} \alpha \\ \nu \end{pmatrix} (\partial^\nu \boldsymbol{L}) \left(\partial^{\alpha-\nu+e_2}\boldsymbol{v}\right). \tag{6.4.5b}$$

As an example, we compute

$$\begin{aligned} \boldsymbol{L}\partial_\tau^2 \boldsymbol{v} &= -2\boldsymbol{u}\partial_\tau \boldsymbol{v} + 2\boldsymbol{u}^2 \boldsymbol{S}_d \partial_\tau \boldsymbol{v} - 2\boldsymbol{v}\partial_\tau \boldsymbol{v} \boldsymbol{S}_o \partial_\tau \boldsymbol{v} + \frac{2\tau \boldsymbol{u}^2}{\boldsymbol{v}} \partial_\tau \boldsymbol{v} \boldsymbol{S}_d \partial_\tau \boldsymbol{v} - \frac{2\tau \boldsymbol{u}^3}{\boldsymbol{v}} \left(\boldsymbol{S}_d \boldsymbol{v}\right)^2 \\ &= \frac{2}{\boldsymbol{v}} \left(\partial_\tau \boldsymbol{v}\right)^2 + 2\boldsymbol{u}^2 \boldsymbol{S}_d \partial_\tau \boldsymbol{v} - \frac{2\tau \boldsymbol{u}^3}{\boldsymbol{v}} \left(\boldsymbol{S}_d \partial_\tau \boldsymbol{v}\right)^2, \end{aligned} \tag{6.4.6}$$

where we used the second relation in (6.4.3) in the second step.

By induction on $|\alpha| = \alpha_1 + \alpha_2$, we prove $\|\boldsymbol{r}_\alpha\|_\infty \lesssim 1$ and $\|\partial^\alpha \boldsymbol{v}\|_\infty \lesssim 1$ simultaneously. From (6.4.5), we conclude that $\boldsymbol{r}_{\alpha+e_1}$ and $\boldsymbol{r}_{\alpha+e_2}$ are bounded in ℓ^∞-norm if $\|\partial^\nu \boldsymbol{v}\|_\infty \lesssim 1$

for all $\nu \le \alpha$ as the first term on the right-hand side of (6.4.5a) and (6.4.5b), respectively, and $\partial^\nu \boldsymbol{L}$ for all $\nu \le \alpha$ are bounded. In order to conclude that $\partial^{\alpha+e_1}\boldsymbol{v}$ and $\partial^{\alpha+e_2}\boldsymbol{v}$ are bounded it suffices to prove that $\|\partial^\alpha \boldsymbol{v}\|_\infty \lesssim \|\boldsymbol{r}_\alpha\|_\infty$ by controlling $\boldsymbol{L}^{-1}$ in (6.4.4).

As in the proof of Proposition 6.3.3 the norm of $\boldsymbol{L}^{-1}$ is bounded, $\|\boldsymbol{L}^{-1}\|_\infty \lesssim 1$, for $\tau \in [1+\tau_*, \tau^*]$ or $\tau \in [0, 1-\tau_*]$ and large η as well as $\tau \in [0, 1-\tau_*]$ and small η separately. We thus focus on the most interesting regime where $\tau \in [0, 1-\tau_*]$ and small η. As for the proof of Proposition 6.3.3 we apply Lemma 6.3.7 in this regime. We only check the condition (6.3.58) here since the others are established in the same way as in the proof of Proposition 6.3.3. Recall the definition of $\boldsymbol{a}$ in (6.3.43). Using $\langle \boldsymbol{e}_- \partial^\alpha \boldsymbol{v}\rangle = 0$ from (6.3.8) for all $\alpha \in \mathbb{N}^2$, we obtain

$$\langle \boldsymbol{a}\, ,\, \boldsymbol{V}\boldsymbol{r}_\alpha\rangle = \left\langle \boldsymbol{L}^*(\boldsymbol{e}_-\boldsymbol{V}^2\boldsymbol{v})\, ,\, \partial^\alpha \boldsymbol{v}\right\rangle = \langle \eta \boldsymbol{e}_-\, , \partial^\alpha \boldsymbol{v}\rangle = 0$$

for all $\alpha \in \mathbb{N}^2$. Here, we used $\boldsymbol{L}^*(\boldsymbol{e}_-\boldsymbol{V}^2\boldsymbol{v}) = \eta\boldsymbol{e}_-$ which is shown in (6.3.65) in the proof of Proposition 6.3.3. This concludes the proof of (6.4.2).

Next, we show the integrability of $\Delta_z \langle v_1^\tau|_{\tau=|z|^2}\rangle$ as a function of η for $z \in \mathbb{D}_<$ for fixed $\tau_* > 0$. Note that $\langle v_1^\tau\rangle = \langle \boldsymbol{v}^\tau\rangle$ by (6.3.8). Using

$$\Delta_z \left(\boldsymbol{v}^\tau|_{\tau=|z|^2}\right) = 4\left(\tau\partial_\tau^2 \boldsymbol{v}^\tau + \partial_\tau \boldsymbol{v}^\tau\right)|_{\tau=|z|^2}$$

together with (6.4.3) and (6.4.6), we obtain

$$\boldsymbol{L}\Delta_z \left(\boldsymbol{v}^\tau|_{\tau=|z|^2}\right) = 4\left(\frac{2\tau}{\boldsymbol{v}}\,(\partial_\tau \boldsymbol{v})^2 + 2\tau\boldsymbol{u}^2\boldsymbol{S}_d\partial_\tau\boldsymbol{v} - \frac{2\tau^2\boldsymbol{u}^3}{\boldsymbol{v}}\,(\boldsymbol{S}_d\partial_\tau\boldsymbol{v})^2 - \boldsymbol{u}\boldsymbol{v}\right). \qquad (6.4.7)$$

From (6.3.9), (6.3.10) and (6.3.26), we conclude that $\boldsymbol{u}\boldsymbol{v} \sim (1+\eta^3)^{-1}$ and hence $|\partial_\tau \boldsymbol{v}| \lesssim (1+\eta^3)^{-1}$ uniformly for $z \in \mathbb{D}_<$ since $\|\partial^\alpha \boldsymbol{v}\|_\infty \lesssim \|\boldsymbol{r}_\alpha\|_\infty$. Therefore, the right-hand side of (6.4.7) is of order $(1+\eta^3)^{-1}$ for $z \in \mathbb{D}_<$ and hence using the control on $\boldsymbol{L}^{-1}$ as before, we conclude that $|\Delta_z \left(\boldsymbol{v}^\tau|_{\tau=|z|^2}\right)| \lesssim (1+\eta^3)^{-1}$ uniformly for $\eta > 0$. Thus, $\Delta_z\langle v_1^\tau|_{\tau=|z|^2}\rangle = \Delta_z\langle \boldsymbol{v}^\tau|_{\tau=|z|^2}\rangle$ as a function of η is integrable on $\mathbb{R}_+$ and the integral is a continuous function of $z \in \mathbb{D}_<$. As $\tau_* > 0$ was arbitrary, this concludes the proof of part (i) of Proposition 6.2.5 and shows that σ is a rotationally invariant function on $\mathbb{C}$ which is continuous on $D(0,1)$.

Now, we establish that for $\tau < 1$, the derivative of the average of $\boldsymbol{u}$ with respect to τ gives an alternative representation of the density of states as follows

$$\sigma(z) = \frac{1}{\pi}\partial_\tau \left(\tau\langle \boldsymbol{u}_0\rangle\right)|_{\tau=|z|^2} = -\frac{2}{\pi}\langle \boldsymbol{S}_o\boldsymbol{v}_0\,, \partial_\tau \boldsymbol{v}_0\rangle|_{\tau=|z|^2}, \tag{6.4.8}$$

where $\boldsymbol{u}_0 := \lim_{\eta\downarrow 0}\boldsymbol{u}(\eta)$ and $\boldsymbol{v}_0 := \lim_{\eta\downarrow 0}\boldsymbol{v}(\eta)$. The first relation in (6.4.8) will be proven below and the second one follows immediately using $\tau\boldsymbol{u}_0 = 1 - \boldsymbol{v}_0\boldsymbol{S}_o\boldsymbol{v}_0$ by (6.3.6) and (6.3.25) for $\eta \downarrow 0$, as well as $\boldsymbol{S}_o^t = \boldsymbol{S}_o$.

We first give a heuristic derivation of the first equality in (6.4.8) (see for example Section 4.6 of [40]). Writing the resolvent $\boldsymbol{G}^z$ of $\boldsymbol{H}^z$ as

$$\boldsymbol{G}^z = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$$

with blocks G_{11}, G_{12}, G_{21} and G_{22} of size $n \times n$, we obtain

$$\begin{aligned} \operatorname{tr} G_{12} &= \operatorname{tr}\left[\Big((X-z)(X^*-\bar z)+\eta^2\Big)^{-1}(X-z)\right] \\ &= -\partial_{\bar z}\operatorname{tr}\log\Big((X-z)(X^*-\bar z)+\eta^2\Big) = -\frac{2}{n}\partial_{\bar z}\log|\det(\boldsymbol{H}^z - \mathrm{i}\eta)| \end{aligned}$$

for the normalized trace of G_{12} (see (6.1.3)). Since $\Delta_z = 4\partial_z\partial_{\bar z}$, taking the ∂_z-derivative of the previous identity, we obtain

$$\frac{1}{2n}\Delta_z \log|\det(\boldsymbol{H}^z - \mathrm{i}\eta)| = -\partial_z \operatorname{tr} G_{12}. \tag{6.4.9}$$

Using (6.2.5), (6.2.14) and $\operatorname{Im} m^z \approx \langle v_1^\tau|_{\tau=|z|^2}\rangle$, the left-hand side of (6.4.9) is approximately $\pi\sigma(z)$ after taking the $\eta \downarrow 0$ limit. On the other hand, $\boldsymbol{G}^z$ converges to $\boldsymbol{M}^z$ for $n \to \infty$. Thus, by (6.3.5) the right-hand side of (6.4.9) can be approximated by $\partial_z\left(z\langle u^\tau|_{\tau=|z|^2}(\eta)\rangle\right)$. Therefore, taking $\eta \downarrow 0$, we conclude

$$\pi\sigma(z) \approx \partial_z z\langle u_0^\tau|_{\tau=|z|^2}\rangle = \left(\partial_\tau\tau\langle u_0^\tau\rangle\right)|_{\tau=|z|^2}.$$

In fact, this approximation holds not only in the $n \to \infty$ limit but it is an identity for any fixed n. This completes the heuristic argument for (6.4.8).

We now turn to the rigorous proof of the first relation in (6.4.8). In fact, for $\tau < 1$, we prove the following integrated version

$$\int_{|z'|^2\leq\tau} \sigma(z')\mathrm{d}^2 z' = \tau\langle \boldsymbol{u}_0^\tau\rangle. \tag{6.4.10}$$

Since σ is a continuous function on $D(0,1)$ differentiating (6.4.10) with respect to τ immediately yields (6.4.8).

In order to justify the existence of the limits of $\boldsymbol{v}$ and $\boldsymbol{u}$ for $\eta \downarrow 0$ and the computations in the proof of (6.4.10), we remark that by (6.4.2), $(\eta, z) \mapsto \boldsymbol{v}^\tau(\eta)|_{\tau=|z|^2}$ can be uniquely extended to a positive C^∞ function on $[0,\infty)\times D(0,1)$. In the following, $\boldsymbol{v}$ and $\boldsymbol{v}_0^\tau :=$ $\boldsymbol{v}^\tau|_{\eta=0}$ denote this function and its restriction to $\{0\}\times[0,1)$, respectively. In particular, the restriction $\boldsymbol{v}_0^\tau|_{\tau=|z|^2}$ is a smooth function on $D(0,1)$ which satisfies

$$\frac{1}{\boldsymbol{v}_0^\tau} = \boldsymbol{S}_o\boldsymbol{v}_0^\tau + \frac{\tau}{\boldsymbol{S}_d\boldsymbol{v}_0^\tau} \tag{6.4.11}$$

with $\tau = |z|^2$. Moreover, derivatives of $\boldsymbol{v}$ in η and τ and limits in η and τ for $\tau < 1$ can be freely interchanged.

For the proof of (6.4.10), we use integration by parts to obtain

$$\int_{|z'|^2\leq\tau} \sigma(z')\mathrm{d}^2 z' = -2\tau\int_0^\infty \partial_\tau\langle\boldsymbol{v}\rangle\mathrm{d}\eta = -\tau\int_0^\infty \partial_\tau\left(\langle\boldsymbol{v}\rangle + \langle\widetilde{\boldsymbol{v}}\rangle\right)\mathrm{d}\eta. \tag{6.4.12}$$

We recall $\widetilde{\boldsymbol{v}} = (v_2, v_1)$ and get

$$\boldsymbol{v} = \frac{\eta + \boldsymbol{S}_d\boldsymbol{v}}{(\eta+\boldsymbol{S}_d\boldsymbol{v})(\eta+\boldsymbol{S}_o\boldsymbol{v})+\tau}, \quad \widetilde{\boldsymbol{v}} = \frac{\eta + \boldsymbol{S}_o\boldsymbol{v}}{(\eta+\boldsymbol{S}_d\boldsymbol{v})(\eta+\boldsymbol{S}_o\boldsymbol{v})+\tau}$$

from (6.3.6). This implies the identity

$$\partial_\eta \log\left((\eta+\boldsymbol{S}_d\boldsymbol{v})(\eta+\boldsymbol{S}_o\boldsymbol{v})+\tau\right) = \boldsymbol{v} + \widetilde{\boldsymbol{v}} + \widetilde{\boldsymbol{v}}\boldsymbol{S}_d\partial_\eta\boldsymbol{v} + \boldsymbol{v}\boldsymbol{S}_o\partial_\eta\boldsymbol{v}.$$

Using

$$\langle\widetilde{\boldsymbol{v}}\boldsymbol{S}_d\partial_\eta\boldsymbol{v}\rangle + \langle\boldsymbol{v}\boldsymbol{S}_o\partial_\eta\boldsymbol{v}\rangle = \langle\boldsymbol{v}\boldsymbol{S}_o\partial_\eta\boldsymbol{v}\rangle + \langle\boldsymbol{v}\boldsymbol{S}_o\partial_\eta\boldsymbol{v}\rangle = \partial_\eta\langle\boldsymbol{v}\boldsymbol{S}_o\boldsymbol{v}\rangle$$

and recalling $\boldsymbol{v}_0 := \lim_{\eta\downarrow 0}\boldsymbol{v}(\eta)$, we find for (6.4.12) the expression

$$\int_0^\infty \partial_\tau\left(\langle\boldsymbol{v}\rangle + \langle\widetilde{\boldsymbol{v}}\rangle\right)\mathrm{d}\eta = -\langle\partial_\tau\log\left((\boldsymbol{S}_d\boldsymbol{v}_0)(\boldsymbol{S}_o\boldsymbol{v}_0)+\tau\right)\rangle + \partial_\tau\langle\boldsymbol{v}_0\boldsymbol{S}_o\boldsymbol{v}_0\rangle. \tag{6.4.13}$$

Hence, due to

$$\langle \partial_\tau \log\left((\boldsymbol{S}_d\boldsymbol{v}_0)(\boldsymbol{S}_o\boldsymbol{v}_0)+\tau\right)\rangle = \langle \boldsymbol{u}\rangle + \langle \widetilde{\boldsymbol{v}}_0 \boldsymbol{S}_d \partial_\tau \boldsymbol{v}_0\rangle + \langle \boldsymbol{v}\boldsymbol{S}_o\partial_\tau \boldsymbol{v}_0\rangle = \langle \boldsymbol{u}\rangle + \partial_\tau\langle \boldsymbol{v}_0\boldsymbol{S}_o\boldsymbol{v}_0\rangle.$$

we obtain (6.4.10) from (6.4.13). The formula (6.4.10) was also obtained in [51] with a different method.

We prove (iii) before (ii). As $\boldsymbol{v}_0$ is infinitely often differentiable in τ and $\tau = |z|^2$, we conclude from (6.4.8) that σ is infinitely often differentiable in z. The following lemma shows (6.2.6) which completes the proof of part (iii).

Lemma 6.4.1 (Positivity and boundedness of σ)**.** *Uniformly for $z \in D(0,1)$, we have*

$$\sigma(z) \sim 1, \tag{6.4.14}$$

where $\sim$ only depends on s_* and s^*.

Proof of Lemma 6.4.1. We will compute the derivative in (6.4.8) and prove the estimate (6.4.14) first for $z \in \mathbb{D}_<$ and arbitrary $\tau_* > 0$ depending only on s_* and s^*. Then we show that there is $\tau_* > 0$ depending only on s_* and s^* such that (6.4.14) holds true for $z \in D(0,1) \setminus \mathbb{D}_<$.

In this proof, we write $\mathscr{D}(y) := \operatorname{diag}(y)$ for $y \in \mathbb{C}^l$ for brevity. Furthermore, we introduce the $2n \times 2n$ matrix

$$\boldsymbol{E} := \begin{pmatrix} \mathbb{1} & \mathbb{1} \\ \mathbb{1} & \mathbb{1} \end{pmatrix}.$$

In the following, $\boldsymbol{v}$ and all related quantities will be evaluated at $\tau = |z|^2$. We start the proof from (6.4.8), recall $\boldsymbol{L} = \boldsymbol{V}^{-1}(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})\boldsymbol{V}$ and use the second relation in (6.4.3) as well as (6.3.31) to obtain

$$\begin{aligned} \sigma(z) &= -\frac{2}{\pi}\langle \boldsymbol{S}_o\boldsymbol{v}_0\,, \partial_\tau \boldsymbol{v}_0\rangle \\ &= \lim_{\eta\downarrow 0}\frac{2}{\pi}\left\langle \boldsymbol{V}^{-1}\frac{\widetilde{\boldsymbol{v}}}{\boldsymbol{u}}\,,\,(\mathbb{1}-\boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{V}(\boldsymbol{v}\boldsymbol{u})\right\rangle \\ &= \lim_{\eta\downarrow 0}\frac{2}{\pi}\left\langle \sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}\,,\,\frac{1}{\sqrt{\boldsymbol{u}}}(\mathbb{1}-\boldsymbol{T}\boldsymbol{F})^{-1}\sqrt{\boldsymbol{u}}\sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}\right\rangle \\ &= \lim_{\eta\downarrow 0}\frac{2}{\pi}\left\langle \sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}\,,\,\Big(\mathbb{1}-\mathscr{D}(\boldsymbol{u}^{-1/2})\boldsymbol{T}\boldsymbol{F}\mathscr{D}(\boldsymbol{u}^{1/2})\Big)^{-1}\sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}\right\rangle. \end{aligned} \tag{6.4.15}$$

Note that the inverses of $\mathbb{1} - \boldsymbol{T}\boldsymbol{F}$ and $\mathbb{1} - \tau\mathscr{D}(\boldsymbol{u}^{-1/2})\boldsymbol{T}\boldsymbol{F}\mathscr{D}(\boldsymbol{u}^{1/2})$ exist by Lemma 6.3.6 and Lemma 6.3.4 as $\eta > 0$ and $\tau < 1$.

Due to (6.3.27a) and (6.3.32), we have $\boldsymbol{T} = -\mathbb{1} + \tau\boldsymbol{u}\boldsymbol{E}$ which implies

$$\begin{aligned}\mathbb{1} - \mathscr{D}(\boldsymbol{u}^{-1/2})\boldsymbol{T}\boldsymbol{F}\mathscr{D}(\boldsymbol{u}^{1/2}) \\ &= \mathbb{1} + \mathscr{D}(\boldsymbol{u}^{-1/2})\boldsymbol{F}\mathscr{D}(\boldsymbol{u}^{1/2}) - \tau\mathscr{D}(\boldsymbol{u}^{1/2})\boldsymbol{E}\boldsymbol{F}\mathscr{D}(\boldsymbol{u}^{1/2}) \\ &= \left(\mathbb{1} - \tau\mathscr{D}(\boldsymbol{u}^{1/2})\boldsymbol{E}\boldsymbol{F}(\mathbb{1}+\boldsymbol{F})^{-1}\mathscr{D}(\boldsymbol{u}^{1/2})\right)\left(\mathbb{1} + \mathscr{D}(\boldsymbol{u}^{-1/2})\boldsymbol{F}\mathscr{D}(\boldsymbol{u}^{1/2})\right).\end{aligned} \tag{6.4.16}$$

From (6.3.33) and (6.3.44), we deduce $\sqrt{\boldsymbol{u}}\boldsymbol{F}\sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}/\boldsymbol{u}} = \sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}} + \mathcal{O}(\eta)$. Hence, due to (6.4.16), (6.4.15) yields

$$\sigma(z) = \lim_{\eta\downarrow 0}\frac{1}{\pi}\left\langle\sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}\,,\,\left(\mathbb{1} - \tau\mathscr{D}(\boldsymbol{u}^{1/2})\boldsymbol{E}\boldsymbol{F}(\mathbb{1}+\boldsymbol{F})^{-1}\mathscr{D}(\boldsymbol{u}^{1/2})\right)^{-1}\sqrt{\boldsymbol{v}\widetilde{\boldsymbol{v}}}\right\rangle. \tag{6.4.17}$$

Defining the matrix $F \in \mathbb{C}^{n\times n}$ through $Fy = \sqrt{v_1 u/v_2}S\sqrt{v_2 u/v_1}\,y$ for $y \in \mathbb{C}^n$, we obtain

$$\boldsymbol{F} = \begin{pmatrix} 0 & F \\ F^t & 0\end{pmatrix}, \qquad (\mathbb{1}+\boldsymbol{F})^{-1} = \begin{pmatrix} (\mathbb{1}-FF^t)^{-1} & -(\mathbb{1}-FF^t)^{-1}F \\ -F^t(\mathbb{1}-FF^t)^{-1} & (\mathbb{1}-F^tF)^{-1}\end{pmatrix}. \tag{6.4.18}$$

Furthermore, we introduce the $n \times n$ matrix A by

$$A := 2\cdot\mathbb{1} + (F^t - \mathbb{1})(\mathbb{1}-FF^t)^{-1} + (F-\mathbb{1})(\mathbb{1}-F^tF)^{-1}.$$

From the computation

$$\boldsymbol{E}\boldsymbol{F}(\mathbb{1}+\boldsymbol{F})^{-1} = \begin{pmatrix} \mathbb{1} + (F^t-\mathbb{1})(\mathbb{1}-FF^t)^{-1} & \mathbb{1} + (F-\mathbb{1})(\mathbb{1}-F^tF)^{-1} \\ \mathbb{1} + (F^t-\mathbb{1})(\mathbb{1}-FF^t)^{-1} & \mathbb{1} + (F-\mathbb{1})(\mathbb{1}-F^tF)^{-1}\end{pmatrix},$$

we conclude that

$$\left(\mathbb{1} - \tau\mathscr{D}(\boldsymbol{u}^{1/2})\boldsymbol{E}\boldsymbol{F}(\mathbb{1}+\boldsymbol{F})^{-1}\mathscr{D}(\boldsymbol{u}^{1/2})\right)^{-1}\begin{pmatrix} x \\ x\end{pmatrix} = \begin{pmatrix} (\mathbb{1}-\tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2}))^{-1}x \\ (\mathbb{1}-\tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2}))^{-1}x\end{pmatrix} \tag{6.4.19}$$

for all $x \in \mathbb{C}^n$. Before applying this relation to (6.4.17), we show that $\mathbb{1}-\tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})$ is invertible for $\tau < 1$. The relations in (6.4.18) yield

$$\langle x\,, Ax\rangle = 2\|x\|_2^2 - 2\left\langle\begin{pmatrix} x \\ x\end{pmatrix}\,,\,(\mathbb{1}+\boldsymbol{F})^{-1}\begin{pmatrix} x \\ x\end{pmatrix}\right\rangle \tag{6.4.20}$$

for all $x \in \mathbb{C}^n$ and $\eta > 0$. In particular, since $\|\boldsymbol{F}\|_2 \leq 1$ by (6.3.33) we conclude $A \leq \mathbb{1}$. Hence, $\tau u = 1 - v_1 v_2/u < 1$ for $\tau < 1$ by (6.3.32) implies that $\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$ is invertible for $\tau < 1$. Thus, we apply (6.4.19) to (6.4.17) and obtain for $z \in D(0,1)$

$$\sigma(z) = \frac{2}{\pi} \lim_{\eta \downarrow 0} \left\langle \sqrt{v_1 v_2}\,, \left(\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})\right)^{-1} \sqrt{v_1 v_2} \right\rangle. \tag{6.4.21}$$

Let $\tau_* > 0$ depend only on s_* and s^*. From (6.3.10) and (6.4.2), we conclude that $|\sigma| \lesssim 1$ uniformly for $z \in \mathbb{D}_<$ because of (6.4.8). This proves the upper bound in (6.4.14) for $z \in \mathbb{D}_<$.

For the proof of the lower bound, we infer some further properties of A and $\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$, respectively, from information about $\boldsymbol{F}$ via (6.4.20). In the following, we use versions of Proposition 6.3.2, (6.3.26) and Lemma 6.3.4 extended to the limiting case $\eta = 0+$. Recalling $\boldsymbol{v}_0 = \lim_{\eta \downarrow 0} \boldsymbol{v}$, these results are a simple consequence of the uniform convergence $\partial^\alpha \boldsymbol{v} \to \partial^\alpha \boldsymbol{v}_0$ for $\eta \downarrow 0$ and all $\alpha \in \mathbb{N}^2$ by (6.4.2).

Since $\boldsymbol{f}_- = (\sqrt{v_1 v_2/u}, -\sqrt{v_1 v_2/u}) + \mathcal{O}(\eta)$ by (6.3.45) there are $\eta_*, \varepsilon \sim 1$ by Lemma 6.3.4 such that $\operatorname{Spec}(\boldsymbol{F}|_W) \subset [-1+\varepsilon, 1]$ on the subspace $W := \{(x,x) | x \in \mathbb{C}^n\} \subset \mathbb{C}^{2n}$ as $\boldsymbol{f}_- \perp W$ uniformly for all $\eta \in [0, \eta_*]$. Therefore, for $\|x\|_2 = 1$, the right-hand side of (6.4.20) is contained in $[2(\varepsilon - 1)/\varepsilon, 1]$. Since $(F^t(\mathbb{1} - FF^t)^{-1})^t = F(\mathbb{1} - F^t F)^{-1}$ the matrix A is real symmetric and hence the spectrum of A is contained in $[2(\varepsilon-1)/\varepsilon, 1]$ for all $\eta \in [0, \eta_*]$ as well.

The real symmetric matrix A has a positive and a negative part, i.e., there are positive matrices A_+ and A_- such that $A = A_+ - A_-$. Hence, we have

$$\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2}) = \mathbb{1} - \tau \mathscr{D}(u^{1/2}) A_+ \mathscr{D}(u^{1/2}) + \tau \mathscr{D}(u^{1/2}) A_- \mathscr{D}(u^{1/2}). \tag{6.4.22}$$

The above statements about (6.4.20) yield $\operatorname{Spec} A_+ \subset [0,1]$ and $\operatorname{Spec} A_- \subset [0, 2(1-\varepsilon)/\varepsilon]$. As $0 \leq u\tau$ we conclude from (6.4.22) that the spectrum of $\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})$ is contained in $(0, 2/\varepsilon]$ for all $\eta \in [0, \eta_*]$. Therefore, using (6.4.21), we obtain

$$\sigma(z) = \frac{2}{\pi} \lim_{\eta \downarrow 0} \left\langle \sqrt{v_1 v_2}\,, \left(\mathbb{1} - \tau \mathscr{D}(u^{1/2}) A \mathscr{D}(u^{1/2})\right)^{-1} \sqrt{v_1 v_2} \right\rangle \geq \frac{\varepsilon}{\pi} \langle \boldsymbol{v}_0 \widetilde{\boldsymbol{v}}_0 \rangle \gtrsim 1$$

uniformly for all $z \in \mathbb{D}_<$. Here, we used (6.3.10) in the last step. This shows (6.4.14) for $z \in \mathbb{D}_<$ for any $\tau_* > 0$ depending only on s_* and s^*.

We now show that there is $\tau_* > 0$ depending only on s_* and s^* such that (6.4.14) holds true for $z \in D(0,1)\backslash\mathbb{D}_<$. This is proven by tracking the blowup of $(\mathbb{1}-\tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2}))^{-1}$ in $1-\tau$ for $\tau \uparrow 1$ in (6.4.21) and establishing a compensation through $v_1 \sim v_2 \sim (1-\tau)^{1/2}$ due to (6.3.10). This yields the upper and lower bound in (6.4.14). Since $\mathbb{1} - \tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})$ in (6.4.21) is also invertible for $\eta = 0$ we may directly set $\eta = 0$ in the following argument.

We multiply the first component of the first relation in (6.3.32) by τ and solve for τu to obtain

$$\tau u = \frac{1}{2}\left(1 + \sqrt{1 - 4\tau v_1 v_2}\right) = 1 - \tau v_1 v_2 + \mathcal{O}\left((1-\tau)^2\right).$$

Therefore, using $v_1 \sim v_2 \sim (1-\tau)^{1/2}$, we have

$$\tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2}) = A - \frac{\tau}{2}\left(\mathscr{D}(v_1v_2)A + A\mathscr{D}(v_1v_2)\right) + \mathcal{O}\left((1-\tau)^2\right).$$

Moreover, from (6.4.20) we conclude that $Aa = a$ for $a := \sqrt{v_1v_2/u}/\|\sqrt{v_1v_2/u}\|_2$. Here, we also used (6.3.44) and (6.3.33) with $\eta = 0$.

Thus, the smallest eigenvalue of the positive operator $\mathbb{1} - \tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})$ satisfies

$$\begin{aligned}\lambda_{\min}\left(\mathbb{1} - \tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})\right) &= \lambda_{\min}\left(\mathbb{1} - A\right) + \tau\langle a^2 v_1 v_2\rangle + \mathcal{O}\left((1-\tau)^2\right) \\ &= \tau\langle a^2 v_1 v_2\rangle + \mathcal{O}\left((1-\tau)^2\right).\end{aligned}$$

Here, we used multiple times that $Aa = a$. Therefore, as A is symmetric we conclude from (6.4.21) that

$$\sigma(z) = \frac{2}{\pi}\left\langle \sqrt{v_1v_2}\,,\,\left(\mathbb{1} - \tau\mathscr{D}(u^{1/2})A\mathscr{D}(u^{1/2})\right)^{-1}\sqrt{v_1v_2}\right\rangle \geq \frac{\langle a\,,\sqrt{v_1v_2}\rangle^2}{\tau\langle a^2v_1v_2\rangle} + \mathcal{O}\left(1-\tau\right).$$

Since $a \sim 1$ and $v_1 \sim v_2 \sim (1-\tau)^{1/2}$ there is $\tau_* \sim 1$ such that the lower bound in (6.4.14) holds true for $z \in D(0,1) \setminus \mathbb{D}_<$. Starting from (6.4.21), we similarly obtain

$$\sigma(z) \leq \frac{\langle v_1v_2\rangle}{\tau\langle a^2v_1v_2\rangle} + \mathcal{O}\left(1-\tau\right).$$

Using the positivity of a, $v_1 \sim v_2 \sim (1-\tau)^{1/2}$ and possibly shrinking $\tau_* \sim 1$ the upper bound in (6.4.14) for $z \in D(0,1) \setminus \mathbb{D}_<$ follows. This concludes the proof of Lemma 6.4.1.

□

As $\sigma(z) = 0$ for $|z| \geq 1$ we conclude from (6.2.6) that σ is nonnegative on $\mathbb{C}$. We use (6.4.10) to compute the total mass of the measure on $\mathbb{C}$ defined by σ. Clearly, $\boldsymbol{u}_0 = \boldsymbol{v}_0/\boldsymbol{S}_d\boldsymbol{v}_0$ and using (6.4.11) and (6.4.10), we obtain

$$\lim_{\tau\uparrow 1} \int_{|z'|^2\leq\tau} \sigma(z')\mathrm{d}^2 z' = 1 - \lim_{\tau\uparrow 1}\langle \boldsymbol{v}_0\,, \boldsymbol{S}_o\boldsymbol{v}_0\rangle = 1.$$

Here, we used that $\lim_{\tau\uparrow 1} \boldsymbol{v}_0 = 0$ by (6.3.10). Hence, as $\sigma(z) = 0$ for $|z| \geq 1$ it defines a probability density on $\mathbb{C}$ which concludes the proof of Proposition 6.2.5. □

Remark 6.4.2 (Jump height)**.** In fact, it is possible to compute the jump height of the density of states σ at the edge $\tau = |z|^2 = 1$. Let s_1 and s_2 be two eigenvectors of S^t and S, respectively, associated to the eigenvalue 1, i.e., $S^t s_1 = s_1$ and $Ss_2 = s_2$. Note that s_1 and s_2 are unique up to multiplication by a scalar.

With this notation, expanding $\boldsymbol{v}^\tau$ for $\tau \leq 1$ around $\tau = 1$ yields

$$\begin{aligned} v_1 &= \sqrt{1-\tau}\left(\frac{\langle s_1 s_2\rangle\langle s_2\rangle}{\langle s_1^2 s_2^2\rangle\langle s_1\rangle}\right)^{1/2} s_1 + \mathcal{O}\left((1-\tau)^{3/2}\right),\\ v_2 &= \sqrt{1-\tau}\left(\frac{\langle s_1 s_2\rangle\langle s_1\rangle}{\langle s_1^2 s_2^2\rangle\langle s_2\rangle}\right)^{1/2} s_2 + \mathcal{O}\left((1-\tau)^{3/2}\right).\end{aligned}$$

Therefore, solving (6.3.32) for τu and expanding in $1-\tau$, we obtain that σ has a jump of height

$$\lim_{|z|^2\uparrow 1} \sigma(z) = \frac{1}{\pi}\lim_{\tau\uparrow 1}\partial_\tau\left(\tau\langle \boldsymbol{u}_0\rangle\right) = \frac{1}{\pi}\frac{\langle s_1 s_2\rangle^2}{\langle s_1^2 s_2^2\rangle}.$$

6.5. Local law

We begin this section with a notion for high probability estimates.

Definition 6.5.1 (Stochastic domination)**.** Let $C\colon \mathbb{R}_+^2 \to \mathbb{R}_+$ be a given function which depends only on a, φ, τ_*, τ^* and the model parameters. If $\Phi = (\Phi^{(n)})_n$ and $\Psi = (\Psi^{(n)})_n$ are two sequences of nonnegative random variables, then we will say that Φ is **stochastically dominated** by Ψ, $\Phi \prec \Psi$, if for all $\varepsilon > 0$ and $D > 0$ we have

$$\mathbb{P}\left(\Phi^{(n)} \geq n^\varepsilon \Psi^{(n)}\right) \leq \frac{C(\varepsilon, D)}{n^D}$$

for all $n \in \mathbb{N}$.

As a trivial consequence of $\mathbb{E}\, x_{ij} = 0$, (6.2.1) and (6.2.2) we remark that

$$|x_{ij}| \prec n^{-1/2}. \tag{6.5.1}$$

6.5.1. Local law for $\boldsymbol{H}^z$. Let (v_1^τ, v_2^τ) be the positive solution of (6.2.4) and u^τ defined as in (6.3.25). In the whole section, we will always evaluate v_1^τ, v_2^τ and u^τ at $\tau = |z|^2$ and mostly suppress the dependence on τ and $|z|^2$, respectively, in our notation. Recall that $\boldsymbol{M}^z$ is defined in (6.3.5). Note that although v_1, v_2 and u are rotationally invariant in $z \in \mathbb{C}$, the dependence of $\boldsymbol{M}^z$ on z is not rotationally symmetric.

For the following theorem, we remark that the sets $\mathbb{D}_<$ and $\mathbb{D}_>$ were introduced in (6.4.1).

Theorem 6.5.2 (Local law for $\boldsymbol{H}^z$)**.** *Let X satisfy (A) and (B) and let $\boldsymbol{G} = \boldsymbol{G}^z$ be the resolvent of $\boldsymbol{H}^z$ as defined in* (6.2.11)*. For fixed $\varepsilon \in (0, 1/2)$, the entrywise local law*

$$\|\boldsymbol{G}^z(\eta) - \boldsymbol{M}^z(\eta)\|_{\max} \prec \begin{cases} \frac{1}{\sqrt{n\eta}} & \text{for } z \in \mathbb{D}_< \,,\ \eta \in [n^{-1+\varepsilon}, 1]\,, \\ \frac{1}{\sqrt{n}} + \frac{1}{n\eta} & \text{for } z \in \mathbb{D}_> \,,\ \eta \in [n^{-1+\varepsilon}, 1]\,, \\ \frac{1}{\sqrt{n}\,\eta^2} & \text{for } z \in \mathbb{D}_< \cup \mathbb{D}_> \,,\ \eta \in [1, \infty)\,, \end{cases} \tag{6.5.2}$$

holds true. In particular,

$$\|\boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\|_\infty \prec \begin{cases} \frac{1}{\sqrt{n\eta}} & \text{for } z \in \mathbb{D}_< \,,\ \eta \in [n^{-1+\varepsilon}, 1]\,, \\ \frac{1}{\sqrt{n}} + \frac{1}{n\eta} & \text{for } z \in \mathbb{D}_> \,,\ \eta \in [n^{-1+\varepsilon}, 1]\,, \\ \frac{1}{\sqrt{n}\,\eta^2} & \text{for } z \in \mathbb{D}_< \cup \mathbb{D}_> \,,\ \eta \in [1, \infty)\,, \end{cases} \tag{6.5.3}$$

where $\boldsymbol{g} = (\langle \boldsymbol{e}_i \,, \boldsymbol{G}\boldsymbol{e}_i \rangle)_{i=1}^{2n}$ denotes the vector of diagonal entries of the resolvent $\boldsymbol{G}^z$.

For a nonrandom vector $\boldsymbol{y} \in \mathbb{C}^{2n}$ with $\|\boldsymbol{y}\|_\infty \leq 1$ we have

$$|\langle \boldsymbol{y} \,, \boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\rangle| \prec \begin{cases} \frac{1}{n\eta} & \text{for } z \in \mathbb{D}_< \,,\ \eta \in [n^{-1+\varepsilon}, 1]\,, \\ \frac{1}{n} + \frac{1}{(n\eta)^2} & \text{for } z \in \mathbb{D}_> \,,\ \eta \in [n^{-1+\varepsilon}, 1]\,, \\ \frac{1}{n\eta^2} & \text{for } z \in \mathbb{D}_< \cup \mathbb{D}_> \,,\ \eta \in [1, \infty)\,. \end{cases} \tag{6.5.4}$$

As an easy consequence we can now prove Corollary 6.2.7.

Proof of Corollary 6.2.7. Let $y \in \mathbb{C}^n$ be an eigenvector of X corresponding to the eigenvalue $z \in \operatorname{Spec} X$ with $|z|^2 \le \rho(S) - \tau_*$. Then the $2n$-vector $(0, y)$ is contained in the kernel of $\boldsymbol{H}^z$. Therefore, (6.2.9) is an easy consequence of (6.5.3) (Compare with the proof of Corollary 1.14 in [7]). □

We recall our normalization of the trace, $\operatorname{tr} \mathbb{1} = 1$, from (6.1.3).

Proof of Theorem 6.5.2. Recall from the beginning of Section 6.3 how our problem can be cast into the setup of [6]. In the regime $z \in \mathbb{D}_<$ we follow the structure of the proof of Theorem 2.9 in [6] and in the regime $z \in \mathbb{D}_>$ the proof of Proposition 7.1 in [6] until the end of Step 1. In fact, the arguments from these proofs can be taken over directly with three important adjustments. The flatness assumption (6.3.1) is used heavily in [6] in order to establish bounds (Theorem 2.5 in [6]) on the deterministic limit of the resolvent and for establishing the stability of the matrix Dyson equation, cf. (6.5.5) below, (Theorem 2.6 in [6]). Since this assumption is violated in our setup we present appropriately adjusted versions of these theorems (Proposition 6.3.2 and Proposition 6.3.3 in [6]). We will also take over the proof of the fluctuation averaging result (Proposition 6.5.5 below) for $\boldsymbol{H}^z$ from [6] since the flatness did not play a role in that proof at all. Note that the η^{-2}-decay in the spectral parameter regime $\eta \ge 1$ was not covered in [6]. But this decay simply follows by using the bounds $\|\boldsymbol{M}^z(\eta)\|_{\max} + \|\boldsymbol{G}^z(\eta)\|_{\max} \le \frac{2}{\eta}$ instead of just $\|\boldsymbol{M}^z(\eta)\|_{\max} + \|\boldsymbol{G}^z(\eta)\|_{\max} \le C$ along the proof.

As in [6] we choose a pseudo-metric d on $\{1, \dots, 2n\}$. Here this pseudo-metric is particularly simple,

$$d(i,j) := \begin{cases} 0 & \text{if } i = j \text{ or } i = j+n \text{ or } j = i+n\,, \\ \infty & \text{otherwise}\,, \end{cases} \qquad i,j = 1, \dots, 2n\,.$$

With this choice of d the matrix $\boldsymbol{H}^z$ satisfies all assumptions in [6] apart from the flatness.

We will now show that as in [6] the resolvent $\boldsymbol{G}^z$ satisfies the *perturbed matrix Dyson equation*

$$-\mathbb{1} = (\mathrm{i}\eta \mathbb{1} - \boldsymbol{A}^z + \widetilde{\mathcal{S}}[\boldsymbol{G}^z(\eta)])\boldsymbol{G}^z(\eta) + \boldsymbol{D}^z(\eta)\,. \tag{6.5.5}$$

Here, $\boldsymbol{A}^z$ is given by (6.3.3),

$$\boldsymbol{D}(\eta) := \boldsymbol{D}^z(\eta) := -(\widetilde{\mathcal{S}}[\boldsymbol{G}^z(\eta)] + \boldsymbol{H}^z - \boldsymbol{A}^z)\boldsymbol{G}^z(\eta)\,, \tag{6.5.6}$$

is a random error matrix and $\widetilde{\mathcal{S}}$ is a slight modification of the operator $\mathcal{S}$ defined in (6.3.3),

$$\widetilde{\mathcal{S}}[\boldsymbol{W}] := \mathbb{E}(\boldsymbol{H}^z - \boldsymbol{A}^z)\boldsymbol{W}(\boldsymbol{H}^z - \boldsymbol{A}^z) = \begin{pmatrix} \operatorname{diag}(Sw_2) & T \odot W_{21}^t \\ T^* \odot W_{12}^t & \operatorname{diag}(S^t w_1) \end{pmatrix}. \tag{6.5.7}$$

Here, $\odot$ denotes the Hadamard product, i.e., for matrices $A = (a_{ij})_{i,j=1}^l$ and $B = (b_{ij})_{i,j=1}^l$, we define their Hadamard product through $(A \odot B)_{ij} := a_{ij}b_{ij}$ for $i, j = 1, \ldots, l$. Moreover, we used the conventions from (6.3.4) for $\boldsymbol{W}$ and introduced the matrix $T \in \mathbb{C}^{n \times n}$ with entries

$$t_{ij} := \mathbb{E}x_{ij}^2\,.$$

Note that in contrast to [6] the matrix $\boldsymbol{M}$ solves (6.3.2), which is given in terms of the operator $\mathcal{S}$ and not $\widetilde{\mathcal{S}}$ (we remark that $\widetilde{\mathcal{S}}$ was denoted by $\mathcal{S}$ in [6]). As we will see below this will not affect the proof, since the entries of the matrix T are of order N^{-1} and thus the off-diagonal terms in (6.5.7) of $\widetilde{\mathcal{S}}$ are negligible.

We will see that $\boldsymbol{D} = \boldsymbol{D}^z$ is small in the entrywise maximum norm

$$\|\boldsymbol{W}\|_{\max} := \max_{i,j=1}^{2n} |w_{ij}|\,,$$

$\boldsymbol{W} = (w_{ij})_{i,j=1}^{2n}$, and use the stability of (6.5.5) to show that $\boldsymbol{G}(\eta) = \boldsymbol{G}^z(\eta)$ approaches $\boldsymbol{M}(\eta) = \boldsymbol{M}^z(\eta)$ defined in (6.3.5) as $n \to \infty$, i.e., we will show that

$$\Lambda(\eta) := \|\boldsymbol{G}(\eta) - \boldsymbol{M}(\eta)\|_{\max}\,, \tag{6.5.8}$$

converges to zero. For simplicity we will only consider the most difficult regime $z \in \mathbb{D}_<$ and $\eta \leq 1$ inside the spectrum. The cases $z \in \mathbb{D}_>$ and $\eta \geq 1$ are similar but simpler and left to the reader. In a more general setup, these regimes are addressed in Chapter 7 below. We simply follow the proof in Section 3 of [6] line by line until the flatness assumption is used. This happens for the first time inside the proof of Lemma 3.3. We therefore replace this lemma by the following modification.

Lemma 6.5.3. *Let $z \in \mathbb{D}_<$. Then*

$$\|\boldsymbol{D}(\eta)\|_{\max} \prec \frac{1}{\sqrt{n}}, \qquad \eta \geq 1\,.$$

Furthermore, we have

$$\|\boldsymbol{D}(\eta)\|_{\max} \mathbf{1}(\Lambda(\eta) \leq n^{-\varepsilon}) \prec \frac{1}{\sqrt{n\eta}}, \qquad \eta \in [n^{-1+\varepsilon}, 1]. \tag{6.5.9}$$

To show Lemma 6.5.3 we follow the proof of its analog, Lemma 3.3 in [6], where the flatness assumption as well as the assumptions that the spectral parameter is in the bulk of the spectrum (formulated as $\rho(\zeta) \geq \delta$ in [6]) are used only implicitly through the upper bound on $\boldsymbol{M}$ (Theorem 2.5 in [6]). However, the conclusion of this theorem clearly still holds in our setup because $\boldsymbol{M}$ has the 2×2-diagonal structure (6.3.5) and the vectors v_1, v_2 and u are bounded by Proposition 6.3.2 and (6.3.26).

We continue following the arguments of Section 3 of [6] using our Lemma 6.5.3 above instead of Lemma 3.3 there. The next step that uses the flatness assumption is the stability of the MDE (Theorem 2.6 in [6]) which shows that the bound (6.5.9) also implies

$$\Lambda(\eta)\, \mathbf{1}(\Lambda(\eta) \leq n^{-\varepsilon}) \prec \frac{1}{\sqrt{n\eta}}\,.$$

In our setup this stability result is replaced by the following lemma whose proof is postponed until the end of the proof of Theorem 6.5.2.

Lemma 6.5.4 (MDE stability)**.** *Suppose that some functions $D_{ab}, G_{ab} : \mathbb{R}_+ \to \mathbb{C}^{n\times n}$ for $a, b = 1, 2$ satisfy* (6.5.5) *with*

$$\boldsymbol{D} := \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}, \qquad \boldsymbol{G} := \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \tag{6.5.10}$$

and the additional constraints

$$\operatorname{tr} G_{11} = \operatorname{tr} G_{22}\,, \qquad \operatorname{Im} \boldsymbol{G} = \frac{1}{2\mathrm{i}}(\boldsymbol{G} - \boldsymbol{G}^*) \textit{ is positive definite}\,. \tag{6.5.11}$$

There is a constant $\lambda_* \gtrsim 1$, depending only on $\mathcal{P}$, such that

$$\|\boldsymbol{G} - \boldsymbol{M}\|_{\max} \chi \lesssim \|\boldsymbol{D}\|_{\max} + \frac{1}{n}, \qquad \chi := \mathbf{1}(\|\boldsymbol{G} - \boldsymbol{M}\|_{\max} \leq \lambda_*)\,, \tag{6.5.12}$$

uniformly for all $z \in \mathbb{D}_< \cup \mathbb{D}_>$*, where* $\boldsymbol{M}(\eta) = \boldsymbol{M}^z(\eta)$ *is defined in* (6.3.5).

Furthermore, there exist eight matrix valued functions $R_{ab}^{(k)} : \mathbb{R}_+ \to \mathbb{C}^{n\times n}$ *with* $a, b, k = 1, 2$*, depending only on* z *and* S*, and satisfying* $\|R_{ab}^{(k)}\|_\infty \lesssim 1$*, such that*

$$\Big|\operatorname{tr}[\operatorname{diag}(\boldsymbol{y})(\boldsymbol{G}-\boldsymbol{M})]\Big|\,\chi \lesssim \max_{a,b,k=1,2}\Big|\operatorname{tr}[\operatorname{diag}(R_{ab}^{(k)}y_k)D_{ab}]\Big| + \|\boldsymbol{y}\|_\infty\Big(\frac{1}{n} + \|\boldsymbol{D}\|_{\max}^2\Big), \tag{6.5.13}$$

uniformly for all $z \in \mathbb{D}_< \cup \mathbb{D}_>$ *and* $\boldsymbol{y} = (y_1, y_2) \in \mathbb{C}^{2n}$.

The important difference between Theorem 2.6 in [6] and Lemma 6.5.4 above is the additional assumption (6.5.11) imposed on the solution of the perturbed MDE. This assumption is satisfied for the resolvent of the matrix $\boldsymbol{H}^z$ because of the 2×2-block structure (6.2.11). In fact, we apply the block decomposition in (6.5.10) to $\boldsymbol{G} = (\boldsymbol{H}^z - \mathrm{i}\eta\mathbb{1})^{-1}$ and obtain

$$G_{11}(\eta) \;=\; \frac{\mathrm{i}\eta\mathbb{1}}{(X-z\mathbb{1})(X-z\mathbb{1})^* + \eta^2\mathbb{1}}\,, \qquad G_{22}(\eta) \;=\; \frac{\mathrm{i}\eta\mathbb{1}}{(X-z\mathbb{1})^*(X-z\mathbb{1}) + \eta^2\mathbb{1}}\,.$$

Using Lemma 6.5.4 in the remainder of the proof of the entrywise local law in Section 3 of [6] completes the proof of (6.5.2).

To see (6.5.4) we use the fluctuation averaging mechanism, which was first established for generalized Wigner matrices with Bernoulli entries in [72]. The following proposition is stated and proven as Proposition 3.4 in [6]. Since the flatness condition was not used in its proof at all, we simply take it over.

Proposition 6.5.5 (Fluctuation averaging)**.** *Let* $z \in \mathbb{D}_< \cup \mathbb{D}_>$*,* $\varepsilon \in (0, 1/2)$*,* $\eta \geq n^{-1}$ *and* Ψ *a nonrandom control parameter such that* $n^{-1/2} \leq \Psi \leq n^{-\varepsilon}$*. Suppose the local law holds true in the form*

$$\|\boldsymbol{G}(\eta) - \boldsymbol{M}(\eta)\|_{\max} \;\prec\; \Psi\,.$$

Then for any nonrandom vector $y \in \mathbb{C}^n$ *with* $\|y\|_\infty \leq 1$ *we have*

$$\max_{a,b=1,2}\Big|\operatorname{tr}[\operatorname{diag}(y)D_{ab}]\Big| \;\prec\; \Psi^2\,,$$

where $D_{ab} \in \mathbb{C}^{n \times n}$*,* $a, b = 1, 2$*, are the blocks of the error matrix*

$$\boldsymbol{D}(\eta) \;=\; \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}\,,$$

which was defined in (6.5.6)*.*

Using this proposition the averaged local law (6.5.4) follows from (6.5.2) and (6.5.13). This completes the proof of Theorem 6.5.2. □

Proof of Lemma 6.5.4. We write (6.5.5) in the 2×2 - block structure

$$\begin{aligned} &\begin{pmatrix} \operatorname{diag}(\mathrm{i}\eta + Sg_2) & z\mathbb{1} \\ \overline{z}\mathbb{1} & \operatorname{diag}(\mathrm{i}\eta + S^t g_1) \end{pmatrix} \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \\ &\qquad = -\begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix} - \begin{pmatrix} D_{11} + (T \odot G_{21}^t)G_{21} & D_{12} + (T \odot G_{21}^t)G_{22} \\ D_{21} + (T^* \odot G_{12}^t)G_{11} & D_{22} + (T^* \odot G_{12}^t)G_{22} \end{pmatrix}, \end{aligned} \tag{6.5.14}$$

where we introduced $\boldsymbol{g} = (g_1, g_2) \in \mathbb{C}^{2n}$, the vector of the diagonal elements of $\boldsymbol{G}$.

We restrict the following calculation to the regime, where $\|\boldsymbol{G}(\eta) - \boldsymbol{M}(\eta)\|_{\max} \le \lambda_*$ for some sufficiently small λ_* in accordance with the characteristic function on the left-hand side of (6.5.12). In particular,

$$\|\boldsymbol{g}(\eta) - \mathrm{i}\boldsymbol{v}(\eta)\|_\infty \;\le\; \lambda_*\,. \tag{6.5.15}$$

Since by (6.2.4) and (6.3.5) the identity

$$\begin{pmatrix} \mathrm{i}\operatorname{diag}(\eta + Sv_2(\eta)) & z\mathbb{1} \\ \overline{z}\mathbb{1} & \mathrm{i}\operatorname{diag}(\eta + S^t v_1(\eta)) \end{pmatrix}^{-1} \;=\; -\boldsymbol{M}(\eta)\,,$$

holds we infer from the smallness of $\|\boldsymbol{g} - \mathrm{i}\boldsymbol{v}\|_{\max}$ that the inverse of the first matrix factor on the left-hand side of (6.5.14) is bounded and satisfies

$$\left\| \begin{pmatrix} \operatorname{diag}(\mathrm{i}\eta + Sg_2) & z\mathbb{1} \\ \overline{z}\mathbb{1} & \operatorname{diag}(\mathrm{i}\eta + S^t g_1) \end{pmatrix}^{-1} + \boldsymbol{M} \right\|_{\max} \;\lesssim\; \|\boldsymbol{g} - \mathrm{i}\boldsymbol{v}\|_{\max}\,. \tag{6.5.16}$$

Using this in (6.5.14) yields

$$\begin{aligned} &\boldsymbol{G} + \begin{pmatrix} \mathrm{diag}(\mathrm{i}\eta + Sg_2) & z\mathbb{1} \\ \bar{z}\mathbb{1} & \mathrm{diag}(\mathrm{i}\eta + S^t g_1) \end{pmatrix}^{-1} \\ &\qquad = \boldsymbol{M}\boldsymbol{D} + \mathcal{O}\Big(\|\boldsymbol{g} - \boldsymbol{v}\|_{\max}\|\boldsymbol{D}\|_{\max} + \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n}\Big), \end{aligned} \tag{6.5.17}$$

where we applied the simple estimate

$$\begin{aligned} \|(T \odot G_{ab}^t)G_{cd}\|_{\max} &\lesssim \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n}\|\boldsymbol{G} - \boldsymbol{M}\|_{\max}\|\boldsymbol{M}\|_{\max} + \frac{1}{n}\|\boldsymbol{M}\|_{\max}^2 \\ &\lesssim \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n}, \end{aligned} \tag{6.5.18}$$

which follows from

$$\|T\|_{\max} \lesssim \frac{1}{n}.$$

Thus the diagonal elements $\boldsymbol{g}$ of $\boldsymbol{G}$ satisfy (6.2.19) with an error term $\boldsymbol{d}$ that is given by

$$\boldsymbol{d} = ((\boldsymbol{M}\boldsymbol{D})_{ii})_{i=1}^{2n} + \mathcal{O}\Big(\|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n}\Big). \tag{6.5.19}$$

Here we used $\|\boldsymbol{D}\|_{\max} \lesssim \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}$, which follows directly from (6.5.5) and (6.3.2). With (6.3.21) and (6.3.22) in Proposition 6.3.3, the stability result on (6.2.19), we conclude that

$$\|\boldsymbol{g} - \mathrm{i}\boldsymbol{v}\|_\infty \lesssim \|\boldsymbol{D}\|_{\max} + \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n}, \tag{6.5.20}$$

and that

$$|\langle \boldsymbol{y}, \boldsymbol{g} - \mathrm{i}\boldsymbol{v}\rangle| \lesssim \Big|\mathrm{tr}[\mathrm{diag}(\boldsymbol{R}\boldsymbol{y})\boldsymbol{M}\boldsymbol{D}]\Big| + \|\boldsymbol{D}\|_{\max}^2 + \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n}, \tag{6.5.21}$$

for some bounded $\boldsymbol{R} \in \mathbb{C}^{2n\times 2n}$ and any $\boldsymbol{y} \in \mathbb{C}^{2n}$ with $\|\boldsymbol{y}\|_\infty \leq 1$, respectively. Combining (6.5.16) with (6.5.17) and (6.5.20) yields

$$\|\boldsymbol{G} - \boldsymbol{M}\|_{\max} \lesssim \|\boldsymbol{D}\|_{\max} + \|\boldsymbol{G} - \boldsymbol{M}\|_{\max}^2 + \frac{1}{n}.$$

By choosing λ_* sufficiently small we may absorb the quadratic term of the difference $\boldsymbol{G}-\boldsymbol{M}$ on the right-hand side into the left-hand side and (6.5.12) follows. Using (6.5.12) in (6.5.21) to estimate the term $\|\boldsymbol{G}-\boldsymbol{M}\|_{\max}^2$ proves (6.5.13). □

We use a standard argument to conclude from (6.5.4) the following statement about the number of eigenvalues $\lambda_i(z)$ of $\boldsymbol{H}^z$ in a small interval centered at zero.

Lemma 6.5.6. *Let $\varepsilon > 0$. Then*

$$\#\{i \,:\, |\lambda_i(z)| \le \eta\} \prec n\eta\,, \tag{6.5.22}$$

uniformly for all $\eta \ge n^{-1+\varepsilon}$ and $z \in \mathbb{D}_<$.

Furthermore, we have

$$\sup_{z\in\mathbb{D}_>} \frac{1}{|\lambda_i(z)|} \prec n^{1/2}\,. \tag{6.5.23}$$

Proof. For the proof of (6.5.22) we realize that (6.5.2) implies a uniform bound on the resolvent elements up to the spectral scale $\eta \ge n^{-1+\varepsilon}$. Thus we have

$$\frac{\#\Sigma_\eta}{2\eta} \le \sum_{i\in\Sigma_\eta} \frac{\eta}{\eta^2+\lambda_i(z)^2} \le 2n \operatorname{Im}\operatorname{tr} \boldsymbol{G}^z(\eta) \prec n\,,$$

where $\Sigma_\eta := \{i : |\lambda_i(z)| \le \eta\}$. Here, we used the normalization of the trace (6.1.3).

Before proving (6.5.23), we first establish that

$$\frac{1}{|\lambda_i(z)|} \prec n^{1/2}\,, \tag{6.5.24}$$

uniformly for $z \in \mathbb{D}_>$. We use (6.5.4) and $\langle \boldsymbol{v}(\eta)\rangle \sim \eta$ to estimate

$$\frac{\eta}{\eta^2+\lambda_i(z)^2} \le 2n \operatorname{Im}\operatorname{tr} \boldsymbol{G}^z(\eta) \prec n\eta + \frac{1}{n\eta^2}\,, \tag{6.5.25}$$

with the choice $\eta := n^{-1/2-\varepsilon}$ for any $\varepsilon > 0$. This immediately implies $|\lambda_i(z)|^{-1} \prec n^{1/2+\varepsilon}$, hence (6.5.24). For the stronger bound (6.5.23) we use that $z \mapsto \operatorname{Im}\operatorname{tr} \boldsymbol{G}^z(\eta)$ is a Lipschitz continuous function (with a Lipschitz constant $C\eta^{-2}$ uniformly in z) and that $\mathbb{D}_>$ is compact, so the second bound in (6.5.25) holds even after taking the supremum over $z \in \mathbb{D}_>$. Thus

$$\sup_{z\in\mathbb{D}_>} \frac{\eta}{\eta^2+\lambda_i(z)^2} \le 2n \sup_{z\in\mathbb{D}_>} \operatorname{Im}\operatorname{tr} \boldsymbol{G}^z(\eta) \prec n\eta + \frac{1}{n\eta^2}$$

holds for $\eta := n^{-1/2-\varepsilon}$. From the last inequality we easily conclude (6.5.23). □

6.5.2. Local inhomogeneous circular law. For the following proof of Theorem 6.2.6 we recall that without loss of generality, we are assuming that $\rho(S) = 1$ which can be obtained by a simple rescaling of X. Moreover, from (6.4.1), for $\tau_* > 0$ and $\tau^* > 1+\tau_*$, we recall the notation

$$\mathbb{D}_< := \{z \in \mathbb{C} \mid |z|^2 \leq 1 - \tau_*\}, \quad \mathbb{D}_> := \{z \in \mathbb{C} \mid 1 + \tau_* \leq |z|^2 \leq \tau^*\}.$$

Proof of Theorem 6.2.6. We start with the proof of part (i) of Theorem 6.2.6. We will estimate each term on the right-hand side of (6.2.15). Let $w \in \mathbb{D}_<$. We suppress the τ dependence of v_1 in this proof but it will always be evaluated at $\tau = |z|^2$.

As supp $f \subset D_\varphi(0)$, $a > 0$ and $w \in \mathbb{D}_<$ we can assume that the integration domains of the $\mathrm{d}^2 z$ integrals in (6.2.15) are $\mathbb{D}_<$ instead of $\mathbb{C}$. Hence, it suffices to prove every bound along the proof of (i) uniformly for $z \in \mathbb{D}_<$.

To begin, we estimate the first term in (6.2.15). Since

$$\log|\det(\boldsymbol{H}^z - \mathrm{i}T\mathbb{1})| = 2n \log T + \sum_{j=1}^n \log\left(1 + \frac{\lambda_j^2}{T^2}\right)$$

and the integral of $\Delta f_{w,a}$ over $\mathbb{C}$ vanishes as $f \in C_0^2(\mathbb{C})$, we obtain

$$\left|\frac{1}{4\pi n}\int_{\mathbb{C}} \Delta f_{w,a}(z) \log|\det(\boldsymbol{H}^z - \mathrm{i}T\mathbb{1})|\mathrm{d}^2 z\right| \leq \frac{1}{2\pi}\int_{\mathbb{C}} |\Delta f_{w,a}(z)| \frac{\mathrm{tr}\,((\boldsymbol{H}^z)^2)}{T^2}\mathrm{d}^2 z. \tag{6.5.26}$$

Here, we used $\log(1+x) \leq x$ for $x \geq 0$. Furthermore, if $|z| \leq 1$, then we have

$$\mathrm{tr}((\boldsymbol{H}^z)^2) = \frac{1}{n}\sum_{i,j=1}^n (x_{ij} - z\delta_{ij})(\overline{x_{ij}} - \bar{z}\delta_{ij}) \leq \frac{2}{n}\sum_{i,j=1}^n |x_{ij}|^2 + 2|z|^2 \prec 1, \tag{6.5.27}$$

where we applied (6.1.3) in the first and (6.5.1) in the last step. Therefore, choosing $T := n^{100}$, we conclude from (6.5.26) and (6.5.27) that the first term in (6.2.15) is stochastically dominated by $n^{-1+2a}\|\Delta f\|_1$.

To control the second term on right-hand side of (6.2.15), we define

$$I(z) := \int_0^T |\mathrm{Im}\, m^z(\mathrm{i}\eta) - \langle v_1(\eta)\rangle|\,\mathrm{d}\eta \tag{6.5.28}$$

for $z \in \mathbb{D}_<$. We will conclude below the following lemma.

Lemma 6.5.7. *For every $\delta > 0$ and $p \in \mathbb{N}$, there is a positive constant C, depending only on δ and p in addition to the model parameters and τ_*, such that*

$$\sup_{z\in\mathbb{D}_<} \mathbb{E} I(z)^p \le C\frac{n^{\delta p}}{n^p}. \tag{6.5.29}$$

We now show that this moment bound on $I(z)$ will yield that the second term in (6.2.15) is $\prec n^{-1+2a}\|\Delta f\|_1$. Indeed, for every $p \in \mathbb{N}$ and $\delta > 0$, using Hölder's inequality, we estimate

$$\begin{aligned} \mathbb{E}\left|\int_{\mathbb{C}} \Delta f_{w,a}(z) \int_0^T [\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta)\rangle]\,\mathrm{d}\eta\,\mathrm{d}^2 z\right|^p & \\ \le \int_{\mathbb{C}} \cdots \int_{\mathbb{C}} \prod_{i=1}^p |\Delta f_{w,a}(\zeta_i)| \prod_{i=1}^p (\mathbb{E} I(\zeta_i)^p)^{1/p}\,\mathrm{d}^2\zeta_1 \ldots \mathrm{d}^2\zeta_p & \\ \le C\|\Delta f\|_1^p \frac{n^{\delta p + 2ap}}{n^p}. & \end{aligned} \tag{6.5.30}$$

Applying Chebyshev's inequality to (6.5.30) and using that $\delta > 0$ and p were arbitrary, we get

$$\left|\int_{\mathbb{C}} \Delta f_{w,a}(z) \int_0^T \operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta)\rangle \mathrm{d}\eta\,\mathrm{d}^2 z\right| \prec n^{-1+2a}\|\Delta f\|_1.$$

Hence, the bound on the second term on the right-hand side of (6.2.15) follows once we have proven (6.5.29).

For the third term in (6.2.15), notice that the integrand is bounded by $C\eta^{-2}$ so it is bounded by $n^{2a}T^{-1}\|\Delta f\|_1$. This concludes the proof of (i) of Theorem 6.2.6 up to the proof of Lemma 6.5.7 which is given below.

We now turn to the proof of (ii). We will use an interpolation between the random matrix X and an independent Ginibre matrix $\widehat{X}$ together with the well-known result that a Ginibre matrix does not have any eigenvalues $|\lambda| \ge 1 + \tau_*$ with very high probability. With the help of (6.5.23) we will control the number of eigenvalues outside of the disk of radius $1 + \tau^*$ along the flow. We fix $\tau^* > 1 + \tau_*$.

Let $(\widehat{x}_{ij})_{i,j=1}^n$ be independent centered complex Gaussians of variance n^{-1}, i.e., $\mathbb{E}\,\widehat{x}_{ij} = 0$ and $\mathbb{E}|\widehat{x}_{ij}|^2 = n^{-1}$. We set $\widehat{X} := (\widehat{x}_{ij})_{i,j=1}^n$, i.e., $\widehat{X}$ is a Ginibre matrix. We denote the eigenvalues of $\widehat{X}$ by $\widehat{z}_1, \ldots, \widehat{z}_n$.

For $t \in [0, 1]$, we denote the spectral radius of the matrix $tS + (1-t)E$ by $\rho_t := \rho(tS + (1-t)E)$, where E is the $n \times n$ matrix with entries $e_{ij} := 1/n$, $E = (e_{ij})_{i,j=1}^n$.

Furthermore, we define

$$X^t := \rho_t^{-1/2}\left(tX + (1-t)\widehat{X}\right), \quad \boldsymbol{H}^{z,t} := \begin{pmatrix} 0 & X^t - z\mathbb{1} \\ (X^t - z\mathbb{1})^* & 0 \end{pmatrix}$$

for $t \in [0,1]$. The eigenvalues of X^t and $\boldsymbol{H}^{z,t}$ are denoted by z_i^t and $\lambda_k^t(z)$, respectively, for $i = 1, \ldots, n$ and $k = 1, \ldots, 2n$. The one-parameter family $t \mapsto X^t$ interpolates between X and $\widehat{X}$ by keeping the spectral radius of the variance matrix at constant one.

Note that $\|(X^t - z\mathbb{1})^{-1}\|_2 = \max_{k=1}^{2n} |\lambda_k^t(z)|^{-1}$. We can apply Lemma 6.5.6 to the matrices X^t for any t to get

$$\sup_{z \in \mathbb{D}_>} \left\|(X^t - z\mathbb{1})^{-1}\right\|_2 \prec n^{1/2}$$

uniformly in t from (6.5.23). In fact, the estimate can be strengthened to

$$\sup_{t\in[0,1]} \sup_{z \in \mathbb{D}_>} \left\|(X^t - z\mathbb{1})^{-1}\right\|_2 \prec n^{1/2} \tag{6.5.31}$$

exactly in the same way as (6.5.24) was strengthened to (6.5.23), we only need to observe that the two-parameter family $(z,t) \mapsto \operatorname{Im}\operatorname{tr}\boldsymbol{G}^{z,t}(\eta)$ is Lipschitz continuous in both variables, where $\boldsymbol{G}^{z,t}$ denotes the resolvent of $\boldsymbol{H}^{z,t}$.

Let γ be the circle in $\mathbb{C}$ centered at zero with radius $1+\tau_*$. For $t \in [0,1]$, we have

$$N(t) := \#\{i \mid |z_i^t| \le 1 + \tau_*\} = \frac{n}{2\pi\mathrm{i}} \int_\gamma \operatorname{tr}\left((X^t - z\mathbb{1})^{-1}\right) \mathrm{d}z,$$

where $\operatorname{tr}\colon \mathbb{C}^{n\times n} \to \mathbb{C}$ denotes the normalized trace, i.e., $\operatorname{tr}\mathbb{1} = 1$. Due to (6.5.31) $N(t)$ is a continuous function of t. Thus, $N(t)$ is constant as a continuous integer-valued function.

Using Corollary 2.3 of [75], we obtain that $\#\{k \mid |\widehat{z}_k| \ge \tau^*\} = 0$ with very high probability. Furthermore, $\#\{k \mid \widehat{z}_k \in \mathbb{D}_>\} = 0$ with very high probability by (6.5.31). Thus,

$$N(1) = N(0) = n - \#\{k \mid \widehat{z}_k \in \mathbb{D}_>\} - \#\{k \mid |\widehat{z}_k| \ge \tau^*\} = n$$

with very high probability which concludes the proof of (ii) and hence of Theorem 6.2.6. □

Remark 6.5.8. In the above proof we showed that $\|\boldsymbol{H}^z\| \le C$ with very high probability via an interpolation argument using the norm-boundedness of a Ginibre matrix and the local law for the entire interpolating family. Robust upper bounds on the norm of random matrices are typically proven by a simple moment method. Such approach also applies here. For example, one may follow the proof of Lemma 7.2 in [70], and estimate every moment $\mathbb{E}|x_{ij}|^k$ by its maximum over all i, j. The final constant estimating $\|\boldsymbol{H}^z\|$ will not be optimal due to these crude bounds, but it will still only depend on s^* and μ_m from (6.2.1) and (6.2.2), respectively. This argument is very robust, in particular it does not use Hermiticity.

In the proof of Lemma 6.5.7, we will need an estimate on the smallest singular value of $X - z\mathbb{1}$ presented in the following Proposition 6.5.9. In fact, it will be used to control the $\mathrm{d}\eta$-integral in the second term on the right-hand side of (6.2.15) for $\eta \leq n^{-1+\varepsilon}$. Notice that Proposition 6.5.9 is the only result in our proof of Theorem 6.2.6 which requires the entries of X to have a bounded density.

Adapting the proof of [40, Lemma 4.12] with the bounded density assumption to our setting, we obtain the following proposition.

Proposition 6.5.9 (Smallest singular value of $X - z\mathbb{1}$)**.** *Under the condition* (6.2.3)*, there is a constant C, depending only on α, such that*

$$\mathbb{P}\left(\min_{i=1}^{2n} |\lambda_i(z)| \leq \frac{u}{n}\right) \leq C u^{2\alpha/(1+\alpha)} n^{\beta+1} \tag{6.5.32}$$

for all $u > 0$ and $z \in \mathbb{C}$.

Proof. We follow the proof in [40] and explain the differences. Let $R_1, \ldots, R_n$ denote the rows of $\sqrt{n}X - z\mathbb{1}$. Proceeding as in [40] but using our normalization conventions, we are left with estimating

$$\mathbb{P}\left(n|\langle R_i, y\rangle| \leq \frac{u}{\sqrt{n}}\right)$$

uniformly for $i \in \{1, \ldots, n\}$ and for arbitrary $y \in \mathbb{C}^n$ satisfying $\|y\|_2 = 1/\sqrt{n}$ and tracking its dependence on $u > 0$. We choose $j \in \{1, \ldots, n\}$ such that $|y_j| \geq 1/\sqrt{n}$ and compute the conditional probability

$$\mathbb{P}_{ij} := \mathbb{P}\Big(n|\langle R_i, y\rangle| \leq \frac{u}{\sqrt{n}}\Big| x_{i1}, \ldots, \widehat{x_{ij}}, \ldots, x_{in}\Big) = \int_{\mathbb{C}} \mathbf{1}\left(\left|\frac{a}{y_j} + \zeta\right| \leq \frac{u}{y_j\sqrt{n}}\right) f_{ij}(\zeta) \mathrm{d}^2\zeta,$$

where a is independent of x_{ij}. Using (6.2.3) and $|y_j| \geq 1/\sqrt{n}$, we get

$$|\mathbb{P}_{ij}| \leq \left|\pi \frac{u}{y_j\sqrt{n}}\right|^{2\alpha/(1+\alpha)} \|f_{ij}\|_{1+\alpha} \leq (\pi u)^{2\alpha/(1+\alpha)} n^{\beta}.$$

Thus, $\mathbb{P}\left(n|\langle R_i\,, y\rangle| \leq u/\sqrt{n}\right) \leq (\pi u)^{2\alpha/(1+\alpha)} n^{\beta}$ which concludes the proof of (6.5.32) as in [40]. □

Proof of Lemma 6.5.7. To show (6.5.29), we use the following estimate which converts a bound in $\prec$ into a moment bound. For every nonnegative random variable satisfying $Y \prec 1/n$ and $Y \leq n^c$ for some $c > 0$ the p-th moment is bounded by

$$\mathbb{E}Y^p \leq \mathbb{E}Y^p \mathbf{1}(Y \leq n^{\delta-1}) + \left(\mathbb{E}Y^{2p}\right)^{1/2} \left(\mathbb{P}\left(Y > n^{\delta-1}\right)\right)^{1/2} \leq C\frac{n^{p\delta}}{n^p}, \tag{6.5.33}$$

for all $p \in \mathbb{N}$, $\delta > 0$ and for some $C > 0$, depending on c, p and δ.

As a first step in the proof of (6.5.29), we choose $\varepsilon \in (0, 1/2)$, split the $\mathrm{d}\eta$ integral in the definition of $I(z)$, (6.5.28), and consider the regimes $\eta \leq n^{-1+\varepsilon}$ and $\eta \geq n^{-1+\varepsilon}$, separately. For $\eta \leq n^{-1+\varepsilon}$, we compute

$$\int_0^{n^{-1+\varepsilon}} \operatorname{Im} m^z(\mathrm{i}\eta)\mathrm{d}\eta = \frac{1}{2n}\sum_{i=1}^{n} \log\left(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right).$$

We recall that $\lambda_1, \ldots, \lambda_{2n}$ are the eigenvalues of $\boldsymbol{H}^z$. Therefore, (6.5.28) yields

$$\begin{aligned}
\int_0^T & \left[\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta)\rangle\right] \mathrm{d}\eta \\
&= \frac{1}{n}\sum_{|\lambda_i|<n^{-l}} \log\left(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right) + \frac{1}{n}\sum_{|\lambda_i|\geq n^{-l}} \log\left(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right) \\
&\quad - \int_0^{n^{-1+\varepsilon}} \langle v_1(\eta)\rangle \mathrm{d}\eta + \int_{n^{-1+\varepsilon}}^1 \left[\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta)\rangle\right] \mathrm{d}\eta \\
&\quad + \int_1^T \left[\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta)\rangle\right] \mathrm{d}\eta.
\end{aligned} \tag{6.5.34}$$

Here, $l \in \mathbb{N}$ is a large fixed integer to be chosen later.

We will estimate each of the terms on the right-hand side of (6.5.34) individually. We will apply (6.5.33) for estimating the absolute value of the second, fourth and fifth term on the right-hand side of (6.5.34). For the first term, we will need a separate argument based on Proposition 6.5.9, which we present now.

For the first term in (6.5.34), we compute

$$\mathbb{E}\left(\frac{1}{n}\sum_{|\lambda_i|\le n^{-l}}\log\left(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right)\right)^p \le \mathbb{E}\left[\log^p\left(1+\frac{n^{-2+2\varepsilon}}{\lambda_n^2}\right)\mathbf{1}(\lambda_n\le n^{-l})\right]$$
$$\le C\mathbb{E}\left[|\log\lambda_n|^p\mathbf{1}(\lambda_n\le n^{-l})\right]$$

for some constant $C>0$ independent of n. We compute the expectation directly

$$\mathbb{E}\left[|\log\lambda_n|^p\mathbf{1}(\lambda_n\le n^{-l})\right] = p\int_{l\log n}^{\infty}\mathbb{P}\left(\lambda_n\le \mathrm{e}^{-t}\right)t^{p-1}\mathrm{d}t$$
$$\le Cn^{\beta+1+2\alpha/(1+\alpha)}\int_{l\log n}^{\infty}t^{p-1}\mathrm{e}^{-2\alpha t/(1+\alpha)}\mathrm{d}t.$$

Here, we applied (6.5.32) in Proposition 6.5.9 with $u=\mathrm{e}^{-t}n$. Choosing l large enough, depending on α, β and p, we obtain that the right-hand side is smaller than n^{-p}. This shows the bound (6.5.29) for the first term in (6.5.34).

To estimate the second term on the right-hand side of (6.5.34), we decompose the sum into three regimes, $n^{-l}\le|\lambda_i|<n^{-1+\varepsilon}$, $n^{-1+\varepsilon}\le|\lambda_i|<n^{-1/2}$ and $n^{-1/2}\le|\lambda_i|$.

For the first regime, we use (6.5.22) with $\eta=n^{-1+\varepsilon}$ and $\log(1+n^{-2+2\varepsilon+l})\le C\log n$ to get

$$\frac{1}{n}\sum_{|\lambda_i|\in[n^{-l},n^{-1+\varepsilon}]}\log\left(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right)\le\frac{C\log n}{n}\#\{i\colon|\lambda_i|\le n^{-1+\varepsilon}\}\prec\frac{n^{\varepsilon}}{n}. \tag{6.5.35}$$

As this sum is clearly polynomially bounded in n we can apply (6.5.33) to conclude that the first regime of the second term in (6.5.34) fulfills the moment bound in (6.5.29).

For the intermediate regime, due to the symmetry $\mathrm{Spec}(\boldsymbol{H}^z)=-\mathrm{Spec}(\boldsymbol{H}^z)$, we only consider the positive eigenvalues. We decompose the interval $[n^{-1+\varepsilon},n^{-1/2}]$ into dyadic intervals of the form $[\eta_k,\eta_{k+1}]$, where $\eta_k:=2^kn^{-1+\varepsilon}$. Thus, we obtain

$$\frac{1}{n}\sum_{|\lambda_i|\in[n^{-1+\varepsilon},n^{-1/2}]}\log\left(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right)\le\frac{2}{n}\sum_{k=0}^{N}\sum_{\lambda_i\in[\eta_k,\eta_{k+1}]}\log\left(1+\frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right)\prec\frac{n^{\varepsilon}}{n}, \tag{6.5.36}$$

where we introduced $N=\mathcal{O}(\log n)$ in the first step. Moreover, we used the monotonicity of the logarithm, $\log(1+x)\le x$ in the last step and the following consequence of (6.5.22):

$$\#\{i\colon\lambda_i\in[\eta_k,\eta_{k+1}]\}\le\#\{i\colon|\lambda_i|\le\eta_{k+1}\}\prec n^{\varepsilon}2^{k+1}.$$

The left-hand side of (6.5.36) is trivially bounded by $\log 2$. Therefore, applying (6.5.33) to the left-hand side of (6.5.36), we conclude that it satisfies the moment estimate in (6.5.29).

For estimating the second term in (6.5.34) in the third regime, employing $|\lambda_i| \geq n^{-1/2}$ and $\log(1+x) \leq x$, we obtain

$$\frac{1}{n} \sum_{|\lambda_i| \geq n^{-1/2}} \log\left(1 + \frac{n^{-2+2\varepsilon}}{\lambda_i^2}\right) \leq \frac{1}{n} \sum_{|\lambda_i| \geq n^{-1/2}} \log\left(1 + n^{-1+2\varepsilon}\right) \leq \frac{2n^{2\varepsilon}}{n}. \tag{6.5.37}$$

Here, we used that $\boldsymbol{H}^z$ has $2n$ eigenvalues (counted with multiplicities). This deterministic bound and (6.5.33) imply that the moments of this sum are bounded by the right-hand side in (6.5.29).

Combining the estimates in these three regimes, (6.5.35) , (6.5.36) and (6.5.37), we conclude that the second term in (6.5.34) satisfies the moment bound in (6.5.29).

We now estimate the third term on the right-hand side of (6.5.34). Since $\boldsymbol{v} \sim 1$ for $z \in \mathbb{D}_<$ and $\eta \leq 1$ by (6.3.10), the p-th power of the third term is immediately bounded by the right-hand side of (6.5.29).

To bound the fourth and fifth term in (6.5.34), we note that $\operatorname{Im} m^z(\mathrm{i}\eta) = \langle \boldsymbol{g}(\eta) \rangle$ for $\eta > 0$ and recalling the choice $T = n^{100}$, we obtain

$$\int_{n^{-1+\varepsilon}}^{1} |\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta) \rangle| \, \mathrm{d}\eta \prec \frac{n^{\varepsilon}}{n}, \qquad \int_1^T |\operatorname{Im} m^z(\mathrm{i}\eta) - \langle v_1(\eta) \rangle| \, \mathrm{d}\eta \prec \frac{1}{n} \tag{6.5.38}$$

from the first and third regime in (6.5.4) with $\boldsymbol{y} = 1$. As the integrands are bounded by n^2 trivially (6.5.33) yields that the moments of the fourth and fifth term in (6.5.34) are bounded by the right-hand side in (6.5.29).

Since $\varepsilon \in (0, 1/2)$ was arbitrary this concludes the proof of (6.5.29). $\square$

6.6. Proof of Lemma 6.2.3

The existence and uniqueness of the solution to (6.2.4) will be a consequence of the existence and uniqueness of the solution to the matrix Dyson equation

$$-\boldsymbol{M}^{-1}(\eta) = \mathrm{i}\eta \mathbb{1} - \boldsymbol{A} + \mathcal{S}[\boldsymbol{M}(\eta)]. \tag{6.6.1}$$

Note that $\boldsymbol{A} \in \mathbb{C}^{2n \times 2n}$ and $\mathcal{S} \colon \mathbb{C}^{2n \times 2n} \to \mathbb{C}^{2n \times 2n}$ were defined in (6.3.3).

The matrix Dyson equation, (6.6.1), has a unique solution under the constraint that the imaginary part

$$\operatorname{Im} \boldsymbol{M} := \frac{1}{2\mathrm{i}}(\boldsymbol{M} - \boldsymbol{M}^*)$$

is positive definite. This was established in [96]. In the context of random matrices, (6.6.1) was studied in [6].

In the following proof, for vectors $a, b, c, d \in \mathbb{C}^n$, we will denote the $2n \times 2n$ matrix having diagonal matrices with diagonals a, b, c, d on its top-left, top-right, lower-left and lower-right $n \times n$ blocks, respectively, by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} \operatorname{diag} a & \operatorname{diag} b \\ \operatorname{diag} c & \operatorname{diag} d \end{pmatrix} \in \mathbb{C}^{2n \times 2n}.$$

Proof of Lemma 6.2.3. We show that there is a bijection between the solutions of (6.6.1) with positive definite imaginary part $\operatorname{Im} \boldsymbol{M}$ and the positive solutions of (6.3.6).

We remark that (6.6.1) implies that there are vector-valued functions $a, b, c, d \colon \mathbb{R}_+ \to \mathbb{C}^n$ such that for all $\eta > 0$ we have

$$\boldsymbol{M}(\eta) = \begin{pmatrix} a(\eta) & b(\eta) \\ c(\eta) & d(\eta) \end{pmatrix}. \tag{6.6.2}$$

First, we show that $\operatorname{Im} \operatorname{diag} \boldsymbol{M}$ is a solution of (6.3.6) satisfying $\operatorname{Im} \operatorname{diag} \boldsymbol{M} > 0$ if $\boldsymbol{M}$ satisfies (6.6.1) and $\operatorname{Im} \boldsymbol{M}$ is positive definite. Due to (6.6.2), multiplying (6.6.1) by $\boldsymbol{M}$ yields that (6.6.1) is equivalent to

$$\begin{aligned} -1 &= \mathrm{i}\eta a + aSd + b\bar{z}, & 0 &= \mathrm{i}\eta b + za + bS^t a, \\ 0 &= \mathrm{i}\eta c + \bar{z} d + cSd, & -1 &= \mathrm{i}\eta d + dS^t a + zc \end{aligned} \tag{6.6.3}$$

Solving the second relation in (6.6.3) for b and the third relation in (6.6.3) for c, we obtain

$$b = -\frac{za}{\mathrm{i}\eta + S^t a}, \qquad c = -\frac{\bar{z} d}{\mathrm{i}\eta + Sd}. \tag{6.6.4}$$

Plugging the first relation in (6.6.4) into the first relation in (6.6.3) and the second relation in (6.6.4) into the fourth relation in (6.6.3) and dividing the results by a and d, respectively, imply

$$-\frac{1}{a} = \mathrm{i}\eta + Sd - \frac{|z|^2}{\mathrm{i}\eta + S^t a}, \qquad -\frac{1}{d} = \mathrm{i}\eta + S^t a - \frac{|z|^2}{\mathrm{i}\eta + Sd}.$$

Therefore, if a and d are purely imaginary then $(\operatorname{Im} a, \operatorname{Im} d) = -\mathrm{i}(a, d)$ will fulfill (6.3.6).

In order to prove that a and d are purely imaginary, we define

$$\widetilde{\boldsymbol{M}} := \begin{pmatrix} \widetilde{a}(\eta) & \widetilde{b}(\eta) \\ \widetilde{c}(\eta) & \widetilde{d}(\eta) \end{pmatrix} := \begin{pmatrix} -\bar{a} & \frac{z}{\bar{z}}\bar{b} \\ \frac{\bar{z}}{z}\bar{c} & -\bar{d} \end{pmatrix}.$$

The goal is to conclude $\boldsymbol{M} = \widetilde{\boldsymbol{M}}$, and hence $a = -\bar{a}$ and $d = -\bar{d}$, from the uniqueness of the solution of (6.6.1) with positive definite imaginary part. Since the relations (6.6.3) are fulfilled if a, b, c, d are replaced by $\widetilde{a}$, $\widetilde{b}$, $\widetilde{c}$, $\widetilde{d}$, respectively, $\widetilde{\boldsymbol{M}}$ satisfies (6.6.1). For $j = 1, \ldots, n$, we define the 2×2 matrices

$$M_j := \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}, \qquad \widetilde{M}_j := \begin{pmatrix} \widetilde{a}_j & \widetilde{b}_j \\ \widetilde{c}_j & \widetilde{d}_j \end{pmatrix}.$$

Note that $\operatorname{Im} \boldsymbol{M}$ is positive definite if and only if $\operatorname{Im} M_j$ is positive definite for all $j = 1, \ldots, n$. Similarly, the positive definiteness of $\operatorname{Im} \widetilde{\boldsymbol{M}}$ is equivalent to the positive definiteness of $\operatorname{Im} \widetilde{M}_j$ for all $j = 1, \ldots, n$. We have

$$\operatorname{Im} M_j = \begin{pmatrix} \operatorname{Im} a_j & \frac{1}{2\mathrm{i}}(b_j - \bar{c}_j) \\ \frac{1}{2\mathrm{i}}(c_j - \bar{b}_j) & \operatorname{Im} d_j \end{pmatrix}, \quad \operatorname{Im} \widetilde{M}_j := \begin{pmatrix} \operatorname{Im} a_j & \frac{z}{2\mathrm{i}\bar{z}}(\bar{b}_j - c_j) \\ \frac{\bar{z}}{2\mathrm{i}z}(\bar{c}_j - b_j) & \operatorname{Im} d_j \end{pmatrix}.$$

As $\operatorname{tr} \operatorname{Im} \widetilde{M}_j = \operatorname{tr} \operatorname{Im} M_j$ and $\det \operatorname{Im} \widetilde{M}_j = \det \operatorname{Im} M_j$ for all $j = 1, \ldots, n$ we get that $\widetilde{\boldsymbol{M}}$ is a solution of (6.6.1) with positive definite imaginary part $\operatorname{Im} \widetilde{\boldsymbol{M}}$. Thus, the uniqueness of the solution of (6.6.1) implies $\boldsymbol{M} = \widetilde{\boldsymbol{M}}$ as well as $a = -\bar{a}$ and $d = -\bar{d}$.

Moreover, since

$$\operatorname{Im} \boldsymbol{M} = \begin{pmatrix} \operatorname{Im} a & (b - \bar{c})/(2\mathrm{i}) \\ (c - \bar{b})/(2\mathrm{i}) & \operatorname{Im} d \end{pmatrix}$$

is positive definite we have that $\operatorname{Im} a > 0$ and $\operatorname{Im} d > 0$. Hence, $(\operatorname{Im} a, \operatorname{Im} d)$ is a positive solution of (6.3.6).

Conversely, let $\boldsymbol{v} = (v_1, v_2) \in \mathbb{C}^{2n}$ be a solution of (6.3.6) satisfying $\boldsymbol{v} > 0$ and u be defined as in (6.3.25). Because of (6.3.25), we obtain that $\boldsymbol{M} = \boldsymbol{M}^z$, defined as in

(6.3.5), is a solution of (6.6.1). To conclude that $\operatorname{Im} \boldsymbol{M}$ is positive definite, it suffices to show that $\det \operatorname{Im} M_j > 0$ for all $j = 1, \dots, n$ with

$$M_j := \begin{pmatrix} \mathrm{i}(v_1)_j & -zu_j \\ -\bar{z}u_j & \mathrm{i}(v_2)_j \end{pmatrix}$$

as $\operatorname{tr} \operatorname{Im} M_j = (v_1)_j + (v_2)_j > 0$. Since $zu_j - \overline{\bar{z}u_j} = 0$ for all $j = 1, \dots, n$ by (6.3.25) we obtain

$$\det \operatorname{Im} M_j = (v_1)_j (v_2)_j - \frac{1}{4} |zu_j - \overline{\bar{z}u_j}|^2 = (v_1)_j (v_2)_j > 0.$$

Therefore, there is a bijection between the solutions of (6.6.1) with positive definite imaginary part and the positive solutions of (6.3.6). Appealing to the existence and uniqueness of (6.6.1) proven in [96] concludes the proof of Lemma 6.2.3. □

6.7. Proof of the Contraction-Inversion Lemma

Proof of Lemma 6.3.7. The bounds (6.3.55) imply that $\mathbb{1} - \boldsymbol{AB}$ is invertible and

$$\|(\mathbb{1} - \boldsymbol{AB})^{-1}\|_2 \leq \frac{1}{c_1 \eta}.$$

The main point of this lemma is to show that $(\mathbb{1} - \boldsymbol{AB})^{-1}\boldsymbol{p}$ can be bounded independently of η for $\boldsymbol{p}$ satisfying (6.3.58). We introduce $\boldsymbol{h} := (\mathbb{1} - \boldsymbol{AB})^{-1}\boldsymbol{p}$. Thus, (6.3.59) is equivalent to $\|\boldsymbol{h}\|_2 \leq C \|\boldsymbol{p}\|_2$ for some $C > 0$ which depends only on c_1, c_2, c_3 and ε. Without loss of generality, we may assume that $\|\boldsymbol{h}\|_2 = 1$. We decompose

$$\boldsymbol{h} = \alpha \boldsymbol{b}_- + \beta \boldsymbol{b}_+ + \gamma \boldsymbol{x}, \tag{6.7.1}$$

where $\alpha = \langle \boldsymbol{b}_-, \boldsymbol{h} \rangle$, $\beta = \langle \boldsymbol{b}_+, \boldsymbol{h} \rangle$ and $\boldsymbol{x} \perp \boldsymbol{b}_\pm$ satisfying $\|\boldsymbol{x}\|_2 = 1$, thus $|\alpha|^2 + |\beta|^2 + |\gamma|^2 = 1$. Since $\boldsymbol{B} = \boldsymbol{B}^*$, we have $\boldsymbol{b}_+ \perp \boldsymbol{b}_-$ and $\boldsymbol{Bx} \perp \boldsymbol{b}_\pm$. Hence, we obtain

$$\|\boldsymbol{ABh}\|_2^2 \leq \|\boldsymbol{Bh}\|_2^2 \leq |\alpha|^2 \|\boldsymbol{B}\|_2 + |\beta|^2 \|\boldsymbol{B}\|_2 + |\gamma|^2 \|\boldsymbol{Bx}\|_2^2 \leq 1 - \varepsilon + \varepsilon(|\alpha|^2 + |\beta|^2),$$

where we used $\|\boldsymbol{A}\|_2 \leq 1$, $\|\boldsymbol{B}\|_2 \leq 1$ and $\|\boldsymbol{Bx}\|_2 \leq 1 - \varepsilon$ in the last step. Therefore, if $|\alpha|^2 + |\beta|^2 \leq 1 - \delta$ for some $\delta > 0$ to be determined later, then $\|\boldsymbol{ABh}\|_2 \leq \sqrt{1 - \varepsilon\delta} \|\boldsymbol{h}\|_2 \leq (1 - \varepsilon\delta/2) \|\boldsymbol{h}\|_2$ and thus

$$1 = \|\boldsymbol{h}\|_2 \leq \frac{2}{\varepsilon\delta} \|\boldsymbol{p}\|_2. \tag{6.7.2}$$

For the rest of the proof, we assume that $|\alpha|^2 + |\beta|^2 \geq 1 - \delta$. In the regime, where $|\alpha|$ is relatively large, we compute $\langle \boldsymbol{b}_-, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{h}\rangle$, capitalize on the positivity of $\langle \boldsymbol{b}_-, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{b}_-\rangle$ and treat all other terms as errors. In the opposite regime, where $|\beta|$ is relatively large, we use the positivity of $\langle \boldsymbol{b}_+, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{b}_+\rangle$.

Using (6.7.1), we compute

$$\langle \boldsymbol{b}_-, \boldsymbol{p}\rangle = \langle \boldsymbol{b}_-, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{h}\rangle = \alpha(1 + \|\boldsymbol{B}\|_2\langle \boldsymbol{b}_-, \boldsymbol{A}\boldsymbol{b}_-\rangle) - \beta\|\boldsymbol{B}\|_2\langle \boldsymbol{b}_-, \boldsymbol{A}\boldsymbol{b}_+\rangle - \gamma\langle \boldsymbol{b}_-, \boldsymbol{AB}\boldsymbol{x}\rangle.$$

From $\|\boldsymbol{A}\|_2 \leq 1$, the Hermiticity of $\boldsymbol{A}$, $\langle \boldsymbol{b}_-, \boldsymbol{B}\boldsymbol{x}\rangle = 0$, (6.3.57) and (6.3.56), we deduce

$$\begin{aligned} |\langle \boldsymbol{b}_-, \boldsymbol{A}\boldsymbol{b}_-\rangle| &\leq 1, \\ |\langle \boldsymbol{b}_-, \boldsymbol{A}\boldsymbol{b}_+\rangle| = |\langle \boldsymbol{b}_- + \boldsymbol{A}\boldsymbol{b}_-, \boldsymbol{b}_+\rangle| &\leq c_2\eta, \\ |\langle \boldsymbol{b}_-, \boldsymbol{AB}\boldsymbol{x}\rangle| = |\langle \boldsymbol{b}_- + \boldsymbol{A}\boldsymbol{b}_-, \boldsymbol{B}\boldsymbol{x}\rangle| &\leq c_2\eta(1-\varepsilon). \end{aligned}$$

Employing these estimates, $\|\boldsymbol{B}\|_2 \leq 1 - c_1\eta$ and (6.3.58), together with $|\gamma|^2 \leq \delta$, we obtain

$$c_3\|\boldsymbol{p}\|_2 \geq |\alpha|c_1 - |\beta|c_2 - \sqrt{\delta}c_2(1-\varepsilon) \tag{6.7.3}$$

after dividing through by $\eta > 0$. If $|\alpha|c_1 \geq c_2|\beta| + \sqrt{\delta}c_2(1-\varepsilon) + \delta\varepsilon c_3/2$ then we obtain (6.7.2).

Therefore, it suffices to show (6.7.2) in the regime

$$|\gamma|^2 \leq \delta, \qquad |\alpha|c_1 \leq c_2|\beta| + \sqrt{\delta}c_2(1-\varepsilon) + \delta\varepsilon c_3/2. \tag{6.7.4}$$

For this regime, we use (6.7.1) and obtain

$$\begin{aligned} \langle \boldsymbol{b}_+, \boldsymbol{p}\rangle &= \langle \boldsymbol{b}_+, (\mathbb{1} - \boldsymbol{AB})\boldsymbol{h}\rangle \\ &= \beta(1 - \|\boldsymbol{B}\|_2\langle \boldsymbol{b}_+, \boldsymbol{A}\boldsymbol{b}_+\rangle) - \alpha\|\boldsymbol{B}\|_2\langle \boldsymbol{b}_+, \boldsymbol{A}\boldsymbol{b}_-\rangle - \gamma\langle \boldsymbol{b}_+, \boldsymbol{AB}\boldsymbol{x}\rangle. \end{aligned} \tag{6.7.5}$$

We employ (6.3.56), (6.3.57), the Hermiticity of $\boldsymbol{A}$ and $\langle \boldsymbol{b}_-, \boldsymbol{b}_+\rangle = 0$ to obtain

$$\begin{aligned} \langle \boldsymbol{b}_+, \boldsymbol{A}\boldsymbol{b}_+\rangle &\leq 1-\varepsilon, \\ |\langle \boldsymbol{b}_+, \boldsymbol{A}\boldsymbol{b}_-\rangle| = |\langle \boldsymbol{b}_+, \boldsymbol{b}_- + \boldsymbol{A}\boldsymbol{b}_-\rangle| &\leq c_2\eta, \\ |\langle \boldsymbol{b}_+, \boldsymbol{AB}\boldsymbol{x}\rangle| &\leq 1-\varepsilon. \end{aligned} \tag{6.7.6}$$

Applying (6.7.6) to (6.7.5), yields

$$\|\boldsymbol{p}\|_2 \geq |\langle \boldsymbol{b}_+ , \boldsymbol{p}\rangle| \geq |\beta|\varepsilon - |\alpha| c_2 \eta - |\gamma|(1-\varepsilon) \geq |\beta|\varepsilon - |\alpha|\frac{\varepsilon c_1}{2c_2} - \sqrt{\delta}(1-\varepsilon), \qquad (6.7.7)$$

where we used the assumption $\eta \leq \varepsilon c_1/2c_2^2$. Since $|\alpha| c_1/c_2 \leq |\beta| + \mathcal{O}(\sqrt{\delta})$ from (6.7.4), we obtain that $\|\boldsymbol{p}\|_2 \geq |\beta|\varepsilon/3$ for any $\delta \leq \delta_0(c_1, c_2, c_3, \varepsilon)$ sufficiently small. Furthermore, $|\alpha|^2 + |\beta|^2 \geq 1 - \delta$ and the fact that $|\beta|$ is large compared with $|\alpha|$ in the sense (6.7.4) guarantee that $|\beta|^2 \geq \frac{1}{3}[1+(c_2/c_1)^2]^{-1}$, if δ is sufficiently small. In particular, $\|\boldsymbol{p}\|_2 \geq \varepsilon\delta/2$ can be achieved with a small δ, i.e., (6.7.2) holds true in the regime (6.7.4) as well. This concludes the proof of Lemma 6.3.7. □

Lemma 6.7.1. *(i) Uniformly for $z \in \mathbb{D}_< \cup \mathbb{D}_>$ and $\eta > 0$, we have*

$$\|\boldsymbol{F}\|_{2\to\infty} \lesssim 1, \quad \|\boldsymbol{T}\boldsymbol{F}\|_{2\to\infty} \lesssim 1, \quad \|\boldsymbol{F}\boldsymbol{T}\|_{2\to\infty} \lesssim 1. \qquad (6.7.8)$$

(ii) If $\zeta \notin \mathrm{Spec}(\boldsymbol{T}\boldsymbol{F}) \cup \{0\}$ and $\|(\zeta\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{y}\|_2 \lesssim \|\boldsymbol{y}\|_2$ for some $\boldsymbol{y} \in \mathbb{C}^{2n}$ then

$$\|(\zeta\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{y}\|_\infty \lesssim \frac{1}{|\zeta|}\|\boldsymbol{y}\|_\infty. \qquad (6.7.9)$$

A similar statement holds true for $(\bar{\zeta}\mathbb{1} - \boldsymbol{F}\boldsymbol{T})^{-1} = [(\zeta\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}]^*$.

(iii) For every $\eta_* > 0$, depending only on τ_* and the model parameters, such that

$$\begin{aligned} \|(\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q}\|_2 &\lesssim 1, & 1 - \|\boldsymbol{F}\|_2 &\gtrsim \eta, \\ \|\boldsymbol{f}_- + \boldsymbol{T}\boldsymbol{f}_-\|_2 &\lesssim \eta, & \|\boldsymbol{f}_-\|_\infty &\lesssim 1 \end{aligned} \qquad (6.7.10)$$

uniformly for all $\eta \in (0, \eta_*]$ and $z \in \mathbb{D}_<$, we have

$$\|\big((\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q}\big)^*\|_\infty \lesssim 1 \qquad (6.7.11)$$

uniformly for $\eta \in (0, \eta_*]$ and $z \in \mathbb{D}_<$. Here, $\boldsymbol{Q}$ denotes the orthogonal projection onto the subspace $\boldsymbol{f}_-^\perp$, i.e., $\boldsymbol{Q}\boldsymbol{y} := \boldsymbol{y} - \langle \boldsymbol{f}_- , \boldsymbol{y}\rangle \boldsymbol{f}_-$ for every $\boldsymbol{y} \in \mathbb{C}^{2n}$.

The estimate (6.7.9) is proven similarly as (5.28) in [4].

Proof. As $\|\boldsymbol{S}_o\|_{2\to\infty} \lesssim 1$ by (6.2.1), we obtain from Proposition 6.3.2 and (6.3.26) the bound

$$\|\boldsymbol{F}\|_{2\to\infty} \le \|\boldsymbol{V}^{-1}\|_\infty \|\boldsymbol{S}_o\|_{2\to\infty} \|\boldsymbol{V}^{-1}\|_2 = \left\| \frac{\boldsymbol{u}\boldsymbol{v}}{\widetilde{\boldsymbol{v}}} \right\|_\infty \|\boldsymbol{S}_o\|_{2\to\infty} \lesssim 1$$

uniformly for all $\eta > 0$ and $z \in \mathbb{D}_< \cup \mathbb{D}_>$. This proves the first estimate in (6.7.8). From Lemma 6.3.6 (i), we conclude the second and the third estimate in (6.7.8).

We set $\boldsymbol{x} \coloneqq (\zeta \mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{y}$. By assumption there is $C \sim 1$ such that

$$\|\boldsymbol{x}\|_2 \le C\|\boldsymbol{y}\|_2 \le C\|\boldsymbol{y}\|_\infty.$$

Moreover, since $\zeta\boldsymbol{x} = \boldsymbol{T}\boldsymbol{F}\boldsymbol{x} + \boldsymbol{y}$ we obtain from the previous estimate

$$|\zeta|\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{T}\boldsymbol{F}\boldsymbol{x}\|_\infty + \|\boldsymbol{y}\|_\infty \le (\|\boldsymbol{T}\boldsymbol{F}\|_{2\to\infty} C + 1)\,\|\boldsymbol{y}\|_\infty.$$

Using the second estimate in (6.7.8), this concludes the proof of (6.7.9). The statement about $(\bar{\zeta}\mathbb{1} - \boldsymbol{F}\boldsymbol{T})^{-1}$ follows in the same way using the third estimate in (6.7.8) instead of the second.

For the proof of (6.7.11), we remark that the first condition in (6.7.10) implies that

$$\left\| \left((\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q} \right)^* \right\|_2 = \left\| (\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q} \right\|_2 \lesssim 1. \tag{6.7.12}$$

The second assumption in (6.7.10) yields

$$\left\| (\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1} \right\|_2 \lesssim \eta^{-1}. \tag{6.7.13}$$

Take $\boldsymbol{y} \in \mathbb{C}^{2n}$ arbitrary. We get $[\boldsymbol{T}, \boldsymbol{Q}]\boldsymbol{y} = \langle \boldsymbol{T}\boldsymbol{f}_- + \boldsymbol{f}_-, \boldsymbol{y}\rangle \boldsymbol{f}_- - \langle \boldsymbol{f}_-, \boldsymbol{y}\rangle(\boldsymbol{T}\boldsymbol{f}_- + \boldsymbol{f}_-)$, where $[\boldsymbol{T}, \boldsymbol{Q}] = \boldsymbol{T}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{T}$ denotes the commutator of $\boldsymbol{T}$ and $\boldsymbol{Q}$. Therefore,

$$\|[\boldsymbol{T}, \boldsymbol{Q}]\|_2 \le 2\|\boldsymbol{f}_- + \boldsymbol{T}\boldsymbol{f}_-\|_2 \lesssim \eta \tag{6.7.14}$$

by the third condition in (6.7.10). We set $\boldsymbol{x} \coloneqq \boldsymbol{Q}(\mathbb{1} - \boldsymbol{F}\boldsymbol{T})^{-1}\boldsymbol{y} = ((\mathbb{1} - \boldsymbol{T}\boldsymbol{F})^{-1}\boldsymbol{Q})^*\,\boldsymbol{y}$ and compute

$$\boldsymbol{x} = \boldsymbol{F}\boldsymbol{T}\boldsymbol{x} + \boldsymbol{Q}\boldsymbol{y} - \boldsymbol{F}[\boldsymbol{T}, \boldsymbol{Q}](\mathbb{1} - \boldsymbol{F}\boldsymbol{T})^{-1}\boldsymbol{y},$$

where we commuted $\mathbb{1} - \boldsymbol{FT}$ and $\boldsymbol{Q}$ and used that $\boldsymbol{F}$ and $\boldsymbol{Q}$ commute. Hence, using $\|\boldsymbol{x}\|_2 \lesssim \|\boldsymbol{y}\|_2 \lesssim \|\boldsymbol{y}\|_\infty$ by (6.7.12) , $\|\boldsymbol{Q}\|_\infty \leq 1 + \|\boldsymbol{f}_-\|_\infty$, (6.7.14) and (6.7.13), we obtain

$$\|\boldsymbol{x}\|_\infty \lesssim \Big(\|\boldsymbol{FT}\|_{2\to\infty} + 1 + \|\boldsymbol{f}_-\|_\infty + \|\boldsymbol{F}\|_{2\to\infty}\Big)\|\boldsymbol{y}\|_\infty \lesssim \|\boldsymbol{y}\|_\infty.$$

Here, we used the fourth assumption in (6.7.10) and (6.7.8). Notice that the η^{-1} factor from the trivial estimate (6.7.13) was compensated by the smallness of the commutator $[\boldsymbol{T}, \boldsymbol{Q}]$ which was a consequence of the third assumption in (6.7.10). This concludes the proof of (6.7.11). □

Proof of Lemma 6.3.5. We first prove that

$$\|\boldsymbol{f}_- - \boldsymbol{a}\|_2 = \mathcal{O}(\eta). \tag{6.7.15}$$

uniformly for $\eta \leq 1$ and $\tau \in [0, 1-\tau_*]$. To that end, we introduce the auxiliary operator

$$\boldsymbol{A} := \|\boldsymbol{F}\|_2 \mathbb{1} + \boldsymbol{F}.$$

Therefore, we obtain from $\boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2 \boldsymbol{f}_-$ and (6.3.45)

$$\boldsymbol{A}\boldsymbol{f}_- = 0, \quad \boldsymbol{A}\boldsymbol{a} = \mathcal{O}(\eta).$$

Let $\boldsymbol{Q}$ be the orthogonal projection onto the subspace $\boldsymbol{f}_-^\perp$ orthogonal to $\boldsymbol{f}_-$, i.e., $\boldsymbol{Q}\boldsymbol{y} := \boldsymbol{y} - \langle \boldsymbol{f}_-, \boldsymbol{y}\rangle \boldsymbol{f}_-$ for $\boldsymbol{y} \in \mathbb{C}^{2n}$. We then obtain $\boldsymbol{AQa} = \mathcal{O}(\eta)$ which implies $\boldsymbol{Qa} = \mathcal{O}(\eta)$ as $\boldsymbol{A}$ is invertible on $\boldsymbol{f}_-^\perp$ and $\|(\boldsymbol{A}|_{\boldsymbol{f}_-^\perp})^{-1}\|_2 \sim 1$ by (6.3.38). We infer (6.7.15).

For the proof of (6.3.46), we follow the proof of (6.7.11), replace $\boldsymbol{T}$ by $-\mathbb{1}$ and use Lemma 6.3.4 (i) instead of the second and fourth condition in (6.7.10). □

CHAPTER 7

Location of the spectrum of Kronecker random matrices

In this chapter, we present the results of the publication [16] which was prepared in joint work with László Erdős, Torben Krüger and Yuriy Nemish. For a general class of large non-Hermitian random block matrices $\boldsymbol{X}$ we prove that there are no eigenvalues away from a deterministic set with very high probability. This set is obtained from the Dyson equation of the Hermitization of $\boldsymbol{X}$ as the self-consistent approximation of the pseudospectrum. We demonstrate that the analysis of the matrix Dyson equation from [6] offers a unified treatment of many structured matrix ensembles.

7.1. Introduction

Large random matrices tend to exhibit deterministic patterns due to the cumulative effects of many independent random degrees of freedom. The Wigner semicircle law [157] describes the deterministic limit of the empirical density of eigenvalues of Wigner matrices, i.e., *Hermitian* random matrices with i.i.d. entries (modulo the Hermitian symmetry). For *non-Hermitian* matrices with i.i.d. entries, the limiting density is Girko's circular law, i.e., the uniform distribution in a disk centered around zero in the complex plane, see [40] for a review.

For more complicated ensembles, no simple formula exists for the limiting behavior, but second order perturbation theory predicts that it may be obtained from the solution to a nonlinear equation, called the *Dyson equation.* While simplified forms of the Dyson equation are present in practically every work on random matrices, its full scope has only recently been analyzed systematically, see [6]. In fact, the proper Dyson equation describes not only the density of states but the entire resolvent of the random matrix. Treating it as a genuine *matrix equation* unifies many previous works that were specific to certain structures imposed on the random matrix. These additional structures often masked a fundamental property of the Dyson equation, its stability against small perturbations, that plays a key role in proving the expected limit theorems, also called *global laws*. Girko's monograph [82] is the most systematic collection of many possible ensembles, yet it analyzes them on a case by case basis.

In this paper, using the setup of the *matrix Dyson equation (MDE)* from [6], we demonstrate a unified treatment for a large class of random matrix ensembles that contain or generalize many of Girko's models. For brevity, we focus only on two basic problems: (i) obtaining the global law and (ii) locating the spectrum. The global law, typically formulated as a weak convergence of linear statistics of the eigenvalues, describes only the overwhelming majority of the eigenvalues. Even local versions of this limit theorem, commonly called *local laws* (see e.g. [44, 60], Chapter 6 and references therein) are typically not sensitive to individual eigenvalues and they do not exclude that a few eigenvalues are located far away from the support of the density of states.

Extreme eigenvalues have nevertheless been controlled in some simple cases. In particular, for the i.i.d. cases, it is known that with a very high probability all eigenvalues lie in an ε-neighborhood of the support of the density of states. These results can be proven with the moment method, see [19, Theorem 2.1.22] for the Hermitian (Wigner) case, and [80] for the non-Hermitian i.i.d. case; see also [24, 25] for the optimal moment condition. More generally, norms of polynomials in large independent random matrices can be computed via free probability; for GUE or GOE Gaussian matrices it was achieved in [87] and generalized to polynomials of general Wigner and Wishart type matrices in [18, 47]. These results have been extended recently to polynomials that include deterministic matrices with the goal of studying outliers, see [31] and references therein.

All these works concern Hermitian matrices either directly or indirectly by considering quantities, such as norms of non-Hermitian polynomials, that can be deduced from related Hermitian problems. For general Hermitian random matrices, the density of states may be supported on several intervals. In this situation, excluding eigenvalues outside of the convex hull of this support is typically easier than excluding possible eigenvalues lying inside the gaps of the support. This latter problem, however, is especially important for studying the spectrum of non-Hermitian random matrices $\boldsymbol{X}$, since the eigenvalues of $\boldsymbol{X}$ around a complex parameter ζ can be understood by studying the spectrum of the

Hermitized matrix

$$\boldsymbol{H}^{\zeta} = \begin{pmatrix} 0 & \boldsymbol{X} - \zeta \\ \boldsymbol{X}^* - \bar{\zeta} & 0 \end{pmatrix} \tag{7.1.1}$$

around 0. Note that for $\zeta \in \mathbb{C}$ away from the spectrum of $\boldsymbol{X}$, zero will typically fall inside a gap of the spectrum of $\boldsymbol{H}^{\zeta}$ by its symmetry.

In this paper, we consider a very general class of structured block matrices $\boldsymbol{X}$ that we call *Kronecker random matrices* since their structure is reminiscent to the Kronecker product of matrices. They have $L \times L$ large blocks and each block consists of a linear combination of random $N \times N$ matrices with centered, independent, not necessarily identically distributed entries; see (7.2.1) later for the precise definition. We will keep L fixed and let N tend to infinity. The matrix $\boldsymbol{X}$ has a correlation structure that stems from allowing the same $N \times N$ matrix to appear in different blocks. This introduces an arbitrary linear dependence among the blocks, while keeping independence inside the blocks. The dependence is thus described by $L \times L$ deterministic *structure matrices*.

Kronecker random ensembles occur in many real-world applications of random matrix theory, especially in evolution of ecosystems [93] and neural networks [123]. These evolutions are described by a large system of ODE's with random coefficients and the spectral radius of the coefficient matrix determines the long time stability, see [113] for the original idea. More recent results are found in [2, 9, 10] and references therein. The ensemble we study here is even more general as it allows for linear dependence among the blocks described by arbitrary structure matrices. This level of generality is essential for another application; to study spectral properties of polynomials of random matrices. These are often studied via the "linearization trick" and the linearized matrix is exactly a Kronecker random matrix. This application is presented in [61], where the results of the current paper are directly used.

We present general results that exclude eigenvalues of Kronecker random matrices away from a deterministic set $\mathbb{D}$ with a very high probability. The set $\mathbb{D}$ is determined by solving the self-consistent Dyson equation. In the Hermitian case, $\mathbb{D}$ is the *self-consistent spectrum* defined as the support of the *self-consistent density of states* ρ which is defined as the imaginary part of the solution to the Dyson equation when restricted to the real line. We also address the general non-Hermitian setup, where the eigenvalues are not confined to the real line. In this case, the set $\mathbb{D} = \mathbb{D}_\varepsilon$ contains an additional cutoff parameter ε and it is the *self-consistent* ε*-pseudospectrum*, given via the Dyson equation for the Hermitized problem $\boldsymbol{H}^\zeta$, see (7.2.7) later. The $\varepsilon \to 0$ limit of the sets $\mathbb{D}_\varepsilon$ is expected not only to contain but to coincide with the support of the density of states in the non-Hermitian case as well, but this has been proven only in some special cases. We provide numerical examples to support this conjecture.

We point out that the global law and the location of the spectrum for $A + X$, where X is an i.i.d. centered random matrix and A is a general deterministic matrix (so-called *deformed ensembles*), have been extensively studied, see [26, 38, 39, 139, 140, 143]. For more references, we refer to the review [40]. In contrast to these papers, the main focus of our work is to allow for general (not necessarily identical) distributions of the matrix elements.

In this paper, we first study arbitrary Hermitian Kronecker matrices $\boldsymbol{H}$; the Hermitization $\boldsymbol{H}^\zeta$ of a general Kronecker matrix is itself a Kronecker matrix and therefore just a special case. Our first result is the global law, i.e., we show that the empirical density of states of $\boldsymbol{H}$ is asymptotically given by the self-consistent density of states ρ determined by the Dyson equation. We then also prove an optimal local law for spectral parameters away from the instabilities of the Dyson equation. The Dyson equation for Kronecker matrices is a system of $2N$ nonlinear equations for $L \times L$ matrices, see (7.2.6) later. In case of identical distribution of the entries within each $N \times N$ matrix, the system reduces to a single equation for a $2L \times 2L$ matrix – a computationally feasible problem. This analysis provides not only the limiting density of states but also a full understanding of the resolvent for spectral parameters z very close to the real line, down to scales $\operatorname{Im} z \gg 1/N$. Although the optimal local law down to scales $\operatorname{Im} z \gg 1/N$ cannot capture individual eigenvalues inside the support of ρ, the key point is that outside of this support a stronger estimate in the local law may be proven that actually detects individual eigenvalues, or rather lack thereof. This observation has been used for simpler models before, in particular [60, Theorem 2.3] already contained this stronger form of the local semicircle law for generalized Wigner matrices, see also [7] for Wigner-type matrices, Chapter 4 for Gram matrices (cf. [14]) and [56] for correlated matrices with a uniform lower bound on the variances. In particular, by running the stability analysis twice, this allows for an extension of the local law for any $\operatorname{Im} z > 0$ outside of the support of ρ.

Finally, applying the local law to the Hermitization $\boldsymbol{H}^\zeta$ of a non-Hermitian Kronecker matrix $\boldsymbol{X}$, we translate local spectral information on $\boldsymbol{H}^\zeta$ around 0 into information about the location of the spectrum of $\boldsymbol{X}$. This is possible since $\zeta \in \operatorname{Spec}(\boldsymbol{X})$ if and only if $0 \in \operatorname{Spec}(\boldsymbol{H}^\zeta)$. In practice, we give a good approximation to the ε-pseudospectrum of $\boldsymbol{X}$ by considering the set of those ζ values in $\mathbb{C}$ for which 0 is at least ε distance away from the support of the self-consistent density of states for $\boldsymbol{H}^\zeta$.

In the main part of the paper, we give a short, self-contained proof that directly aims at locating the Hermitian spectrum under the weakest conditions for the most general setup. We split the proof into two well-separated parts; a random and a deterministic one. In Section 7.4 and 7.5 as well as Section 7.8 below we give a model-independent probabilistic proof of the main technical result, the local law (Theorem 7.4.7 and Lemma 7.8.1), assuming only two explicit conditions, boundedness and stability, on the solution of the Dyson equation that can be checked separately for concrete models. In Section 7.3.2 we prove that these two conditions are satisfied for Kronecker matrices away from the self-consistent spectrum. The key inputs behind the stability are (i) a matrix version of the Perron-Frobenius theorem and (ii) a sophisticated symmetrization procedure that is much more transparent in the matrix formulation. In particular, the global law is an immediate consequence of this approach. Moreover, the analysis reveals that outside of the spectrum the stability holds without any lower bound on the variances, in contrast to local laws inside the bulk spectrum that typically require some non-degeneracy condition on the matrix of variances.

We stress that only the first part involves randomness and we follow the Schur complement method and concentration estimates for linear and quadratic functionals of independent random variables. Alternatively, we could have used the cumulant expansion method that is typically better suited for ensembles with correlation [56]. We opted for the former path to demonstrate that correlations stemming from the block structure can still be handled with the more direct Schur complement method as long as the noncommutativity of the $L \times L$ structure matrices is properly taken into account. Utilizing a powerful *tensor matrix structure* generated by the correlations between blocks resolves this issue automatically.

Acknowledgement. The authors are grateful to David Renfrew for several discussions and for calling their attention to references on applications of non-Hermitian models.

7.1.1. Notation. Owing to the tensor product structure of Kronecker random matrices (see Definition 7.2.1 below), we need to introduce different spaces of matrices. In order to make the notation more transparent to the reader, we collect the conventions used on these spaces in this subsection.

For $K, N \in \mathbb{N}$, we will consider the spaces $\mathbb{C}^{K\times K}$, $(\mathbb{C}^{K\times K})^N$ and $\mathbb{C}^{K\times K} \otimes \mathbb{C}^{N\times N}$, i.e., we consider $K \times K$ matrices, N-vectors of $K \times K$ matrices and $N \times N$ matrices with $K \times K$ matrices as entries. For brevity, we denote $\mathcal{M} := \mathbb{C}^{K\times K} \otimes \mathbb{C}^{N\times N}$. Elements of $\mathbb{C}^{K\times K}$ are usually denoted by small roman letters, elements of $(\mathbb{C}^{K\times K})^N$ by small boldface roman letters and elements of $\mathcal{M}$ by capitalized boldface roman letters.

For $\alpha \in \mathbb{C}^{K\times K}$, we denote by $|\alpha|$ the matrix norm of α induced by the Euclidean distance on $\mathbb{C}^K$. Moreover, we define two different norms on the N-vectors of $K \times K$ matrices. For any $\boldsymbol{r} = (r_1, \ldots, r_N) \in (\mathbb{C}^{K\times K})^N$ we define $\|\boldsymbol{r}\| := \max_{i=1}^N |r_i|$, and

$$\|\boldsymbol{r}\|_{\text{hs}}^2 := \frac{1}{NK} \sum_{i=1}^N \operatorname{Tr}(r_i^* r_i). \tag{7.1.2}$$

These are the analogues of the maximum norm and the Euclidean norm for vectors in $\mathbb{C}^N$ which corresponds to $K = 1$. Note that $\|\boldsymbol{r}\|_{\text{hs}} \leq \|\boldsymbol{r}\|$.

For any function $f : U \to \mathbb{C}^{K\times K}$ from $U \subset \mathbb{C}^{K\times K}$ to $\mathbb{C}^{K\times K}$, we lift f to U^N by defining $f(\boldsymbol{r}) \in (\mathbb{C}^{K\times K})^N$ entrywise for any $\boldsymbol{r} = (r_1, \ldots, r_N) \in U^N \subset (\mathbb{C}^{K\times K})^N$, i.e.,

$$f(\boldsymbol{r}) := (f(r_1), \ldots, f(r_N)). \tag{7.1.3}$$

We will in particular apply this definition for f being the matrix inversion map and the imaginary part. Moreover, for $\boldsymbol{x} = (x_1, \ldots, x_N)$, $\boldsymbol{y} = (y_1, \ldots, y_N) \in (\mathbb{C}^{K\times K})^N$ we introduce their entrywise product $\boldsymbol{xy} \in (\mathbb{C}^{K\times K})^N$ through

$$\boldsymbol{xy} := (x_1 y_1, \ldots, x_N y_N) \in (\mathbb{C}^{K\times K})^N. \tag{7.1.4}$$

Note that for $K \neq 1$, in general, $\boldsymbol{xy} \neq \boldsymbol{yx}$.

If $a \in \mathbb{C}^{K\times K}$ or $\boldsymbol{A} \in \mathcal{M}$ are positive semidefinite matrices, then we write $a \geq 0$ or $\boldsymbol{A} \geq 0$, respectively. Similarly, for $\boldsymbol{a} \in (\mathbb{C}^{K\times K})^N$, we write $\boldsymbol{a} \geq 0$ to indicate that all components of $\boldsymbol{a}$ are positive semidefinite matrices in $\mathbb{C}^{K\times K}$. The identity matrix in $\mathbb{C}^{K\times K}$ and $\mathcal{M}$ is denoted by $\mathbb{1}$.

We also use two norms on $\mathcal{M}$. These are the operator norm $\|\cdot\|_2$ induced by the Euclidean distance on $\mathbb{C}^{KN} \cong \mathbb{C}^K \otimes \mathbb{C}^N$ and the norm $\|\cdot\|_{\mathrm{hs}}$ induced by the scalar product $\langle\,\cdot\,,\,\cdot\,\rangle$ on $\mathcal{M}$ defined through

$$\langle \boldsymbol{R}, \boldsymbol{T} \rangle := \frac{1}{NK}\,\mathrm{Tr}\,(\boldsymbol{R}^*\boldsymbol{T})\,, \qquad \|\boldsymbol{R}\|_{\mathrm{hs}} := \sqrt{\langle \boldsymbol{R}, \boldsymbol{R} \rangle}, \tag{7.1.5}$$

for $\boldsymbol{R}, \boldsymbol{T} \in \mathcal{M}$. In particular, all orthogonality statements on $\mathcal{M}$ are understood with respect to this scalar product. Furthermore, we introduce $\langle \boldsymbol{R} \rangle := \langle \mathbb{1}\,, \boldsymbol{R} \rangle$, the normalized trace for $\boldsymbol{R} \in \mathcal{M}$.

We also consider linear maps on $(\mathbb{C}^{K\times K})^N$ and $\mathcal{M}$, respectively. We follow the convention that the symbols $\mathscr{S}$, $\mathscr{L}$ and $\mathscr{T}$ label linear maps $(\mathbb{C}^{K\times K})^N \to (\mathbb{C}^{K\times K})^N$ and $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{L}}$ or $\boldsymbol{\mathcal{T}}$ denote linear maps $\mathcal{M} \to \mathcal{M}$. The symbol Id refers to the identity map on $\mathcal{M}$. For any linear map $\mathscr{T} \colon (\mathbb{C}^{K\times K})^N \to (\mathbb{C}^{K\times K})^N$, let $\|\mathscr{T}\|$ denote the operator norm of $\mathscr{T}$ induced by $\|\cdot\|$ and let $\|\mathscr{T}\|_{\mathrm{sp}}$ denote the operator norm induced by $\|\cdot\|_{\mathrm{hs}}$. Similarly, for a linear map $\boldsymbol{\mathcal{T}} \colon \mathcal{M} \to \mathcal{M}$, we write $\|\boldsymbol{\mathcal{T}}\|$ for the operator norm induced by $\|\cdot\|_2$ on $\mathcal{M}$ and $\|\boldsymbol{\mathcal{T}}\|_{\mathrm{sp}}$ for its operator norm induced by $\|\cdot\|_{\mathrm{hs}}$ on $\mathcal{M}$.

We use the notation $[n] := \{1, \ldots, n\}$ for $n \in \mathbb{N}$. For $i, j \in [N]$, we introduce the matrix $E_{ij} \in \mathbb{C}^{N\times N}$ which has a one at its (i, j) entry and only zeros otherwise, i.e.,

$$E_{ij} := (\delta_{ik}\delta_{jl})_{k,l=1}^N. \tag{7.1.6}$$

For $i, j \in [N]$, the linear map $P_{ij} \colon \mathcal{M} \to \mathbb{C}^{K\times K}$ is defined through

$$P_{ij}\boldsymbol{R} = r_{ij}, \tag{7.1.7}$$

for any $\boldsymbol{R} = \sum_{i,j=1}^N r_{ij} \otimes E_{ij} \in \mathcal{M}$ with $r_{ij} \in \mathbb{C}^{K\times K}$.

7.2. Main results

Our main object of study are Kronecker random matrices which we define first. To that end, we recall the definition of E_{ij} from (7.1.6).

Definition 7.2.1 (Kronecker random matrix)**.** A random matrix $\boldsymbol{X} \in \mathbb{C}^{L\times L} \otimes \mathbb{C}^{N\times N}$ is called *Kronecker random matrix* if it is of the form

$$\boldsymbol{X} = \sum_{\mu=1}^{\ell} \widetilde{\alpha}_\mu \otimes X_\mu + \sum_{\nu=1}^{\ell} (\widetilde{\beta}_\nu \otimes Y_\nu + \widetilde{\gamma}_\nu \otimes Y_\nu^*) + \sum_{i=1}^{N} \widetilde{a}_i \otimes E_{ii}, \qquad \ell \in \mathbb{N}, \tag{7.2.1}$$

where $X_\mu = X_\mu^* \in \mathbb{C}^{N\times N}$ are Hermitian random matrices with centered independent entries (up to the Hermitian symmetry) and $Y_\nu \in \mathbb{C}^{N\times N}$ are random matrices with centered independent entries; furthermore $X_1, \ldots, X_\ell, Y_1, \ldots, Y_\ell$ are independent. The "coefficient" matrices $\widetilde{\alpha}_\mu, \widetilde{\beta}_\nu, \widetilde{\gamma}_\nu \in \mathbb{C}^{L\times L}$ are deterministic and they are called *structure matrices*. Finally, $\widetilde{a}_1, \ldots, \widetilde{a}_N \in \mathbb{C}^{L\times L}$ are also deterministic.

We remark that the number of X_μ and Y_ν matrices effectively present in $\boldsymbol{X}$ may differ by choosing some structure matrices zero. Furthermore, note that $\mathbb{E}\boldsymbol{X} = \sum_{i=1}^N \widetilde{a}_i \otimes E_{ii}$, i.e., the deterministic matrices $\widetilde{a}_i$ encode the expectation of $\boldsymbol{X}$.

Our main result asserts that all eigenvalues of a Kronecker random matrix $\boldsymbol{X}$ are contained in the *self-consistent ε-pseudospectrum* for any $\varepsilon > 0$, with a very high probability if N is sufficiently large. The self-consistent ε-pseudospectrum, $\mathbb{D}_\varepsilon$, is a deterministic subset of the complex plane that can be defined and computed via the self-consistent solution to the *Hermitized Dyson equation.* Hermitization entails doubling the dimension and studying the matrix $\boldsymbol{H}^\zeta$ defined in (7.1.1) for any spectral parameter $\zeta \in \mathbb{C}$ associated with $\boldsymbol{X}$. We introduce an additional spectral parameter $z \in \mathbb{H} := \{w \in \mathbb{C} \colon \operatorname{Im} w > 0\}$ that will be associated with the Hermitian matrix $\boldsymbol{H}^\zeta$. The Hermitized Dyson equation is used to study the resolvent $(\boldsymbol{H}^\zeta - z\mathbb{1})^{-1}$.

We first introduce some notation necessary to write up the Hermitized Dyson equation. For $\mu, \nu \in [\ell]$, we define

$$\alpha_\mu := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \widetilde{\alpha}_\mu + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \widetilde{\alpha}_\mu^*, \qquad \beta_\nu := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes (\widetilde{\beta}_\nu + \widetilde{\gamma}_\nu^*). \tag{7.2.2}$$

We set

$$s_{ij}^{\mu} := \mathbb{E}\,|x_{ij}^{\mu}|^2, \qquad t_{ij}^{\nu} := \mathbb{E}\,|y_{ij}^{\nu}|^2, \tag{7.2.3}$$

where x_{ij}^{μ} and y_{ij}^{ν} are the (scalar) entries of the random matrices X_μ and Y_ν, respectively, i.e., $X_\mu = (x_{ij}^{\mu})_{i,j=1}^N$ and $Y_\nu = (y_{ij}^{\nu})_{i,j=1}^N$. We define a linear map $\mathscr{S}$ on $(\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L})^N$, i.e., on N-vectors of $(2L)\times(2L)$ matrices as follows. For any $\boldsymbol{r} = (r_1, \dots, r_N) \in (\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L})^N$ we set

$$\mathscr{S}[\boldsymbol{r}] = (\mathscr{S}_1[\boldsymbol{r}], \mathscr{S}_2[\boldsymbol{r}], \dots, \mathscr{S}_N[\boldsymbol{r}]) \in (\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L})^N,$$

where the i-th component is given by

$$\mathscr{S}_i[\boldsymbol{r}] := \sum_{k=1}^{N} \left(\sum_{\mu=1}^{\ell} s_{ik}^{\mu} \alpha_\mu r_k \alpha_\mu + \sum_{\nu=1}^{\ell} (t_{ik}^{\nu} \beta_\nu r_k \beta_\nu^* + t_{ki}^{\nu} \beta_\nu^* r_k \beta_\nu) \right) \in \mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L}, \qquad i \in [N]. \tag{7.2.4}$$

For $j \in [N]$ and $\zeta \in \mathbb{C}$, we define $a_j^{\zeta} \in \mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L}$ through

$$a_j^{\zeta} := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \widetilde{a}_j + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \widetilde{a}_j^* - \begin{pmatrix} 0 & \zeta \\ \bar{\zeta} & 0 \end{pmatrix} \otimes \mathbb{1}. \tag{7.2.5}$$

The Hermitized Dyson equation is the following system of equations

$$-\frac{1}{m_j^{\zeta}(z)} = z\mathbb{1} - a_j^{\zeta} + \mathscr{S}_j[\boldsymbol{m}^{\zeta}(z)], \qquad j = 1, 2, \dots N, \tag{7.2.6}$$

for the vector

$$\boldsymbol{m}^{\zeta}(z) = (m_1^{\zeta}(z), \dots, m_N^{\zeta}(z)) \in (\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L})^N.$$

Here, $\mathbb{1}$ denotes the identity matrix in $\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L}$ and $\zeta \in \mathbb{C}$ as well as $z \in \mathbb{H}$ are spectral parameters associated to $\boldsymbol{X}$ and $\boldsymbol{H}^{\zeta}$, respectively.

Lemma 7.2.2. *For any $z \in \mathbb{H}$ and $\zeta \in \mathbb{C}$ there exists a unique solution to* (7.2.6) *with the additional condition that the matrices $\operatorname{Im} m_j^{\zeta}(z) := \frac{1}{2i}(m_j^{\zeta}(z) - m_j^{\zeta}(z)^*)$ are positive definite for all $j \in [N]$. Moreover, for $j \in [N]$, there are measures v_j^{ζ} on $\mathbb{R}$ with values in the positive semidefinite matrices in $\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L}$ such that*

$$m_j^{\zeta}(z) = \int_{\mathbb{R}} \frac{v_j^{\zeta}(\mathrm{d}\tau)}{\tau - z}$$

for all $z \in \mathbb{H}$ and $\zeta \in \mathbb{C}$.

Lemma 7.2.2 is proven after Proposition 7.3.10 below. Throughout the paper $\boldsymbol{m}^\zeta$ will always denote the unique solution to the Hermitized Dyson equation defined in Lemma 7.2.2. The *self-consistent density of states* ρ^ζ of $\boldsymbol{H}^\zeta$ is given by

$$\rho^\zeta(\mathrm{d}\tau) := \frac{1}{2LN}\sum_{j=1}^N \operatorname{Tr} v_j^\zeta(\mathrm{d}\tau)$$

(cf. Definition 7.3.3 below). The *self-consistent spectrum* of $\boldsymbol{H}^\zeta$ is the set $\operatorname{supp}\rho^\zeta = \bigcup_{j=1}^N \operatorname{supp} v_j^\zeta$. Finally, for any $\varepsilon > 0$ the *self-consistent ε-pseudospectrum of $\boldsymbol{X}$* is defined by

$$\mathbb{D}_\varepsilon := \{\zeta \in \mathbb{C}\colon\ \operatorname{dist}(0, \operatorname{supp}\rho^\zeta) \leq \varepsilon\}. \tag{7.2.7}$$

The eigenvalues of $\boldsymbol{X}$ will concentrate on the set $\mathbb{D}_\varepsilon$ for any fixed $\varepsilon > 0$ if N is large. The motivation for this definiton (7.2.7) is that ζ is in the ε-pseudospectrum of $\boldsymbol{X}$ if and only if 0 is in the ε-vicinity of the spectrum of $\boldsymbol{H}^\zeta$, i.e., $\operatorname{dist}(0, \operatorname{Spec}(\boldsymbol{H}^\zeta)) \leq \varepsilon$. We recall that the *$\varepsilon$-pseudospectrum* $\operatorname{Spec}_\varepsilon(\boldsymbol{X})$ of $\boldsymbol{X}$ is defined through

$$\operatorname{Spec}_\varepsilon(\boldsymbol{X}) := \operatorname{Spec}(\boldsymbol{X}) \cup \{\zeta \in \mathbb{C} \setminus \operatorname{Spec}(\boldsymbol{X})\colon \|(\boldsymbol{X} - \zeta\mathbb{1})^{-1}\|_2 \geq \varepsilon^{-1}\}. \tag{7.2.8}$$

In accordance with Subsection 7.1.1, $\|\cdot\|_2$ denotes the operator norm on $\mathbb{C}^{L\times L} \otimes \mathbb{C}^{N\times N}$ induced by the Euclidean norm on $\mathbb{C}^L \otimes \mathbb{C}^N$ and $\mathbb{1}$ is the identity matrix in $\mathbb{C}^{L\times L} \otimes \mathbb{C}^{N\times N}$.

The precise statement is given in Theorem 7.2.4 below whose conditions we collect next.

Assumptions 7.2.3. (i) (Upper bound on variances) There is $\kappa_1 > 0$ such that

$$s_{ij}^\mu \leq \frac{\kappa_1}{N}, \qquad t_{ij}^\nu \leq \frac{\kappa_1}{N} \tag{7.2.9}$$

for all $i, j \in [N]$ and $\mu, \nu \in [\ell]$.

(ii) (Bounded moments) For each $p \in \mathbb{N}$, $p \geq 3$, there is $\varphi_p > 0$ such that

$$\mathbb{E}|x_{ij}^\mu|^p \leq \varphi_p N^{-p/2}, \qquad \mathbb{E}|y_{ij}^\nu|^p \leq \varphi_p N^{-p/2} \tag{7.2.10}$$

for all $i, j \in [N]$ and $\mu, \nu \in [\ell]$.

(iii) (Upper bound on structure matrices) There is $\kappa_2 > 0$ such that

$$\max_{\mu\in[\ell]}|\widetilde{\alpha}_\mu| \leq \kappa_2, \qquad \max_{\nu\in[\ell]}|\widetilde{\beta}_\nu| \leq \kappa_2, \tag{7.2.11}$$

where $|\alpha|$ denotes the operator norm induced by the Euclidean norm on $\mathbb{C}^L$.

(iv) (Bounded expectation) Let $\kappa_3 > 0$ be such that the matrices $\widetilde{a}_i \in \mathbb{C}^{L\times L}$ satisfy

$$\max_{i=1}^{N}|\widetilde{a}_i| \leq \kappa_3. \tag{7.2.12}$$

The constants L, ℓ, κ_1, κ_2, κ_3 and $(\varphi_p)_{p\in\mathbb{N}}$ are called *model parameters*. Our estimates will be uniform in all models possessing the same model parameters, in particular the bounds will be uniform in N, the large parameter in our problem. Now we can formulate our main result:

Theorem 7.2.4 (All eigenvalues of $\boldsymbol{X}$ are inside self-consistent ε-pseudospectrum)**.** *Fix $L \in \mathbb{N}$. Let $\boldsymbol{X}$ be a Kronecker random matrix as in* (7.2.1) *such that the bounds* (7.2.9) – (7.2.12) *are satisfied.*

Then for each $\varepsilon > 0$ and $D > 0$, there is a constant $C_{\varepsilon,D} > 0$ such that

$$\mathbb{P}(\operatorname{Spec}(\boldsymbol{X}) \subset \mathbb{D}_\varepsilon) \geq 1 - \frac{C_{\varepsilon,D}}{N^D}. \tag{7.2.13}$$

The constant $C_{\varepsilon,D}$ in (7.2.13) *only depends on the model parameters in addition to ε and D.*

Remark 7.2.5. (i) Theorem 7.2.4 follows from the slightly stronger Lemma 7.6.1 below; we show that not only the spectrum of $\boldsymbol{X}$ but also its $\varepsilon/2$-pseudospectrum lies in the self-consistent ε-pseudospectrum.

(ii) By carefully following the proof of Lemma 7.6.1, one can see that ε can be replaced by $N^{-\delta}$ with a small universal constant $\delta > 0$. The constant C in (7.2.13) will depend only on D and the model parameters.

(iii) (Only finitely many moments) If (7.2.10) holds true only for $p \leq P$ and some $P \in \mathbb{N}$ then there is a $D_0(P) \in \mathbb{N}$ such that the bound (7.2.13) is valid for all $D \leq D_0(P)$.

(iv) The self-consistent ε-pseudospectrum $\mathbb{D}_\varepsilon$ from (7.2.7) is defined in terms of the support of the self-consistent density of states of the Hermitized Dyson equation

(7.2.6). In particular, to determine $\mathbb{D}_\varepsilon$ one needs to solve the Dyson equation for spectral parameters z in a neighborhood of $z = 0$. There is an alternative definition for a deterministic ε-regularized set that is comparable to $\mathbb{D}_\varepsilon$ and requires to solve the Dyson equation solely on the imaginary axis $z = \mathrm{i}\eta$, namely

$$\widetilde{\mathbb{D}}_\varepsilon = \Big\{\zeta \colon \limsup_{\eta\downarrow 0} \frac{1}{\eta} \max_j |\mathrm{Im}\, m_j^\zeta(\mathrm{i}\eta)| \geq \frac{1}{\varepsilon}\Big\}. \tag{7.2.14}$$

Hence, (7.2.13) is true if $\mathbb{D}_\varepsilon$ is replaced by $\widetilde{\mathbb{D}}_\varepsilon$. For more details we refer the reader to Section 7.7 below.

(v) (Hermitian matrices) If $\boldsymbol{X}$ is a Hermitian random matrix, $\boldsymbol{X} = \boldsymbol{X}^*$, i.e., $\widetilde{\alpha}_\mu = \widetilde{\alpha}_\mu^*$ and $\widetilde{\beta}_\nu^* = \widetilde{\gamma}_\nu$ for all $\mu, \nu \in [\ell]$ and $\widetilde{a}_i^* = \widetilde{a}_i$ for all $i \in [N]$, then the Hermitization is superfluous and the Dyson equation may be formulated directly for $\boldsymbol{X}$. One may easily show that the support of the self-consistent density of states ρ is the intersection of all self-consistent ε-pseudospectra:

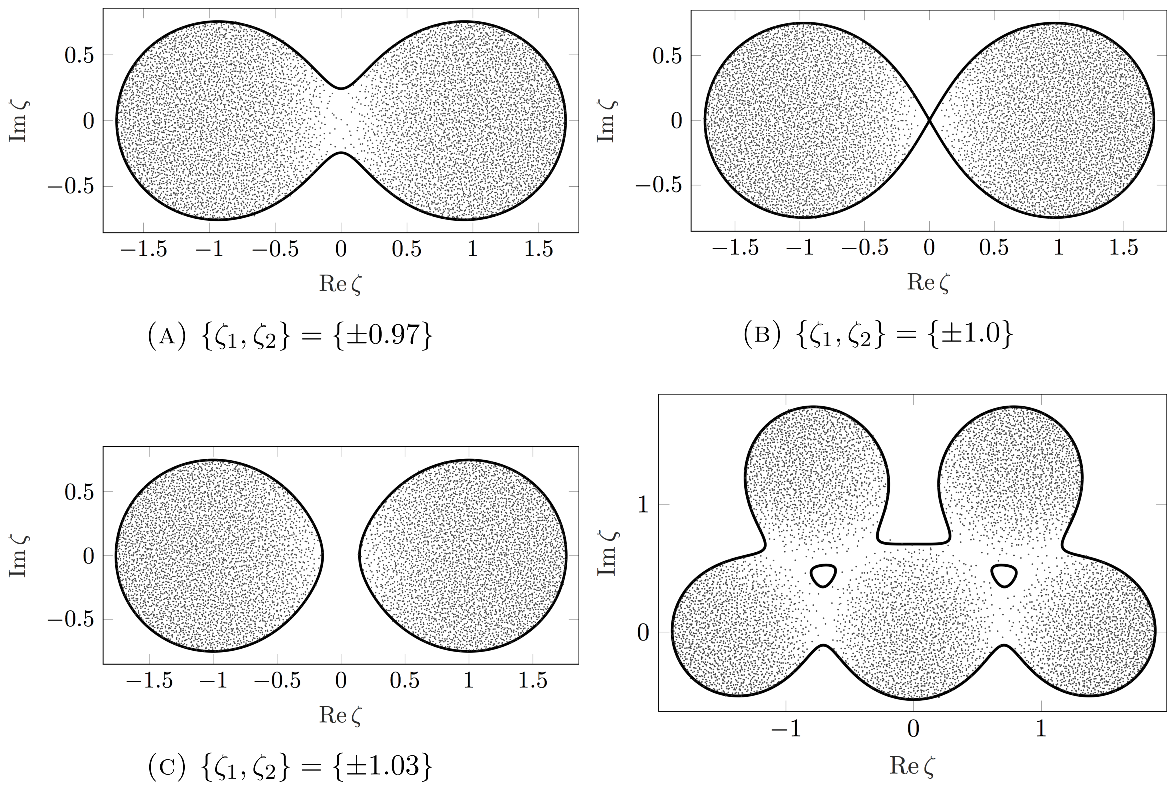
$$\mathrm{supp}\, \rho = \bigcap_{\varepsilon>0} \mathbb{D}_\varepsilon.$$

(vi) Theorem 7.2.4 as well as its stronger version for the Hermitian case, Theorem 7.4.7, identify a deterministic superset of the spectrum of $\boldsymbol{X}$. In fact, it is expected that for a large class of Kronecker matrices the set $\bigcap_{\varepsilon>0} \mathbb{D}_\varepsilon$ is the smallest deterministic set that still contains the entire $\mathrm{Spec}(\boldsymbol{X})$ up to a negligible distance. For $L = 1$ this has been proven for many Hermitian ensembles and for the circular ensemble. Example 7.2.6 below presents numerics for the $L \geq 2$ case.

Example 7.2.6. Fix $L \in \mathbb{N}$. Let $\zeta_1, \ldots, \zeta_L \in \mathbb{C}$ and $a \in \mathbb{C}^{L\times L}$ denote the diagonal matrix with $\zeta_1, \ldots, \zeta_L$ on its diagonal. We set $\boldsymbol{X} := a \otimes \mathbb{1} + \boldsymbol{W}$, where $\boldsymbol{W}$ has centered i.i.d. entries with variance $1/(NL)$. Clearly, $\boldsymbol{X}$ is a Kronecker matrix. In this case the Dyson equation can be directly solved and one easily finds that

$$\bigcap_{\varepsilon>0} \mathbb{D}_\varepsilon = \Big\{\zeta \in \mathbb{C} \colon \sum_{i=1}^{L} \frac{1}{|\zeta_i - \zeta|^2} \geq L\Big\} \tag{7.2.15}$$

(To our knowledge, the formula on the r.h.s. first appeared in [100]). Figure 7.1 shows the set (7.2.15) and the actual eigenvalues of $\boldsymbol{X}$ for $N = 8000$ and different matrices a.



(A) $\{\zeta_1, \zeta_2\} = \{\pm 0.97\}$

(B) $\{\zeta_1, \zeta_2\} = \{\pm 1.0\}$

(C) $\{\zeta_1, \zeta_2\} = \{\pm 1.03\}$

(D) $\{\zeta_1, \ldots, \zeta_5\} = \{0, \pm 1.4, \pm 0.8 + \mathrm{i}1.26\}$

FIGURE 7.1. Eigenvalues of sample random matrix with $N = 8000$ and $\cap_{\varepsilon>0}\mathbb{D}_\varepsilon$.

The *empirical density of states* of a Hermitian matrix $\boldsymbol{H} \in \mathbb{C}^{L\times L} \otimes \mathbb{C}^{N\times N}$ is defined through

$$\mu_{\boldsymbol{H}}(\mathrm{d}\tau) := \frac{1}{NL} \sum_{\lambda\in\mathrm{Spec}(\boldsymbol{H})} \delta_\lambda(\mathrm{d}\tau). \tag{7.2.16}$$

Theorem 7.2.7 (Global law for Hermitian Kronecker matrices)**.** *Fix $L \in \mathbb{N}$. For $N \in \mathbb{N}$, let $\boldsymbol{H}_N \in \mathbb{C}^{L\times L} \otimes \mathbb{C}^{N\times N}$ be a Hermitian Kronecker random matrix as in* (7.2.1) *such that the bounds* (7.2.9) – (7.2.12) *are satisfied. Then there exists a sequence of deterministic probability measures ρ_N on $\mathbb{R}$ such that the difference of ρ_N and the empirical spectral measure $\mu_{\boldsymbol{H}_N}$, defined in* (7.2.16)*, of $\boldsymbol{H}_N$ converges to zero weakly in probability, i.e.,*

$$\lim_{N\to\infty} \int_{\mathbb{R}} f(\tau)(\mu_{\boldsymbol{H}_N} - \rho_N)(\mathrm{d}\tau) = 0 \tag{7.2.17}$$

for all $f \in C_0(\mathbb{R})$ *in probability. Here,* $C_0(\mathbb{R})$ *denotes the continuous functions on* $\mathbb{R}$ *vanishing at infinity.*

Furthermore, there is a compact subset of $\mathbb{R}$ *which contains the supports of all* ρ_N*. This compact set depends only on the model parameters.*

Theorem 7.2.7 is proven in Section 7.8 below. The measure ρ_N, the self-consistent density of states, can be obtained by solving the corresponding Dyson equation, see Definition 7.3.3 later. If the function f is sufficiently regular then our proof combined with the Helffer-Sjöstrand formula yields an effective convergence rate of order $N^{-\delta}$ in (7.2.17).

7.3. Solution and stability of the Dyson equation

The general matrix Dyson equation (MDE) has been extensively studied in [6], but under conditions that exclude general Kronecker random matrices. Here, we relax these conditions and show how to extend some key results of [6] to our current setup. Our analysis of the MDE on the space of $n \times n$ matrices, $\mathcal{M} = \mathbb{C}^{n\times n}$, will then be applied to (7.2.6) with $n = 2LN = KN$. On $\mathcal{M} = \mathbb{C}^{n\times n}$, we use the norms as defined in Subsection 7.1.1 and require the pair $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$ to have the following properties:

Definition 7.3.1 (Data pair)**.** We call $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$ a *data pair* if

- The imaginary part $\operatorname{Im} \boldsymbol{A} = \frac{1}{2\mathrm{i}}(\boldsymbol{A} - \boldsymbol{A}^*)$ of the matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is negative semidefinite.
- The linear operator $\boldsymbol{\mathcal{S}} : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is self-adjoint with respect to the scalar product
$$\langle \boldsymbol{R}, \boldsymbol{T} \rangle := \frac{1}{n} \operatorname{Tr}[\boldsymbol{R}^* \boldsymbol{T}] \,,$$
and preserves the cone of positive semidefinite matrices, i.e., it is positivity preserving.

For any data pair $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$, the MDE then takes the form

$$-\boldsymbol{M}^{-1}(z) = z\mathbb{1} - \boldsymbol{A} + \boldsymbol{\mathcal{S}}[\boldsymbol{M}(z)], \quad z \in \mathbb{H}, \tag{7.3.1}$$

for a solution matrix $\boldsymbol{M}(z) \in \mathbb{C}^{n\times n}$. It was shown in this generality that the MDE, (7.3.1), has a unique solution under the constraint that the imaginary part $\operatorname{Im} \boldsymbol{M}(z) := (\boldsymbol{M}(z) -$

$\boldsymbol{M}(z)^*)/(2\mathrm{i})$ is positive definite [96]. We remark that $\mathrm{Im}\,\boldsymbol{A}$ being negative semidefinite is the most general condition for which our analysis is applicable. Furthermore, in [6], properties of the solution of (7.3.1) and the stability of (7.3.1) against small perturbations were studied in the general setup with Hermitian $\boldsymbol{A}$ and under the so-called *flatness* assumption,

$$\frac{c}{n}\,\mathrm{Tr}(\boldsymbol{R})\mathbb{1} \leq \boldsymbol{\mathcal{S}}[\boldsymbol{R}] \leq \frac{C}{n}\,\mathrm{Tr}(\boldsymbol{R})\mathbb{1}\,, \tag{7.3.2}$$

for all positive definite $\boldsymbol{R} \in \mathbb{C}^{n\times n}$ with some constants $C > c > 0$. Within Section 7.3 we will generalize certain results from [6] by dropping the flatness assumption (7.3.2) and the Hermiticity of $\boldsymbol{A}$. The results in this section, apart from (7.3.4b) below, follow by combining and modifying several arguments from [6]. We will only explain the main steps and refer to [6] for details. At the end of the section we translate these general results back to the setup of Kronecker matrices with the associated Dyson equation (7.2.6).

7.3.1. Solution of the Dyson equation. According to Proposition 2.1 in [6] the solution $\boldsymbol{M}$ to (7.3.1) has a Stieltjes transform representation

$$\boldsymbol{M}(z) \;=\; \int_{\mathbb{R}} \frac{\boldsymbol{V}(\mathrm{d}\tau)}{\tau - z}\,, \qquad z \in \mathbb{H}\,, \tag{7.3.3}$$

where $\boldsymbol{V}$ is a compactly supported measure on $\mathbb{R}$ with values in positive semidefinite $n \times n$-matrices such that $\boldsymbol{V}(\mathbb{R}) = \mathbb{1}$, provided $\boldsymbol{A}$ is Hermitian. The following lemma strengthens the conclusion about the support properties for this measure compared to Proposition 2.1 in [6].

Lemma 7.3.2. *Let* $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$ *be a data pair as in Definition 7.3.1 and* $\boldsymbol{M} : \mathbb{H} \to \mathbb{C}^{n\times n}$ *be the unique solution to* (7.3.1) *with positive definite imaginary part. Then*

(i) There is a unique measure $\boldsymbol{V}$ *on* $\mathbb{R}$ *with values in positive semidefinite matrices and* $\boldsymbol{V}(\mathbb{R}) = \mathbb{1}$ *such that* (7.3.3) *holds true.*

(ii) If $\boldsymbol{A}$ *is Hermitian, then*

$$\mathrm{supp}\,\boldsymbol{V} \;\subset\; \mathrm{Spec}\,\boldsymbol{A} + [-2\|\boldsymbol{\mathcal{S}}\|^{1/2}, 2\|\boldsymbol{\mathcal{S}}\|^{1/2}], \tag{7.3.4a}$$

$$\mathrm{Spec}\,\boldsymbol{A} \;\subset\; \mathrm{supp}\,\boldsymbol{V} + [-\|\boldsymbol{\mathcal{S}}\|^{1/2}, \|\boldsymbol{\mathcal{S}}\|^{1/2}]. \tag{7.3.4b}$$

Proof of Lemma 7.3.2. The representation (7.3.3) follows exactly as in the proof of Proposition 2.1 in [6] even for $\boldsymbol{A}$ with negative semidefinite imaginary part. We now prove (7.3.4a) motivated by the same proof in [6]. For a matrix $\boldsymbol{R} \in \mathbb{C}^{n\times n}$, its smallest singular value is denoted by $\sigma_{\min}(\boldsymbol{R})$. Note that $\sigma_{\min}(z\mathbb{1} - \boldsymbol{A}) = \operatorname{dist}(z, \operatorname{Spec} \boldsymbol{A})$ since $\boldsymbol{A}$ is Hermitian. In the following, we fix $z \in \mathbb{H}$ such that $\operatorname{dist}(z, \operatorname{Spec} \boldsymbol{A}) = \sigma_{\min}(z\mathbb{1} - \boldsymbol{A}) > 2\|\boldsymbol{\mathcal{S}}\|^{1/2}$.

Under the condition $\|\boldsymbol{M}(z)\|_2 \leq \sigma_{\min}(z\mathbb{1} - \boldsymbol{A})/(2\|\boldsymbol{\mathcal{S}}\|)$, we obtain from (7.3.1)

$$\begin{aligned} \|\boldsymbol{M}(z)\|_2 = \frac{1}{\sigma_{\min}(z\mathbb{1} - \boldsymbol{A} + \boldsymbol{\mathcal{S}}[\boldsymbol{M}(z)])} &\leq \frac{1}{\sigma_{\min}(z\mathbb{1} - \boldsymbol{A}) - \|\boldsymbol{\mathcal{S}}\|\|\boldsymbol{M}(z)\|_2} \\ &\leq \frac{2}{\operatorname{dist}(z, \operatorname{Spec} \boldsymbol{A})}. \end{aligned} \tag{7.3.5}$$

Therefore, using $\sigma_{\min}(z\mathbb{1} - \boldsymbol{A}) > 2\|\boldsymbol{\mathcal{S}}\|^{1/2}$, we find a gap in the values $\|\boldsymbol{M}(z)\|_2$ can achieve

$$\|\boldsymbol{M}(z)\|_2 \notin \Big(\frac{2}{\sigma_{\min}(z\mathbb{1} - \boldsymbol{A})}, \frac{\sigma_{\min}(z\mathbb{1} - \boldsymbol{A})}{2\|\boldsymbol{\mathcal{S}}\|}\Big).$$

For large values of $\eta = \operatorname{Im} z$, $\|\boldsymbol{M}(z)\|_2$ is smaller than the lower bound of this interval. Thus, since $\|\boldsymbol{M}(z)\|_2$ is a continuous function of z and the set $\{w \in \mathbb{H}\colon \operatorname{dist}(w, \operatorname{Spec} \boldsymbol{A}) > 2\|\boldsymbol{\mathcal{S}}\|^{1/2}\}$ is path-connected, we conclude that (7.3.5) holds true for all $z \in \mathbb{H}$ satisfying $\operatorname{dist}(z, \operatorname{Spec} \boldsymbol{A}) > 2\|\boldsymbol{\mathcal{S}}\|^{1/2}$.

We take the imaginary part of (7.3.1) and use $\boldsymbol{A} = \boldsymbol{A}^*$ to obtain $\operatorname{Im} \boldsymbol{M} = \eta \boldsymbol{M}^*\boldsymbol{M} + \boldsymbol{M}^*\boldsymbol{\mathcal{S}}[\operatorname{Im} \boldsymbol{M}]\boldsymbol{M}$. Solving this relation for $\operatorname{Im} \boldsymbol{M}$ and estimating its norm yields

$$\|\operatorname{Im} \boldsymbol{M}\|_2 \leq \frac{\eta\|\boldsymbol{M}\|_2^2}{1 - \|\boldsymbol{\mathcal{S}}\|\|\boldsymbol{M}\|_2^2} \leq \frac{4\eta}{\operatorname{dist}(z, \operatorname{Spec} \boldsymbol{A})^2 - 4\|\boldsymbol{\mathcal{S}}\|}.$$

Here, we employed $\|\boldsymbol{M}\|_2^2\|\boldsymbol{\mathcal{S}}\| < 1$ by (7.3.5) and $\operatorname{dist}(z, \operatorname{Spec} \boldsymbol{A}) > 2\|\boldsymbol{\mathcal{S}}\|^{1/2}$. Hence, $\operatorname{Im} \boldsymbol{M}$ converges to zero locally uniformly on the set $\{z \in \mathbb{H}\colon \operatorname{dist}(z, \operatorname{Spec} \boldsymbol{A}) > 2\|\boldsymbol{\mathcal{S}}\|^{1/2}\}$ for $\eta \downarrow 0$. Therefore, $E \notin \operatorname{supp} \boldsymbol{V}$ if $\operatorname{dist}(E, \operatorname{Spec} \boldsymbol{A}) > 2\|\boldsymbol{\mathcal{S}}\|^{1/2}$. This concludes the proof of (7.3.4a).

We now prove (7.3.4b). From (7.3.1), we obtain

$$\boldsymbol{A} - z\mathbb{1} = \boldsymbol{M}^{-1}(\mathbb{1} + \boldsymbol{M}\boldsymbol{\mathcal{S}}[\boldsymbol{M}]) \tag{7.3.6}$$

for $z \in \mathbb{H}$. Since $\boldsymbol{V}(\mathbb{R}) = 1$, we have

$$\|\boldsymbol{M}\|_2 \leq \frac{1}{\operatorname{dist}(z, \operatorname{supp} \boldsymbol{V})}. \tag{7.3.7}$$

Therefore, taking the inverse in (7.3.6) and applying (7.3.7) yield

$$\|(\boldsymbol{A} - z\mathbb{1})^{-1}\|_2 \leq \frac{1}{\operatorname{dist}(z, \operatorname{supp} \boldsymbol{V})(1 - \|\boldsymbol{\mathcal{S}}\| \operatorname{dist}(z, \operatorname{supp} \boldsymbol{V})^{-2})} \tag{7.3.8}$$

for all $z \in \mathbb{H}$ satisfying $\operatorname{dist}(z, \operatorname{supp} \boldsymbol{V})^2 > \|\boldsymbol{\mathcal{S}}\|$. Taking $\operatorname{Im} z \downarrow 0$ in (7.3.8), we see that the matrix $\boldsymbol{A} - E\mathbb{1}$ is invertible for all $E \in \mathbb{R}$ satisfying $\operatorname{dist}(E, \operatorname{supp} \boldsymbol{V})^2 > \|\boldsymbol{\mathcal{S}}\|$, showing (7.3.4b). □

In accordance with Definition 2.3 in [6] we define the self-consistent density of states as the unique measure whose Stieltjes transform is $n^{-1} \operatorname{Tr} \boldsymbol{M}$.

Definition 7.3.3 (Self-consistent density of states)**.** The measure

$$\rho(\mathrm{d}\tau) := \frac{1}{n} \operatorname{Tr} \boldsymbol{V}(\mathrm{d}\tau) \; = \; \langle \boldsymbol{V}(\mathrm{d}\tau) \rangle \tag{7.3.9}$$

is called the *self-consistent density of states*. Clearly, $\operatorname{supp} \rho = \operatorname{supp} \boldsymbol{V}$. For the following lemma, we also define the harmonic extension of the self-consistent density of states $\rho\colon \mathbb{H} \to \mathbb{R}_+$ through

$$\rho(z) := \frac{1}{\pi} \langle \operatorname{Im} \boldsymbol{M}(z) \rangle. \tag{7.3.10}$$

In the following we will use the short hand notation

$$d_\rho(z) := \operatorname{dist}(z, \operatorname{supp} \rho)\,.$$

Lemma 7.3.4 (Bounds on $\boldsymbol{M}$ and $\boldsymbol{M}^{-1}$)**.** *Let $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$ be a data pair as in Definition 7.3.1.*

(i) For $z \in \mathbb{H}$, we have the bounds

$$\|\boldsymbol{M}\|_2 \leq \frac{1}{d_\rho(z)}, \tag{7.3.11a}$$

$$(\operatorname{Im} z)\|\boldsymbol{M}^{-1}\|_2^{-2} \mathbb{1} \leq \operatorname{Im} \boldsymbol{M} \leq \frac{\operatorname{Im} z}{d_\rho^2(z)} \mathbb{1}, \tag{7.3.11b}$$

$$\|\boldsymbol{M}^{-1}\|_2 \leq |z| + \|\boldsymbol{A}\|_2 + \|\boldsymbol{\mathcal{S}}\| \|\boldsymbol{M}\|_2. \tag{7.3.11c}$$

(ii) For $z \in \mathbb{H}$, we have the bound

$$\rho(z) \leq \frac{\operatorname{Im} z}{\pi d_\rho^2(z)}. \tag{7.3.12}$$

PROOF. Using (7.3.3) immediately yields (7.3.11a) and the upper bound in (7.3.11b) since $\boldsymbol{V}(\mathbb{R}) = \mathbb{1}$. With $\eta = \operatorname{Im} z$ and taking the imaginary part of (7.3.1), we obtain

$$\operatorname{Im} \boldsymbol{M} = \eta \boldsymbol{M}^* \boldsymbol{M} - \boldsymbol{M}^* (\operatorname{Im} \boldsymbol{A}) \boldsymbol{M} + \boldsymbol{M}^* \boldsymbol{\mathcal{S}}[\operatorname{Im} \boldsymbol{M}] \boldsymbol{M} \geq \eta \boldsymbol{M}^* \boldsymbol{M}$$

as $\operatorname{Im} \boldsymbol{A} \leq 0$, $\operatorname{Im} \boldsymbol{M} \geq 0$ and $\boldsymbol{\mathcal{S}}$ is positivity preserving. Since $\boldsymbol{R}^* \boldsymbol{R} \geq \|\boldsymbol{R}^{-1}\|_2^{-2} \mathbb{1}$ for any $\boldsymbol{R} \in \mathbb{C}^{n \times n}$ the lower bound in (7.3.11b) follows. From (7.3.1), we obtain (7.3.11c). Since $\rho(z) = \pi^{-1} \langle \operatorname{Im} \boldsymbol{M}(z) \rangle$ the upper bound in (7.3.11b) implies (7.3.12). □

7.3.2. Stability of the Dyson equation. The goal of studying the stability of the Dyson equation in matrix form, (7.3.1), is to show that if some $\boldsymbol{G}$ satisfies

$$-\mathbb{1} = (z\mathbb{1} - \boldsymbol{A} + \boldsymbol{\mathcal{S}}[\boldsymbol{G}])\boldsymbol{G} + \boldsymbol{D} \tag{7.3.13}$$

for some small $\boldsymbol{D}$, then $\boldsymbol{G}$ is close to $\boldsymbol{M}$. It turns out that to a large extent this is a question about the invertibility of the stability operator $\boldsymbol{\mathcal{L}} := \operatorname{Id} - \boldsymbol{M}\boldsymbol{\mathcal{S}}[\,\cdot\,]\boldsymbol{M}$ acting on $\mathbb{C}^{n \times n}$. From (7.3.1) and (7.3.13), we obtain the following equation

$$\boldsymbol{\mathcal{L}}[\boldsymbol{G} - \boldsymbol{M}] = \boldsymbol{M}\boldsymbol{D} + \boldsymbol{M}\boldsymbol{\mathcal{S}}[\boldsymbol{G} - \boldsymbol{M}](\boldsymbol{G} - \boldsymbol{M}) \tag{7.3.14}$$

relating the difference $\boldsymbol{G} - \boldsymbol{M}$ with $\boldsymbol{D}$. We will call (7.3.14) the *stability equation.* Under the assumption that $\boldsymbol{G}$ is not too far from $\boldsymbol{M}$, the question whether $\boldsymbol{G} - \boldsymbol{M}$ is comparable with $\boldsymbol{D}$ is determined by the invertibility of $\boldsymbol{\mathcal{L}}$ in (7.3.14) and the boundedness of the inverse.

In this subsection, we show that $\|\boldsymbol{\mathcal{L}}^{-1}\|$ is bounded, provided $\operatorname{dist}(z, \operatorname{supp} \boldsymbol{V})$ is bounded away from zero. In order to prove this bound on $\boldsymbol{\mathcal{L}}^{-1}$, we follow the symmetrization procedure for $\boldsymbol{\mathcal{L}}$ introduced in [6]. We introduce the operators $\mathcal{C}_{\boldsymbol{R}} \colon \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$ and $\boldsymbol{\mathcal{F}} \colon \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$ through

$$\mathcal{C}_{\boldsymbol{R}}[\boldsymbol{Q}] = \boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}, \qquad \boldsymbol{\mathcal{F}} := \mathcal{C}_{\boldsymbol{W}} \mathcal{C}_{\sqrt{\operatorname{Im} \boldsymbol{M}}} \boldsymbol{\mathcal{S}} \mathcal{C}_{\sqrt{\operatorname{Im} \boldsymbol{M}}} \mathcal{C}_{\boldsymbol{W}},$$

for $\boldsymbol{Q} \in \mathbb{C}^{n\times n}$. Furthermore, the matrix $\boldsymbol{T} \in \mathbb{C}^{n\times n}$, the unitary matrix $\boldsymbol{U} \in \mathbb{C}^{n\times n}$ and the positive definite matrix $\boldsymbol{W} \in \mathbb{C}^{n\times n}$ are defined through

$$\boldsymbol{T} := \mathcal{C}^{-1}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}[\operatorname{Re}\boldsymbol{M}] - \mathrm{i}\mathbb{1}, \qquad \boldsymbol{U} := \frac{\boldsymbol{T}}{|\boldsymbol{T}|}, \qquad \boldsymbol{W} := |\boldsymbol{T}|^{1/2}.$$

With these notations, a direct calculation yields

$$\boldsymbol{\mathcal{L}} = \mathrm{Id} - \mathcal{C}_{\boldsymbol{M}}\boldsymbol{\mathcal{S}} = \mathcal{C}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}\mathcal{C}_{\boldsymbol{W}}\mathcal{C}_{\boldsymbol{U}^*}(\mathcal{C}_{\boldsymbol{U}} - \boldsymbol{\mathcal{F}})\mathcal{C}_{\boldsymbol{W}}^{-1}\mathcal{C}^{-1}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}, \tag{7.3.15}$$

as in (4.39) of [6].

We remark that $\mathcal{C}_{\boldsymbol{R}}$ for $\boldsymbol{R} \in \mathbb{C}^{n\times n}$ is invertible if and only if $\boldsymbol{R}$ is invertible and $\mathcal{C}_{\boldsymbol{R}}^{-1} = \mathcal{C}_{\boldsymbol{R}^{-1}}$ in this case. Similarly, $\mathcal{C}_{\boldsymbol{R}}^* = \mathcal{C}_{\boldsymbol{R}^*}$.

Our goal is to verify $\|\boldsymbol{\mathcal{F}}\|_{\mathrm{sp}} \le 1 - c$ for some positive constant c which yields $\|(\mathcal{C}_{\boldsymbol{U}} - \boldsymbol{\mathcal{F}})^{-1}\|_{\mathrm{sp}} \le c^{-1}$ as $\|\mathcal{C}_{\boldsymbol{U}}\|_{\mathrm{sp}} = 1$. Then the boundedness of the other factors in (7.3.15) implies the bound on the inverse of the stability operator $\boldsymbol{\mathcal{L}}$.

Convention 7.3.5 (Comparison relation)**.** *For nonnegative scalars or vectors f and g, we will use the notation $f \lesssim g$ if there is a constant $c > 0$, depending only on $\|\boldsymbol{\mathcal{S}}\|_{\mathrm{hs}\to\|\cdot\|}$ such that $f \le cg$ and $f \sim g$ if $f \lesssim g$ and $f \gtrsim g$ both hold true. If the constant c depends on an additional parameter (e.g. $\varepsilon > 0$), then we will indicate this dependence by a subscript (e.g. $\lesssim_\varepsilon$).*

Lemma 7.3.6. *Let $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$ be a data pair as in Definition 7.3.1.*

(i) Uniformly for any $z \in \mathbb{H}$, we have

$$d_\rho^4(z)\|\boldsymbol{M}^{-1}\|_2^{-2}\mathbb{1} \lesssim \boldsymbol{W}^4(\operatorname{Im} z)^2 \lesssim \|\boldsymbol{M}\|_2^2\|\boldsymbol{M}^{-1}\|_2^4\mathbb{1}. \tag{7.3.16}$$

(ii) There is a positive semidefinite $\boldsymbol{F} \in \mathbb{C}^{n\times n}$ such that $\|\boldsymbol{F}\|_{\mathrm{hs}} = 1$ and $\boldsymbol{\mathcal{F}}[\boldsymbol{F}] = \|\boldsymbol{\mathcal{F}}\|_{\mathrm{sp}}\boldsymbol{F}$. Moreover,

$$1 - \|\boldsymbol{\mathcal{F}}\|_{\mathrm{sp}} = (\operatorname{Im} z)\frac{\langle \boldsymbol{F}, \mathcal{C}_{\boldsymbol{W}}[\operatorname{Im}\boldsymbol{M}]\rangle}{\langle \boldsymbol{F}, \boldsymbol{W}^{-2}\rangle}. \tag{7.3.17}$$

(iii) Uniformly for $z \in \mathbb{H}$, we have

$$1 - \|\boldsymbol{\mathcal{F}}\|_{\mathrm{sp}} \gtrsim d_\rho^4(z)\|\boldsymbol{M}^{-1}\|_2^{-4}. \tag{7.3.18}$$

The proof of this lemma is motivated by the proofs of Lemma 4.6 and Lemma 4.7 (i) in [6].

Proof. We set $\eta := \operatorname{Im} z$. We rewrite the definition of $\boldsymbol{W}$ and use the upper bound in (7.3.11b) to obtain

$$\boldsymbol{W}^4 = \mathcal{C}^{-1}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}(\mathcal{C}_{\operatorname{Im}\boldsymbol{M}} + \mathcal{C}_{\operatorname{Re}\boldsymbol{M}})[(\operatorname{Im}\boldsymbol{M})^{-1}] \geq \eta^{-1} d_\rho^2(z)\mathcal{C}^{-1}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}[\boldsymbol{M}\boldsymbol{M}^* + \boldsymbol{M}^*\boldsymbol{M}]$$
$$\gtrsim \|\boldsymbol{M}^{-1}\|_2^{-2}\eta^{-2}d_\rho^4(z)\mathbb{1}.$$

Here, we also applied $\boldsymbol{M}\boldsymbol{M}^* + \boldsymbol{M}^*\boldsymbol{M} \geq 2\|\boldsymbol{M}^{-1}\|_2^{-2}\mathbb{1}$ and the upper bound in (7.3.11b) again. This proves the lower bound in (7.3.16). Similarly, using $\boldsymbol{M}\boldsymbol{M}^* + \boldsymbol{M}^*\boldsymbol{M} \leq 2\|\boldsymbol{M}\|_2^2\mathbb{1}$ and the lower bound in (7.3.11b) we obtain the upper bound in (7.3.16).

For the proof of (ii), we remark that $\boldsymbol{\mathcal{F}}$ preserves the cone of positive semidefinite matrices. Thus, by a version of the Perron-Frobenius theorem of cone preserving operators there is a positive semidefinite $\boldsymbol{F}$ such that $\|\boldsymbol{F}\|_{\mathrm{hs}} = 1$ and $\boldsymbol{\mathcal{F}}\boldsymbol{F} = \|\boldsymbol{\mathcal{F}}\|_{\mathrm{sp}}\boldsymbol{F}$. Following the proof of (4.24) in [6] and noting that this proof uses neither the uniqueness of $\boldsymbol{F}$ nor its positive definiteness, we obtain (7.3.17).

The bound in (7.3.18) is obtained by plugging the lower bound in (7.3.16) and the lower bound in (7.3.11b) into (7.3.17). We start by estimating the numerator in (7.3.17). Using $\boldsymbol{F} \geq 0$, the cyclicity of the trace, (7.3.11b) and the lower bound in (7.3.16), we get

$$\langle \boldsymbol{F}, \mathcal{C}_{\boldsymbol{W}}[\operatorname{Im}\boldsymbol{M}]\rangle \geq \eta\langle\sqrt{\boldsymbol{F}}\boldsymbol{W}^2\sqrt{\boldsymbol{F}}\rangle\|\boldsymbol{M}^{-1}\|_2^{-2} \gtrsim \|\boldsymbol{M}^{-1}\|_2^{-3}d_\rho^2(z)\langle\boldsymbol{F}\rangle. \tag{7.3.19}$$

Similarly, we have

$$\langle \boldsymbol{F}, \boldsymbol{W}^{-2}\rangle = \langle\sqrt{\boldsymbol{F}}\boldsymbol{W}^{-2}\sqrt{\boldsymbol{F}}\rangle \lesssim \frac{\eta}{d_\rho^2(z)}\|\boldsymbol{M}^{-1}\|_2\langle\boldsymbol{F}\rangle. \tag{7.3.20}$$

Combining (7.3.19) and (7.3.20) in (7.3.17) yields (7.3.18) and concludes the proof of the lemma. □

Lemma 7.3.7 (Bounds on the inverse of the stability operator)**.** *Let* $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$ *be a data pair as in Definition 7.3.1.*

(i) *The stability operator* $\boldsymbol{\mathcal{L}}$ *is invertible for all* $z \in \mathbb{H}$. *For fixed* $E \in \mathbb{R}$ *and uniformly for* $\eta \geq \max\{1, |E|, \|\boldsymbol{A}\|_2\}$, *we have*

$$\|\boldsymbol{\mathcal{L}}^{-1}(E + \mathrm{i}\eta)\| \lesssim 1. \tag{7.3.21}$$

(ii) *Uniformly for* $z \in \mathbb{H}$, *we have*

$$\|\boldsymbol{\mathcal{L}}^{-1}(z)\|_{\mathrm{sp}} \lesssim \frac{\|\boldsymbol{M}(z)\|_2 \|\boldsymbol{M}^{-1}(z)\|_2^9}{d_\rho^8(z)}. \tag{7.3.22}$$

(iii) *Uniformly for* $z \in \mathbb{H}$, *we have*

$$\|\boldsymbol{\mathcal{L}}^{-1}(z)\| + \|(\boldsymbol{\mathcal{L}}^{-1}(z))^*\| \lesssim 1 + \|\boldsymbol{M}(z)\|_2^2 + \|\boldsymbol{M}(z)\|_2^4 \|\boldsymbol{\mathcal{L}}^{-1}(z)\|_{\mathrm{sp}}. \tag{7.3.23}$$

PROOF. We start with the proof of (7.3.22). From the upper and lower bounds in (7.3.16) and (7.3.11b), respectively, we obtain

$$\|\mathcal{C}_{\boldsymbol{W}}\| \lesssim \frac{1}{\eta}\|\boldsymbol{M}\|_2 \|\boldsymbol{M}^{-1}\|_2^2, \qquad \|\mathcal{C}_{\boldsymbol{W}}^{-1}\| \lesssim \frac{\eta}{d_\rho^2(z)}\|\boldsymbol{M}^{-1}\|_2, \tag{7.3.24a}$$

$$\|\mathcal{C}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}\| \lesssim \frac{\eta}{d_\rho^2(z)}, \qquad \|\mathcal{C}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}^{-1}\| \lesssim \frac{1}{\eta}\|\boldsymbol{M}^{-1}\|_2^2. \tag{7.3.24b}$$

Since $\|\mathcal{C}_{\boldsymbol{T}}\|_{\mathrm{sp}} \leq \|\mathcal{C}_{\boldsymbol{T}}\|$ for Hermitian $\boldsymbol{T} \in \mathbb{C}^{n\times n}$ we conclude from (7.3.24), (7.3.18) and (7.3.11a) the bound

$$\|\boldsymbol{\mathcal{L}}^{-1}\|_{\mathrm{sp}} \lesssim \frac{\|\boldsymbol{M}\|_2\|\boldsymbol{M}^{-1}\|_2^5}{d_\rho^4(z)} \|(\mathcal{C}_{\boldsymbol{U}} - \boldsymbol{\mathcal{F}})^{-1}\|_{\mathrm{sp}} \lesssim \frac{\|\boldsymbol{M}\|_2\|\boldsymbol{M}^{-1}\|_2^9}{d_\rho^8(z)}.$$

For the proof of (7.3.23), we remark that $\|\boldsymbol{\mathcal{S}}\|_{\mathrm{hs}\to\|\cdot\|} \lesssim 1$ implies $\|\boldsymbol{\mathcal{S}}\|_{\|\cdot\|\to\mathrm{hs}} \lesssim 1$. Therefore, exactly as in the proof of (4.53) in [6], we obtain the first bound in (7.3.23). We similarly conclude the second bound from $\|(\boldsymbol{\mathcal{L}}^{-1})^*\|_{\mathrm{sp}} = \|\boldsymbol{\mathcal{L}}^{-1}\|_{\mathrm{sp}}$.

We conclude the proof of Lemma 7.3.7 by remarking that (7.3.21) is a consequence of (7.3.22), (7.3.11a), (7.3.23) and (7.3.11c). ∎

Corollary 7.3.8 (Lipschitz-continuity of $\boldsymbol{M}$). *If* $(\boldsymbol{A}, \boldsymbol{\mathcal{S}})$ *is a data pair as in Definition 7.3.1 then there exists* $c > 0$ *such that for each (possibly* N*-dependent)* $\varepsilon \in (0, 1]$ *we have*

$$\|\boldsymbol{M}(z_1) - \boldsymbol{M}(z_2)\|_2 \lesssim (\varepsilon^{-c} + \|\boldsymbol{A}\|_2^c)|z_1 - z_2| \tag{7.3.25}$$

for all $z_1, z_2 \in \mathbb{H}$ *such that* $\operatorname{Im} z_1, \operatorname{Im} z_2 \geq \varepsilon$.

Proof. We differentiate (7.3.1) with respect to z and obtain $\boldsymbol{\mathcal{L}}[\partial_z \boldsymbol{M}] = \boldsymbol{M}^2$. We invert $\boldsymbol{\mathcal{L}}$, use (7.3.22), (7.3.11a) and (7.3.11c) and follow the proof of (7.3.23). This yields (7.3.25) and hence concludes the proof of Corollary 7.3.8. □

7.3.3. Translation to results for Kronecker matrices. Here we translate the results of Subsections 7.3.1 and 7.3.2 into results about (7.2.6). In fact, we study (7.2.6) in a slightly more general setup. Motivated by the identification $\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L} \cong \mathbb{C}^{2L\times 2L}$, we consider (7.2.6) on $\mathbb{C}^{K\times K}$ for some $K \in \mathbb{N}$ instead. The results of Subsections 7.3.1 and 7.3.2 are applied with $n = KN$. Moreover, the special a_j^ζ defined in (7.2.5) are replaced by general $a_j \in \mathbb{C}^{K\times K}$. Therefore, the parameter ζ will not be present throughout this subsection. We thus look at the *Dyson equation in vector form*

$$-\frac{1}{m_j(z)} = z\mathbb{1} - a_j + \mathscr{S}_j[\boldsymbol{m}(z)], \tag{7.3.26}$$

where $z \in \mathbb{H}$, $m_j(z) \in \mathbb{C}^{K\times K}$ for $j \in [N]$, $\boldsymbol{m}(z) := (m_1(z), \dots m_N(z))$ and $\mathscr{S}_j$ is defined as in (7.2.4).

Recall that the definition of $\mathscr{S}_j$ involves coefficients s_{ij}^μ and t_{ij}^ν as well as matrices α_μ and β_ν. Next, we formulate assumptions on $\mathscr{S}$ in terms of these data as well as assumptions on $a_1, \dots, a_N$.

Assumptions 7.3.9.
(i) For all $\mu, \nu \in [\ell]$ and $i, j \in [N]$, we have nonnegative scalars $s_{ij}^\mu \in \mathbb{R}$ and $t_{ij}^\nu \in \mathbb{R}$ satisfying (7.2.9). Furthermore, $s_{ij}^\mu = s_{ji}^\mu$ for all $i, j \in [N]$ and $\mu \in [\ell]$.
(ii) For $\mu, \nu \in [\ell]$, we have $\alpha_\mu, \beta_\nu \in \mathbb{C}^{K\times K}$ and α_μ is Hermitian. There is $\alpha^* > 0$ such that

$$\max_{\mu\in[\ell]} |\alpha_\mu| \le \alpha^*, \qquad \max_{\nu\in[\ell]} |\beta_\nu| \le \alpha^*. \tag{7.3.27}$$

(iii) The matrices $a_1, \dots, a_N \in \mathbb{C}^{K\times K}$ have a negative semidefinite imaginary part, $\operatorname{Im} a_j \le 0$.

The conditions in (i) of Assumptions 7.3.9 are motivated by the definition of the variances in (7.2.3). In particular, since X_μ is Hermitian the variances from (7.2.3) satisfy $s_{ij}^\mu = s_{ji}^\mu$.

In order to apply the results of Subsections 7.3.1 and 7.3.2 to (7.3.26), we now relate it to the matrix Dyson equation (MDE) (7.3.1). It turns out that (7.3.26) is a special case when the MDE on $\mathcal{M} = \mathbb{C}^{K\times K} \otimes \mathbb{C}^{N\times N}$ is restricted to the *block diagonal matrices*

$$\mathcal{D} := \operatorname{span}\{a \otimes D \colon a \in \mathbb{C}^{K\times K}, D \in \mathbb{C}^{N\times N} \text{ diagonal}\} \subset \mathcal{M}. \tag{7.3.28}$$

We recall E_{ll}, $\mathscr{S}_l$ and P_{ll} from (7.1.6), (7.2.4) and (7.1.7), respectively, and define $\boldsymbol{A} \in \mathcal{M}$ and $\boldsymbol{\mathcal{S}} \colon \mathcal{M} \to \mathcal{M}$ through

$$\boldsymbol{A} := \sum_{l=1}^N a_l \otimes E_{ll}, \qquad \boldsymbol{\mathcal{S}}[\boldsymbol{R}] := \sum_{l=1}^N \mathscr{S}_l[(P_{11}\boldsymbol{R}, \ldots, P_{NN}\boldsymbol{R})] \otimes E_{ll}. \tag{7.3.29}$$

With these definitions, the Dyson equation in vector form, (7.3.26), can be rewritten in the matrix form (7.3.1) for a solution matrix $\boldsymbol{M} \in \mathcal{M}$. In the following, we will refer to (7.3.1) with these choices of $\mathcal{M}$, $\boldsymbol{A}$ and $\boldsymbol{\mathcal{S}}$ as the *Dyson equation in matrix form.*

In the remainder of the paper, we will consider the Dyson equation in matrix form, (7.3.1), exclusively with the choices of $\boldsymbol{A}$ and $\boldsymbol{\mathcal{S}}$ from (7.3.29). We have the following connection between (7.3.26) and (7.3.1). If $\boldsymbol{M}$ is a solution of (7.3.1) then, since the range of $\boldsymbol{\mathcal{S}}$ is contained in $\mathcal{D}$ and $\boldsymbol{A} \in \mathcal{D}$, we have $\boldsymbol{M} \in \mathcal{D}$, i.e, it can be written as

$$\boldsymbol{M}(z) = \sum_{j=1}^N m_j(z) \otimes E_{jj} \tag{7.3.30}$$

for some unique $m_1(z), \ldots, m_N(z) \in \mathbb{C}^{K\times K}$. Moreover, these m_i solve (7.3.26). Conversely, if $\boldsymbol{m} = (m_1, \ldots, m_N) \in (\mathbb{C}^{K\times K})^N$ solves (7.3.26) then $\boldsymbol{M}$ defined via (7.3.30) is a solution of (7.3.1). Furthermore, if $\boldsymbol{M}$ satisfies (7.3.30) then $\operatorname{Im} \boldsymbol{M}$ is positive definite if and only if $\operatorname{Im} m_j$ is positive definite for all $j \in [N]$. This correspondence yields the following translation of Lemma 7.3.2 to the setting for Kronecker random matrices, Proposition 7.3.10 below.

For part (ii), we recall $\|\boldsymbol{r}\| = \max_{i=1}^N |r_i|$ for $\boldsymbol{r} = (r_1, \ldots, r_N) \in (\mathbb{C}^{K\times K})^N$ and that $\|\mathscr{S}\|$ denotes the operator norm of $\mathscr{S} \colon (\mathbb{C}^{K\times K})^N \to (\mathbb{C}^{K\times K})^N$ induced by $\|\cdot\|$. We also used that $\|\mathscr{S}\| = \|\boldsymbol{\mathcal{S}}\|$, which is easy to see since $\mathscr{S} = \boldsymbol{\mathcal{S}}$ on the block diagonal matrices $(\mathbb{C}^{K\times K})^N \cong \mathcal{D}$ and $\boldsymbol{\mathcal{S}} = 0$ on the orthogonal complement $\mathcal{D}^\perp$. The orthogonal complement is defined with respect to the scalar product on $\mathcal{M}$ introduced in (7.1.5).

Furthermore, we remark that the identity (7.3.30) implies

$$\|\boldsymbol{M}\|_2 = \|\boldsymbol{m}\|.$$

Proposition 7.3.10 (Existence, uniqueness of $\boldsymbol{m}$)**.** *Under Assumptions 7.3.9 we have*

(i) There is a unique function $\boldsymbol{m}\colon \mathbb{H} \to (\mathbb{C}^{K\times K})^N$ *such that the components* $\boldsymbol{m}(z) = (m_1(z), \ldots, m_N(z))$ *satisfy* (7.3.26) *for* $z \in \mathbb{H}$ *and all* $j \in [N]$ *and* $\operatorname{Im} m_j(z)$ *is positive definite for all* $z \in \mathbb{H}$ *and all* $j \in [N]$*. Furthermore, for each* $j \in [N]$*, there is a measure* v_j *on* $\mathbb{R}$ *with values in the positive semidefinite matrices of* $\mathbb{C}^{K\times K}$ *such that* $v_j(\mathbb{R}) = \mathbb{1}$ *and for all* $z \in \mathbb{H}$*, we have*

$$m_j(z) = \int_{\mathbb{R}} \frac{v_j(\mathrm{d}\tau)}{\tau - z}. \tag{7.3.31}$$

(ii) If a_j *is Hermitian, i.e.,* $a_j = a_j^*$ *for all* $j \in [N]$ *then the union of the supports of* v_j *is comparable with the union of the spectra of the* a_j *in the following sense*

$$\bigcup_{j=1}^N \operatorname{supp} v_j \subset \bigcup_{j=1}^N \operatorname{Spec} a_j + [-2\|\mathscr{S}\|^{1/2}, 2\|\mathscr{S}\|^{1/2}], \tag{7.3.32a}$$

$$\bigcup_{j=1}^N \operatorname{Spec} a_j \subset \bigcup_{j=1}^N \operatorname{supp} v_j + [-\|\mathscr{S}\|^{1/2}, \|\mathscr{S}\|^{1/2}]. \tag{7.3.32b}$$

Proof of Lemma 7.2.2. Using the identification $\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L} \cong \mathbb{C}^{K\times K}$ for $K = 2L$ and the definitions in (7.2.2) and (7.2.5), the lemma is an immediate consequence of Proposition 7.3.10 with $a_j = a_j^\zeta$ for $j \in [N]$ since the proof of the proposition only uses the qualitative conditions in Assumptions 7.3.9. □

Proposition 7.3.10 asserts that there is a measure $V_{\boldsymbol{M}}$ on $\mathbb{R}$ with values in the positive semidefinite elements of $\mathcal{D} \subset \mathcal{M}$ such that for $z \in \mathbb{H}$, we have

$$V_{\boldsymbol{M}}(\mathrm{d}\tau) := \sum_{j=1}^{N} v_j(\mathrm{d}\tau) \otimes E_{jj}, \qquad \boldsymbol{M}(z) = \int_{\mathbb{R}} \frac{1}{\tau - z} V_{\boldsymbol{M}}(\mathrm{d}\tau). \tag{7.3.33}$$

Clearly, we have $V_{\boldsymbol{M}} = \boldsymbol{V}$ for the unique measure $\boldsymbol{V}$ with values in positive semidefinite matrices that satisfies (7.3.3). And we have $\operatorname{supp} V_{\boldsymbol{M}} = \operatorname{supp} \rho$ with the self-consistent density of states defined in (7.3.9). Note that in this setup

$$\rho(\mathrm{d}\tau) = \frac{1}{NK} \sum_{j=1}^{N} \operatorname{Tr} v_j(\mathrm{d}\tau)\,, \tag{7.3.34}$$

with the $\mathbb{C}^{K\times K}$-matrix valued measures v_j defined through (7.3.31).

In the remainder of the paper, $\boldsymbol{m} = (m_1, \ldots, m_N)$ and $\boldsymbol{M}$ always denote the unique solutions of (7.3.26) and (7.3.1), respectively, connected via (7.3.30). We now modify the concept of comparison relation introduced in Convection 7.3.5 so that inequalities are understood up to constants depending only on the model parameters from Assumption 7.3.9.

Convention 7.3.11 (Comparison relation)**.** *From here on we use the comparison relation introduced in Convection 7.3.5 so that the constants implicitly hidden in this notation may depend only on K, ℓ, κ_1 from* (7.2.9) *and α^* from* (7.3.27)*.*

Lemma 7.3.12 (Bounds on $\mathscr{S}$)**.** *Assumptions 7.3.9 imply*

$$\|\mathscr{S}\|_{\mathrm{sp}} \lesssim 1, \qquad \|\mathscr{S}\| \lesssim 1. \tag{7.3.35}$$

Proof. Direct estimates of $\mathscr{S}[\boldsymbol{a}]$ for $\boldsymbol{a} \in (\mathbb{C}^{K\times K})^N$ starting from the definition of $\mathscr{S}_i$, (7.2.4), and using the assumptions (7.2.9) and (7.3.27) yield the bounds in (7.3.35). □

Similarly to $\boldsymbol{\mathcal{L}}$, we now introduce the stability operator of the Dyson equation in vector form, (7.3.26). In fact, it is defined through

$$\mathscr{L}\colon (\mathbb{C}^{K\times K})^N \to (\mathbb{C}^{K\times K})^N, \quad \mathscr{L}(r_1, \ldots, r_N) := (r_i - m_i \mathscr{S}_i[\boldsymbol{r}] m_i)_{i=1}^N. \tag{7.3.36}$$

We remark that $\boldsymbol{\mathcal{S}}$ and thus $\boldsymbol{\mathcal{L}}$ leave the set of block diagonal matrices $\mathcal{D}$ defined in (7.3.28) invariant. The operators $\mathscr{S}$ and $\mathscr{L}$ are the restrictions of $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{L}}$ to $\mathcal{D}$. In particular, we have

$$\|\mathscr{L}^{-1}\|_{\mathrm{sp}} \leq \|\boldsymbol{\mathcal{L}}^{-1}\|_{\mathrm{sp}}, \qquad \|\boldsymbol{\mathcal{L}}^{-1}\|_{\mathrm{sp}} \leq \max\{1, \|\mathscr{L}^{-1}\|_{\mathrm{sp}}\}, \qquad \|\mathscr{L}^{-1}\| \leq \|\boldsymbol{\mathcal{L}}^{-1}\|, \quad (7.3.37)$$

since $\boldsymbol{\mathcal{L}}$ acts as the identity map on the orthogonal complement $\mathcal{D}^\perp$ of the block diagonal matrices. Here, the orthogonal complement is defined with respect to the scalar product on $\mathcal{M}$ introduced in (7.1.5). Moreover, $\mathscr{L}$ is invertible if and only if $\boldsymbol{\mathcal{L}}$ is invertible. Using (7.3.37) the bounds on $\boldsymbol{\mathcal{L}}$ from Lemma 7.3.7 can be translated into bounds on $\mathscr{L}$

7.4. Hermitian Kronecker matrices

The analysis of a non-Hermitian random matrix usually starts with Girko's Hermitization procedure. It provides a technique to extract spectral information about a non-Hermitian matrix $\boldsymbol{X}$ from a family of Hermitian matrices $(\boldsymbol{H}^\zeta)_{\zeta\in\mathbb{C}}$ defined through

$$\boldsymbol{H}^\zeta := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \boldsymbol{X} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes \boldsymbol{X}^* - \begin{pmatrix} 0 & \zeta \\ \bar{\zeta} & 0 \end{pmatrix} \otimes \mathbb{1}, \qquad \zeta \in \mathbb{C}. \tag{7.4.1}$$

Applying Girko's Hermitization procedure to a Kronecker random matrix $\boldsymbol{X}$ as in (7.2.1) generates a Hermitian Kronecker matrix $\boldsymbol{H}^\zeta \in \mathbb{C}^{2\times 2}\otimes\mathbb{C}^{L\times L}\otimes\mathbb{C}^{N\times N}$. However, similarly to our analysis in Section 7.3, we study more general Kronecker matrices $\boldsymbol{H} \in \mathbb{C}^{K\times K}\otimes\mathbb{C}^{N\times N}$ as in (7.4.2) below for $K, N \in \mathbb{N}$. This is motivated by the identification $\mathbb{C}^{2\times 2} \otimes \mathbb{C}^{L\times L} \cong \mathbb{C}^{2L\times 2L}$.

For $K, N \in \mathbb{N}$, let the random matrix $\boldsymbol{H} \in \mathbb{C}^{K\times K} \otimes \mathbb{C}^{N\times N}$ be defined through

$$\boldsymbol{H} := \sum_{\mu=1}^{\ell} \alpha_\mu \otimes X_\mu + \sum_{\nu=1}^{\ell} (\beta_\nu \otimes Y_\nu + \beta_\nu^* \otimes Y_\nu^*) + \sum_{i=1}^{N} a_i \otimes E_{ii}. \tag{7.4.2}$$

Furthermore, we make the following assumptions. Let $\ell \in \mathbb{N}$. For $\mu \in [\ell]$, let $\alpha_\mu \in \mathbb{C}^{K\times K}$ be a deterministic Hermitian matrix and $X_\mu = X_\mu^* \in \mathbb{C}^{N\times N}$ a Hermitian random matrix with centered and independent entries (up to the Hermitian symmetry constraint). For $\nu \in [\ell]$, let $\beta_\nu \in \mathbb{C}^{K\times K}$ be a deterministic matrix and Y_ν a random matrix with centered and independent entries. We also assume that $X_1, \ldots, X_\ell, Y_1, \ldots, Y_\ell$ are independent. Let $a_1, \ldots, a_N \in \mathbb{C}^{K\times K}$ be some deterministic matrices with negative semidefinite imaginary part. We recall that E_{ii} was defined in (7.1.6) and introduce the expectation $\boldsymbol{A} := \mathbb{E}\boldsymbol{H} = \sum_{i=1}^N a_i \otimes E_{ii}$.

If $\boldsymbol{A}$ is a Hermitian matrix then $\boldsymbol{H}$ as in (7.4.2) with the above properties is a Hermitian Kronecker random matrix in the sense of Definition 7.2.1. As in the setup from (7.2.1), the matrices $\alpha_1, \dots \alpha_\ell, \beta_1, \dots, \beta_\ell$ are called *structure matrices*.

Since the imaginary parts of $a_1, \dots, a_N$ are negative semidefinite, the same holds true for the imaginary part of $\boldsymbol{A}$ and $\boldsymbol{H}$. Hence, the matrix $\boldsymbol{H} - z\mathbb{1}$ is invertible for all $z \in \mathbb{H}$. For $z \in \mathbb{H}$, we therefore introduce the resolvent $\boldsymbol{G}(z)$ of $\boldsymbol{H}$ and its "matrix elements" $G_{ij}(z) := P_{ij}\boldsymbol{G} \in \mathbb{C}^{K\times K}$ for $i, j \in [N]$ defined through

$$\boldsymbol{G}(z) := (\boldsymbol{H} - z\mathbb{1})^{-1}, \qquad \boldsymbol{G}(z) = \sum_{i,j=1}^N G_{ij}(z) \otimes E_{ij}.$$

We recall that P_{ij} has been defined in (7.1.7). Our goal is to show that G_{ij} is small for $i \neq j$ and G_{ii} is well approximated by the deterministic matrix $m_i(z) \in \mathbb{C}^{K\times K}$ in the regime where $K \in \mathbb{N}$ is fixed and $N \in \mathbb{N}$ is large.

Apart from the above listed qualitative assumptions, we will need the following quantitative assumptions. To formulate them we use the same notation as before, i.e., the entries of X_μ and Y_ν are denoted by $X_\mu = (x_{ij}^\mu)_{i,j=1}^N$ and $Y_\nu = (y_{ij}^\nu)_{i,j=1}^N$ and their variances by $s_{ij}^\mu := \mathbb{E}|x_{ij}^\mu|^2$ and $t_{ij}^\nu := \mathbb{E}|y_{ij}^\nu|^2$ (cf. (7.2.3)).

Assumptions 7.4.1. We assume that all variances s_{ij}^μ and t_{ij}^μ satisfy (7.2.9) and the entries x_{ij}^μ and y_{ij}^ν of the random matrices fulfill the moment bounds (7.2.10). Furthermore, the structure matrices satisfy (7.3.27).

In this section, the model parameters are defined to be K, ℓ, κ_1 from (7.2.9), the sequence $(\varphi_p)_{p\in\mathbb{N}}$ from (7.2.10) and α^* from (7.3.27), so the relation $\lesssim$ indicates an inequality up to a multiplicative constant depending on these model parameters. Moreover, for the real and imaginary part of the spectral parameter z we will write $E = \operatorname{Re} z$ and $\eta = \operatorname{Im} z$, respectively.

7.4.1. Error term in the perturbed Dyson equation. We introduce the notion of *stochastic domination*, a high probability bound up to N^ε factors.

Definition 7.4.2 (Stochastic domination)**.** If $\Phi = (\Phi^{(N)})_N$ and $\Psi = (\Psi^{(N)})_N$ are two sequences of nonnegative random variables, then we say that Φ is **stochastically dominated** by Ψ, $\Phi \prec \Psi$, if for all $\varepsilon > 0$ and $D > 0$ there is a constant $C(\varepsilon, D)$ such that

$$\mathbb{P}\left(\Phi^{(N)} \geq N^\varepsilon \Psi^{(N)}\right) \leq \frac{C(\varepsilon, D)}{N^D} \tag{7.4.3}$$

for all $N \in \mathbb{N}$ and the function $(\varepsilon, D) \mapsto C(\varepsilon, D)$ depends only on the model parameters. If Φ or Ψ depend on some additional parameter δ and the function $(\varepsilon, D) \mapsto C(\varepsilon, D)$ additionally depends on δ then we write $\Phi \prec_\delta \Psi$.

We set $h_{ij} := P_{ij}\boldsymbol{H} \in \mathbb{C}^{K\times K}$. Using $P_{lm}\boldsymbol{A} = a_l\delta_{lm}$, $\mathbb{E}\, x^\mu_{ik} = 0$, $\mathbb{E}\, y^\nu_{ik} = 0$, (7.2.9), (7.3.27) and (7.2.10) we trivially obtain

$$|P_{ik}\left(\boldsymbol{H} - \boldsymbol{A}\right)| = |h_{ik} - a_i\delta_{ik}| \prec N^{-1/2}. \tag{7.4.4}$$

For $B \subset [N]$ we set

$$\boldsymbol{H}^B := \sum_{i,j=1}^N h^B_{ij} \otimes E_{ij}, \qquad h^B_{ij} := h_{ij}\mathbf{1}(i, j \notin B),$$

and denote the resolvent of $\boldsymbol{H}^B$ by $\boldsymbol{G}^B(z) := \left(\boldsymbol{H}^B - z\mathbb{1}\right)^{-1}$ for $z \in \mathbb{H}$. Since $\operatorname{Im}\boldsymbol{H}^B = \operatorname{Im}\boldsymbol{A}^B \leq 0$ for $B \subset [N]$, the matrix $(\boldsymbol{H}^B - z\mathbb{1})$ is invertible for all $z \in \mathbb{H}$ and

$$\|\boldsymbol{G}^B(z)\|_2 \leq \frac{1}{\operatorname{Im} z}. \tag{7.4.5}$$

In the following, we will use the convention

$$\sum_{k\in A}^{B} := \sum_{k\in A\setminus B}$$

for $A, B \subset [N]$ and $B \subset A$. If $A = [N]$ then we simply write $\sum^B_k$.

For $i \in [N]$, starting from the Schur complement formula,

$$-\frac{1}{G_{ii}} = z - h_{ii} + \sum_{k,l}^{\{i\}} h_{ik}G^{\{i\}}_{kl}h_{li}, \tag{7.4.6}$$

and using the definition of $\mathscr{S}_i$ in (7.2.4), we obtain the perturbed Dyson equation

$$-\frac{1}{g_i} = z\mathbb{1} - a_i + \mathscr{S}_i[\boldsymbol{g}] + d_i. \tag{7.4.7}$$

Here, we introduced

$$g_i := G_{ii}, \qquad \boldsymbol{g} := (g_1, \dots, g_N) \in (\mathbb{C}^{K\times K})^N \tag{7.4.8}$$

and the error term $d_i \in \mathbb{C}^{K\times K}$. We remark that (7.4.7) is a perturbed version of the Dyson equation in vector form, (7.3.26), and recall that $\boldsymbol{m}$ denotes its unique solution (cf. Proposition 7.3.10). To represent the error term d_i in (7.4.7), we use $h_{ik} = a_i\delta_{ik} + \sum_\mu x^\mu_{ik}\alpha_\mu + \sum_\nu (y^\nu_{ik}\beta_\nu + \overline{y^\nu_{ki}}\beta^*_\nu)$ and write $d_i := d_i^{(1)} + \ldots + d_i^{(8)}$, where

$$d_i^{(1)} := -h_{ii} + a_i, \tag{7.4.9a}$$

$$\begin{aligned} d_i^{(2)} := \sum_k^{\{i\}} \Big(&\sum_\mu \alpha_\mu G^{\{i\}}_{kk} \alpha_\mu \left(|x^\mu_{ik}|^2 - s^\mu_{ik} \right) \\ &+ \sum_\nu \left((|y^\nu_{ik}|^2 - t^\nu_{ik}) \beta_\nu G^{\{i\}}_{kk} \beta^*_\nu + (|y^\nu_{ki}|^2 - t^\nu_{ki}) \beta^*_\nu G^{\{i\}}_{kk} \beta_\nu \right) \Big), \end{aligned} \tag{7.4.9b}$$

$$d_i^{(3)} := \sum_\nu \sum_k^{\{i\}} \left(y^\nu_{ik} \beta_\nu G^{\{i\}}_{kk} \beta_\nu y^\nu_{ki} + \overline{y^\nu_{ki}} \beta^*_\nu G^{\{i\}}_{kk} \beta^*_\nu \overline{y^\nu_{ik}} \right) \tag{7.4.9c}$$

$$d_i^{(4)} := \Big(\sum_{\mu=\mu'} \sum_{k\neq l}^{\{i\}} + \sum_{\mu\neq\mu'} \sum_{k,l}^{\{i\}} \Big) \alpha_\mu x^\mu_{ik} G^{\{i\}}_{kl} x^{\mu'}_{li} \alpha_{\mu'}, \tag{7.4.9d}$$

$$d_i^{(5)} := \Big(\sum_{\nu=\nu'} \sum_{k\neq l}^{\{i\}} + \sum_{\nu\neq\nu'} \sum_{k,l}^{\{i\}} \Big) (y^\nu_{ik}\beta_\nu + \overline{y^\nu_{ki}}\beta^*_\nu)\, G^{\{i\}}_{kl} \left(y^{\nu'}_{li} \beta_{\nu'} + \overline{y^{\nu'}_{il}} \beta^*_{\nu'} \right), \tag{7.4.9e}$$

$$d_i^{(6)} := \sum_{k,l}^{\{i\}} \sum_\mu \sum_\nu \left(\alpha_\mu x^\mu_{ik} G^{\{i\}}_{kl} (y^\nu_{li}\beta_\nu + \overline{y^\nu_{il}}\beta^*_\nu) + (y^\nu_{ik}\beta_\nu + \overline{y^\nu_{ki}}\beta^*_\nu)\, G^{\{i\}}_{kl} x^\mu_{li} \alpha_\mu \right), \tag{7.4.9f}$$

$$\begin{aligned} d_i^{(7)} := \sum_k^{\{i\}} \Big(&\sum_\mu \alpha_\mu s^\mu_{ik} \left(G^{\{i\}}_{kk} - G_{kk} \right) \alpha_\mu \\ &+ \sum_\nu \left(t^\nu_{ik} \beta_\nu \left(G^{\{i\}}_{kk} - G_{kk} \right) \beta^*_\nu + t^\nu_{ki} \beta^*_\nu \left(G^{\{i\}}_{kk} - G_{kk} \right) \beta_\nu \right) \Big), \end{aligned} \tag{7.4.9g}$$

$$d_i^{(8)} := -\Big(\sum_\mu s^\mu_{ii} \alpha_\mu G_{ii} \alpha_\mu + \sum_\nu t^\nu_{ii} \left(\beta_\nu G_{ii} \beta^*_\nu + \beta^*_\nu G_{ii} \beta_\nu \right) \Big). \tag{7.4.9h}$$

In the remainder of this section, we consider $E = \operatorname{Re} z$ to be fixed and view quantities like $\boldsymbol{m}$ and $\boldsymbol{G}$ only as a function of $\eta = \operatorname{Im} z$. In the following lemma, we will use the following random control parameters to bound the error terms introduced in (7.4.9):

$$\begin{aligned} \Lambda_{\rm hs}(\eta) &:= \frac{1}{N}\Big[\operatorname{Tr} \boldsymbol{G}(E+\mathrm{i}\eta)^*\boldsymbol{G}(E+\mathrm{i}\eta)\Big]^{1/2} \\ \Lambda_{\rm w}(\eta) &:= \frac{1}{\sqrt{2N}} \max_{i=1}^{N} \Big[\operatorname{Tr} P_{ii}[\boldsymbol{G}(E+\mathrm{i}\eta)^*\boldsymbol{G}(E+\mathrm{i}\eta) + \boldsymbol{G}(E+\mathrm{i}\eta)\boldsymbol{G}(E+\mathrm{i}\eta)^*]\Big]^{1/2}, \quad (7.4.10) \\ \Lambda(\eta) &:= \max_{i,j=1}^{N} |G_{ij}(E+\mathrm{i}\eta) - m_i(E+\mathrm{i}\eta)\,\delta_{ij}|\,. \end{aligned}$$

We remark that due to our conventions, we have

$$\|\boldsymbol{m}\| = \max_{i=1}^{N}|m_i|, \qquad \|\boldsymbol{m}^{-1}\| = \max_{i=1}^{N}|m_i^{-1}|.$$

Lemma 7.4.3. *(i) Uniformly for $\eta \geq 1$ and $i \neq j$, we have*

$$|d_i| \prec 1, \tag{7.4.11a}$$

$$|G_{ij}| \prec \eta^{-2}. \tag{7.4.11b}$$

(ii) Uniformly for $\eta > 0$, we have

$$(|d_i^{(1)}| + \ldots + |d_i^{(6)}|)\,\chi \prec \frac{1}{\sqrt{N}} + \Lambda_{\rm hs} + \|\boldsymbol{m}^{-1}\|\Lambda_{\rm w}^2, \tag{7.4.12a}$$

$$(|d_i^{(7)}| + |d_i^{(8)}|)\,\chi \prec \|\boldsymbol{m}^{-1}\|\Lambda_{\rm w}^2 + \frac{1}{N}|G_{ii}|, \tag{7.4.12b}$$

where χ is the characteristic function $\chi := \mathbf{1}(\Lambda \leq (4\|\boldsymbol{m}^{-1}\|)^{-1})$.

Moreover, uniformly for $\eta > 0$ and $i \neq j$, we have

$$|G_{ij}|\,\chi \prec \|\boldsymbol{m}\|\Lambda_{\rm w}. \tag{7.4.13}$$

In the proof of Lemma 7.4.3, we use the following relation between the entries of $\boldsymbol{G}^T$ and $\boldsymbol{G}^{T\cup\{k\}}$

$$G_{ij}^T = G_{ij}^{T\cup\{k\}} + G_{ik}^T \frac{1}{G_{kk}^T} G_{kj}^T \tag{7.4.14}$$

for $T \subset [N]$, $k \notin T$ and $i, j \notin T \cup \{k\}$. This is an identity of $K \times K$ matrices and $1/G_{kk}^T$ is understood as the inverse matrix of G_{kk}^T. The proof of (7.4.14) follows from the Schur complement formula.

PROOF. We will prove the bounds in (7.4.12) in parallel with the estimate

$$|d_i^{(1)}| + \ldots + |d_i^{(8)}| \prec \frac{1}{\sqrt{N}} + \frac{1}{N}\Big(\sum_{k,l}^{\{i\}} |G_{kl}^{\{i\}}|^2\Big)^{1/2} + \frac{1}{N}\sum_{k}^{\{i\}} |G_{kk}^{\{i\}}| + \frac{1}{N}\sum_{k} |G_{kk}| \tag{7.4.15}$$

that we will use to show (7.4.11a).

The trivial estimate (7.4.4) implies that $|d_i^{(1)}| \prec 1/\sqrt{N}$.

In the remaining part of the proof, we will often apply the large deviation bounds with scalar valued random variables from Theorem C.1 in [60]. In our case, they will be applied to sums or quadratic forms of independent random variables, whose coefficients are $K \times K$ matrices; this generalization clearly follows from the scalar case [60] if applied to each entry separately.

We first show the following estimate

$$|d_i^{(2)}| + |d_i^{(3)}| \prec \frac{1}{\sqrt{N}}\Big(\frac{1}{N}\sum_{k}^{\{i\}} |G_{kk}^{\{i\}}|^2\Big)^{1/2}. \tag{7.4.16}$$

From the linear large deviation bound (C.2) in [60], we conclude that the first term in (7.4.9b) is bounded by

$$\sum_{\mu} |\alpha_\mu| \Big| \sum_{k}^{\{i\}} G_{kk}^{\{i\}} (|x_{ik}^\mu|^2 - s_{ik}^\mu) \Big| |\alpha_\mu| \prec \frac{1}{N}\Big(\sum_{k}^{\{i\}} |G_{kk}^{\{i\}}|^2\Big)^{1/2}.$$

The second and third term in (7.4.9b) are estimated similarly with the help of (C.2) in [60] which yields (7.4.16) for $|d_i^{(2)}|$. We apply the linear large deviation bound (C.2) in [60] and bound the first term in (7.4.9c) as follows:

$$\Big| \sum_{\nu} \Big(\sum_{k}^{\{i\}} y_{ik}^\nu y_{ki}^\nu \beta_\nu G_{kk}^{\{i\}} \beta_\nu \Big) \Big| \prec \frac{1}{N}\Big(\sum_{k}^{\{i\}} |G_{kk}^{\{i\}}|^2\Big)^{1/2}.$$

The bound on the second term in (7.4.9c) is obtained in the same way. Consequently, we have proven (7.4.16).

Using the quadratic large deviation bounds (C.4) and (C.3) in [60], we obtain

$$|d_i^{(4)}| + |d_i^{(5)}| + |d_i^{(6)}| \prec \Big(\frac{1}{N^2}\sum_{k,l}^{\{i\}} |G_{kl}^{\{i\}}|^2\Big)^{1/2}. \tag{7.4.17}$$

Moreover, (7.4.16) and (7.4.17) also imply that $|d_i^{(2)}| + \ldots + |d_i^{(6)}|$ are bounded by the second term on the right-hand side of (7.4.15).

Using (7.4.14), (7.2.9) and (7.3.27), we conclude

$$|d_i^{(7)}| \lesssim \min\Big\{\frac{1}{N}\sum_k^{\{i\}}|G_{ki}|\Big|\frac{1}{G_{ii}}\Big||G_{ik}|, \frac{1}{N}\sum_k^{\{i\}}(|G_{kk}^{\{i\}}| + |G_{kk}|)\Big\}. \tag{7.4.18}$$

The assumptions (7.2.9) and (7.3.27) imply

$$|d_i^{(8)}| \lesssim |G_{ii}|/N. \tag{7.4.19}$$

This concludes the proof of (7.4.15). Applying (7.4.5) to (7.4.15), we obtain (7.4.11a).

For all $k, l \notin \{i\}$, we now show that

$$\Big|G_{kl}^{\{i\}}\Big|\chi \leq |G_{kl}| + \frac{4}{3}\|\boldsymbol{m}^{-1}\||G_{ki}||G_{il}|. \tag{7.4.20}$$

This immediately yields (7.4.12a) using (7.4.16) and (7.4.17). For the proof of (7.4.20), we conclude from (7.4.14) by dividing and multiplying the second term by m_i that

$$G_{kl}^{\{i\}} = G_{kl} - G_{ki}\frac{1}{G_{ii}}m_i\frac{1}{m_i}G_{il}. \tag{7.4.21}$$

From the definition of χ in Lemma 7.4.3, we see that

$$\Big|\frac{1}{m_i}G_{ij} - \delta_{ij}\Big|\chi \leq \frac{1}{4}, \qquad \Big|\frac{1}{G_{ii}}m_i\Big|\chi \leq \frac{4}{3}, \tag{7.4.22}$$

which proves (7.4.20) and hence (7.4.12a).

Since (7.4.12b) is established for $|d_i^{(8)}|$ (cf. (7.4.19)), it suffices to use the second bound in (7.4.22) to finish the proof of (7.4.12b) by estimating $|d_i^{(7)}|$ via the first term in (7.4.18).

We now show (7.4.13) and (7.4.11b). The identity

$$G_{ij} = -\sum_k^{\{j\}} G_{ik}^{\{j\}} h_{kj} G_{jj}$$

and the linear large deviation bound (C.2) in [60] imply

$$|G_{ij}| \prec \Big(\frac{1}{N}\sum_k^{\{j\}}|G_{ik}^{\{j\}}|^2\Big)^{1/2}|G_{jj}|. \tag{7.4.23}$$

Using (7.4.5) to estimate $|G_{ik}^{\{j\}}|$ and $|G_{jj}|$, we obtain (7.4.11b). Applying the estimate (7.4.20) and the definition of χ in (7.4.23) yield $|G_{ij}|\chi \prec |G_{jj}|\chi\Lambda_{\mathrm{w}}$. Hence, the second bound in (7.4.22) implies (7.4.13) and conclude the proof of Lemma 7.4.3. □

For the following computations, we recall the definition of the product and the imaginary part on $(\mathbb{C}^{K\times K})^N$ from (7.1.3) and (7.1.4), respectively.

The proof of the following Lemma 7.4.4 is based on inverting the stability operator in the difference equation describing $\boldsymbol{g}-\boldsymbol{m}$ in terms of $\boldsymbol{d}$. We derive this equation first. Subtracting (7.3.26) from (7.4.7) and multiplying the result from the left by m_i and from the right by g_i yield

$$g_i - m_i = m_i\mathscr{S}_i[\boldsymbol{g}-\boldsymbol{m}]m_i + m_i d_i g_i + m_i\mathscr{S}_i[\boldsymbol{g}-\boldsymbol{m}](g_i - m_i)$$

for $i \in [N]$. Introducing $\boldsymbol{d} = (d_1, \ldots, d_N) \in (\mathbb{C}^{K\times K})^N$ as well as recalling $\mathscr{S}[\boldsymbol{r}] = (\mathscr{S}_i[\boldsymbol{r}])_{i=1}^N$, the definition of $\mathscr{S}_i$ from (7.2.4) and $\mathscr{L}[\boldsymbol{r}] = \boldsymbol{r} - \boldsymbol{m}\mathscr{S}[\boldsymbol{r}]\boldsymbol{m}$ from (7.3.36), we can write

$$\mathscr{L}(\boldsymbol{g}-\boldsymbol{m}) = \boldsymbol{m}\boldsymbol{d}\boldsymbol{g} + \boldsymbol{m}\mathscr{S}[\boldsymbol{g}-\boldsymbol{m}](\boldsymbol{g}-\boldsymbol{m}). \tag{7.4.24}$$

Since $\mathscr{L}$ is invertible for $z \in \mathbb{H}$ by Lemma 7.3.7 ((i)) and (7.3.37), applying the inverse of $\mathscr{L}$ on both sides of (7.4.24) and estimating the norm yields

$$\|\boldsymbol{g}-\boldsymbol{m}\| \;\le\; \|\mathscr{L}^{-1}\|\|\boldsymbol{m}\|(\|\boldsymbol{d}\|\|\boldsymbol{g}\| + \|\mathscr{S}\|\|\boldsymbol{g}-\boldsymbol{m}\|^2) \tag{7.4.25}$$

We recall the definition of ρ from (7.3.10).

Lemma 7.4.4. *(i) Uniformly for $\eta \ge \max\{1, |E|, \|\boldsymbol{A}\|_2\}$, we have*

$$\Lambda \prec \eta^{-2}. \tag{7.4.26}$$

(ii) Uniformly for $\eta > 0$, we have

$$\|\boldsymbol{g}-\boldsymbol{m}\|\,\mathbf{1}(\Lambda \le \vartheta) \;\prec\; \|\mathscr{L}^{-1}\|\|\boldsymbol{m}\|^2\left(\frac{1}{\sqrt{N}} + \Lambda_{\mathrm{hs}} + \|\boldsymbol{m}^{-1}\|\Lambda_{\mathrm{w}}^2\right), \tag{7.4.27}$$

where

$$\vartheta \;:=\; \frac{1}{4(\|\mathscr{L}^{-1}\|\|\boldsymbol{m}\|\|\mathscr{S}\| + \|\boldsymbol{m}^{-1}\|)}\,. \tag{7.4.28}$$

(iii) Let $a_1, \dots, a_N$ be Hermitian. We define

$$\psi := \|\mathscr{L}^{-1}\| \|\boldsymbol{m}\|^2 \|\boldsymbol{m}^{-1}\| \frac{1}{N\eta},$$

$$\varphi := \|\mathscr{L}^{-1}\| \|\boldsymbol{m}\|^2 \left(\frac{1}{\sqrt{N}} + \sqrt{\frac{\rho}{N\eta}} + \|\mathscr{L}^{-1}\| \|\boldsymbol{m}\|^2 \frac{1}{N\eta} + \frac{\|\boldsymbol{m}^{-1}\|}{N\eta} \|\mathrm{Im}\, \boldsymbol{m}\| \right) + \|\boldsymbol{m}\| \left(\sqrt{\frac{\|\mathrm{Im}\, \boldsymbol{m}\|}{N\eta}} + \frac{\|\boldsymbol{m}\|}{N\eta} \right).$$

Then for all $\delta > 0$ and uniformly for all $\eta > 0$ such that $\psi(\eta) \le N^{-\delta}$ we have

$$\Lambda\, \mathbf{1}(\Lambda \le \vartheta) \prec_\delta \varphi. \tag{7.4.29}$$

Note that the proof of (iii) of Lemma 7.4.4 requires $\boldsymbol{H}$ to be Hermitian because of the use of the Ward identity, $\boldsymbol{G}(\eta)^* \boldsymbol{G}(\eta) = \eta^{-1} \mathrm{Im}\, \boldsymbol{G}(\eta)$. The Ward identity implies $P_{ii} \boldsymbol{G}^* \boldsymbol{G} = P_{ii} \boldsymbol{G} \boldsymbol{G}^* = \mathrm{Im}\, G_{ii}/\eta$ and hence,

$$\Lambda_{\mathrm{hs}} = \sqrt{\frac{\langle \mathrm{Im}\, \boldsymbol{G} \rangle}{N\eta}}, \qquad \Lambda_{\mathrm{w}} = \max_i \sqrt{\frac{\mathrm{Im}\, \mathrm{Tr}\, G_{ii}}{N\eta}}. \tag{7.4.30}$$

Proof. We start with the proof of (7.4.26). We remark that $\|\boldsymbol{g}\| + \|\boldsymbol{m}\| \le 2/\eta$ by (7.4.5) and (7.3.11a). Therefore, for $\eta \ge \max\{1, |E|, \|\boldsymbol{A}\|_2\}$, we conclude from (7.4.25) that

$$\|\boldsymbol{g} - \boldsymbol{m}\| \lesssim \frac{1}{\eta^2} \|\boldsymbol{d}\| + \frac{1}{\eta^3}.$$

Here, we also used (7.3.21), (7.3.37) and (7.3.35). Since $\|\boldsymbol{d}\| \prec 1$ by (7.4.11a), we get $\|\boldsymbol{g} - \boldsymbol{m}\| \prec \eta^{-2}$ in this η-regime. Hence, combined with the bound (7.4.11b) for the offdiagonal terms, we obtain (7.4.26).

For the proof of (ii), we also start from (7.4.25). Since $2\|\mathscr{L}^{-1}\| \|\boldsymbol{m}\| \|\mathscr{S}\| \vartheta \le 1$ by definition of ϑ (cf. (7.4.28)) and $\|\boldsymbol{g}\| \mathbf{1}(\Lambda \le \vartheta) \le \|\boldsymbol{m}\| \|\boldsymbol{m}^{-1} \boldsymbol{g}\| \mathbf{1}(\Lambda \le \vartheta) \le 4\|\boldsymbol{m}\|/3$ by the second bound in (7.4.22), we conclude that

$$\|\boldsymbol{g} - \boldsymbol{m}\|\, \mathbf{1}(\Lambda \le \vartheta) \le 8 \|\mathscr{L}^{-1}\| \|\boldsymbol{m}\| \|\boldsymbol{d}\| \|\boldsymbol{m}\|/3. \tag{7.4.31}$$

Applying (7.4.12) to the right-hand side and using $|G_{ii}| \le \sqrt{N} \Lambda_{\mathrm{hs}}$, we obtain (7.4.27).

For the proof of (iii), let now $\boldsymbol{H}$ be Hermitian. Therefore, (7.4.30) is applicable and yields

$$\begin{aligned}\Lambda_{\mathrm{hs}} &= \sqrt{\frac{\langle \operatorname{Im} \boldsymbol{G}\rangle}{N\eta}} \lesssim \sqrt{\frac{\rho}{N\eta}} + \frac{1}{\varepsilon}\frac{1}{N\eta} + \varepsilon \|\boldsymbol{g}-\boldsymbol{m}\|\,,\\ \Lambda_{\mathrm{w}}^2 &= \Big(\max_{i=1}^{N} \sqrt{\frac{\operatorname{Im}\operatorname{Tr} G_{ii}}{N\eta}}\Big)^2 \le \frac{\|\operatorname{Im}\boldsymbol{m}\|}{N\eta} + \frac{\|\boldsymbol{g}-\boldsymbol{m}\|}{N\eta}.\end{aligned}$$

Here, we used $\langle \operatorname{Im} \boldsymbol{G}\rangle \le \langle \operatorname{Im} \boldsymbol{M}\rangle + \|\boldsymbol{g}-\boldsymbol{m}\|$, $\langle \operatorname{Im} \boldsymbol{M}\rangle = \pi\rho$ and Young's inequality as well as introduced an arbitrary $\varepsilon > 0$ in the first estimate. We plug these estimates into the right-hand side of (7.4.27) and choose $\varepsilon := N^{-\gamma}/(\|\mathscr{L}^{-1}\|\|\boldsymbol{m}\|^2)$ for arbitrary $\gamma > 0$. Thus, we can absorb $\|\boldsymbol{g}-\boldsymbol{m}\|$ in the estimate on Λ_{hs} into the left-hand side of (7.4.27). Similarly, using $\psi(\eta) \le N^{-\delta}$ we absorb $\|\boldsymbol{g}-\boldsymbol{m}\|$ in the estimate on Λ_{w} into the left-hand side of (7.4.27). This yields (7.4.29) for the contribution of the diagonal entries to Λ.

For the offdiagonal entries, we use the second relation in (7.4.30) and get as before

$$\Lambda_{\mathrm{w}} = \max_{i=1}^{N} \sqrt{\frac{\operatorname{Im}\operatorname{Tr} G_{ii}}{N\eta}} \le \sqrt{\frac{\|\operatorname{Im}\boldsymbol{m}\|}{N\eta}} + \frac{1}{\varepsilon}\frac{1}{N\eta} + \varepsilon\Lambda.$$

Using this estimate in (7.4.13) and choosing $\varepsilon := N^{-\gamma}/\|\boldsymbol{m}\|$ to absorb Λ into the left-hand side, we obtain (7.4.29) for diagonal and offdiagonal entries of $\boldsymbol{G}$. This concludes the proof of Lemma 7.4.4. □

Lemma 7.4.5 (Averaged local law)**.** *Suppose for some deterministic control parameter* $0 < \Phi \le N^{-\varepsilon}$ *a local law holds in the form*

$$\Lambda \prec \frac{\Phi}{\|\boldsymbol{m}^{-1}\|}\,. \tag{7.4.32}$$

Then for any deterministic $c_1, \dots, c_N \in \mathbb{C}^{K\times K}$ *with* $\max_i |c_i| \le 1$ *we have*

$$\begin{aligned}\Big|\frac{1}{N}\sum_{i=1}^{N} c_i^*(G_{ii}-m_i)\Big| \prec \|(\mathscr{L}^{-1})^*\|\|\boldsymbol{m}\|\Big(&\frac{\Phi^2}{\|\boldsymbol{m}^{-1}\|^2} + \max\Big\{\frac{1}{\sqrt{N}},\Phi\Big\}\Phi +\\ &\frac{\|\boldsymbol{m}\|^2}{N} + \Lambda_{\mathrm{w}}^2\|\boldsymbol{m}\|\|\boldsymbol{m}^{-1}\|\Big).\end{aligned} \tag{7.4.33}$$

In (7.4.33), the adjoint of $\mathscr{L}^{-1}$ is understood with respect to the scalar product $\operatorname{Tr}(\boldsymbol{x}\cdot\boldsymbol{y})$, where we defined the dot-product $\boldsymbol{x}\cdot\boldsymbol{y}$ for $\boldsymbol{x} = (x_1, \dots, x_N)$, $\boldsymbol{y} = (y_1, \dots, y_N) \in$

$(\mathbb{C}^{K\times K})^N$ via

$$\boldsymbol{x}\cdot\boldsymbol{y} := \frac{1}{N}\sum_{i=1}^N x_i^* y_i \in \mathbb{C}^{K\times K}. \tag{7.4.34}$$

It is easy to see that $\boldsymbol{x}\cdot\mathscr{L}^{-1}\boldsymbol{y} = ((\mathscr{L}^{-1})^*\boldsymbol{x})\cdot\boldsymbol{y}$.

Proof. We set $\boldsymbol{c} := (c_1,\ldots,c_N)$ and recall $\boldsymbol{g} = (G_{11},\ldots,G_{NN}) \in (\mathbb{C}^{K\times K})^N$. Using (7.4.24), we compute

$$\frac{1}{N}\sum_{i=1}^N c_i^*(G_{ii}-m_i) = \boldsymbol{c}\cdot(\boldsymbol{g}-\boldsymbol{m}) = (\boldsymbol{m}^*(\mathscr{L}^{-1})^*[\boldsymbol{c}])\cdot(\boldsymbol{dg}+\mathscr{S}[\boldsymbol{g}-\boldsymbol{m}](\boldsymbol{g}-\boldsymbol{m})). \tag{7.4.35}$$

We rewrite the term $\boldsymbol{dg}$ next. Indeed, a straightforward computation starting from the Schur complement formula (7.4.6) shows that

$$\begin{aligned} d_iG_{ii} &= \Big(Q_i\frac{1}{G_{ii}}\Big)G_{ii} + (d_i^{(7)}+d_i^{(8)})G_{ii} \\ &= \Big(Q_i\frac{1}{G_{ii}}\Big)m_i + \Big(Q_i\frac{1}{G_{ii}}\Big)(G_{ii}-m_i) + (d_i^{(7)}+d_i^{(8)})G_{ii}, \end{aligned} \tag{7.4.36}$$

where we defined $Q_iZ := Z - \mathbb{E}_iZ$ and the conditional expectation

$$\mathbb{E}_iZ := \mathbb{E}[Z|\boldsymbol{H}^{\{i\}}] = \mathbb{E}[Z|\{x_{kl}^\mu, y_{kl}^\nu \colon k,l\in[N]\setminus\{i\}, \mu,\nu\in[\ell]\}]$$

for any random variable Z.

The advantage of the representation (7.4.36) is that we can apply the following proposition to the first term on the right-hand side. It shows that when $Q_i(1/G_{ii})$ is averaged in i, there are certain cancellations taking place such that the average has a smaller order than $Q_i(1/G_{ii}) = \mathcal{O}(\Lambda)$. The first statement of this type was proven for generalized Wigner matrices in [72]. The complete proof in our setup will be presented in Section 7.5.

Proposition 7.4.6 (Fluctuation Averaging)**.** *Let* Φ *be a deterministic control parameter such that* $0<\Phi\le N^{-\varepsilon}$*. If*

$$\max_{i,j}\Big|\frac{1}{m_i}G_{ij}-\delta_{ij}\Big| \prec \Phi, \tag{7.4.37}$$

then for any deterministic $c_1, \ldots, c_N \in \mathbb{C}^{K\times K}$ *satisfying* $\max_i |c_i| \leq 1$ *we have*

$$\Big|\frac{1}{N}\sum_{i=1}^N c_i\, Q_i \frac{1}{G_{ii}} m_i\Big| \prec \max\Big\{\frac{1}{\sqrt{N}}, \Phi\Big\}\Phi\,. \tag{7.4.38}$$

Note that the assumption (7.4.32) directly implies (7.4.37). Moreover, (7.4.37) yields

$$\Big|\Big(Q_i \frac{1}{G_{ii}}\Big)\,(G_{ii} - m_i)\Big| \leq \Big|Q_i\Big(\frac{1}{G_{ii}}m_i - \mathbb{1}\Big)\Big| \|\boldsymbol{m}^{-1}\|\Lambda \prec \Phi^2.$$

Thus, we obtain from (7.4.35) and (7.4.36) the relation

$$\begin{aligned}|\boldsymbol{c}\cdot(\boldsymbol{g}-\boldsymbol{m})| \prec \|(\mathscr{L}^{-1})^*\|\|\boldsymbol{m}\|\Big(\frac{1}{N}\Big|\sum_{i=1}^N \widetilde{c}_i Q_i \frac{1}{G_{ii}} m_i\Big| + \Phi^2 + \\ \max_{i=1}^N(|d_i^{(7)}| + |d_i^{(8)}|)|G_{ii}| + \|\mathscr{S}\|\Lambda^2\Big),\end{aligned} \tag{7.4.39}$$

where $\widetilde{\boldsymbol{c}} = (\widetilde{c}_1, \ldots, \widetilde{c}_N) \in (\mathbb{C}^{K\times K})^N$ is a multiple of $\boldsymbol{m}^*(\mathscr{L}^{-1})^*[\boldsymbol{c}]$ and $\|\widetilde{\boldsymbol{c}}\| \leq 1$. From this estimate, we now conclude (7.4.33). Since (7.4.37) is satisfied by (7.4.32) the bound (7.4.38) implies that the first term on the right-hand side of (7.4.39) is controlled by the right-hand side of (7.4.33). For the third term, we use (7.4.12b) and $|G_{ii}| \leq \|\boldsymbol{m}\| + \Phi/\|\boldsymbol{m}^{-1}\|$ as well as $\Phi \leq 1 \leq \|\boldsymbol{m}\|\|\boldsymbol{m}^{-1}\|$. Hence, (7.3.35) concludes the proof of (7.4.33) and Lemma 7.4.5. □

7.4.2. No eigenvalues away from self-consistent spectrum. We now state and prove our result for Hermitian Kronecker matrices $\boldsymbol{H}$, Theorem 7.4.7 below. The theorem has two parts. For simplicity, we state the first part under the condition that $\boldsymbol{A} = \sum_i a_i \otimes E_{ii}$ is bounded. We relax this condition in the second part for the purpose of our main result, Theorem 7.2.4. In this application, $\boldsymbol{A} = \boldsymbol{A}^\zeta = \sum_i a_i^\zeta \otimes E_{ii}$, where a_i^ζ are given in (7.2.5), and we need to deal with unbounded ζ as well.

We recall that $\boldsymbol{m} = (m_1, \ldots, m_N)$ is the unique solution of (7.3.26) with positive imaginary part. Moreover, the function $\rho\colon \mathbb{H} \to \mathbb{R}_+$ was defined in (7.3.10), the set $\operatorname{supp}\rho$ in Definition 7.3.3 and $d_\rho(z) := \operatorname{dist}(z, \operatorname{supp}\rho)$. We denote $E := \operatorname{Re} z$ and $\eta := \operatorname{Im} z$. For a matrix $\boldsymbol{B}$, we write $\sigma_{\min}(\boldsymbol{B})$ to denote its smallest singular value.

Theorem 7.4.7 (No eigenvalues away from $\operatorname{supp}\rho$)**.** *Fix* $K \in \mathbb{N}$. *Let* $\boldsymbol{A} = \sum_{i=1}^N a_i \otimes E_{ii}$ *be a Hermitian matrix and* $\boldsymbol{H}$ *be a Hermitian Kronecker random matrix as in* (7.4.2) *such that* (7.2.9), (7.2.10) *and* (7.3.27) *are satisfied.*

(i) Assume that $\boldsymbol{A}$ *is bounded, i.e.,* $\|\boldsymbol{A}\|_2 \leq \kappa_4$*. Then there is a universal constant* $\delta > 0$ *such that for each* $D > 0$*, there is a constant* $C_D > 0$ *such that*

$$\mathbb{P}\Big(\operatorname{Spec}(\boldsymbol{H}) \subset \{\tau \in \mathbb{R}\colon \operatorname{dist}(\tau, \operatorname{supp}\rho) \leq N^{-\delta}\}\Big) \geq 1 - \frac{C_D}{N^D}. \tag{7.4.40}$$

(ii) Assume now only the weaker bound

$$\|\boldsymbol{A}\|_2 = \max_{i=1}^{N} |a_i| \leq N^{\kappa_7} \tag{7.4.41}$$

Let $\mathbb{H}_{\mathrm{out}}^{(2)}$ *be defined through*

$$\mathbb{H}_{\mathrm{out}}^{(2)} := \Big\{ w \in \mathbb{H}\colon \operatorname{dist}(w, \operatorname{Spec}\boldsymbol{A}) \geq 2\|\mathscr{S}\|^{1/2} + 1,\ \frac{\|\boldsymbol{A} - w\mathbb{1}\|_2}{\sigma_{\min}(\boldsymbol{A} - w\mathbb{1})} \leq \kappa_9 \Big\}. \tag{7.4.42}$$

Then for each $D > 0$*, there is a constant* $C_D > 0$ *such that*

$$\mathbb{P}\Big(\operatorname{Spec}(\boldsymbol{H}) \cap \mathbb{H}_{\mathrm{out}}^{(2)} = \emptyset\Big) \geq 1 - \frac{C_D}{N^D}. \tag{7.4.43}$$

The constants C_D *in* (7.4.40) *and* (7.4.43) *only depend on* K*,* κ_1*,* $(\varphi_p)_{p\geq 3}$*,* α_**,* κ_4*,* κ_7 *and* κ_9 *in addition to* D*.*

We will prove Theorem 7.4.7 as a consequence of the following Lemma 7.4.8. This lemma is a type of local law. Its general comprehensive version, Lemma 7.8.1 below, is a standard application of Lemma 7.4.4, Lemma 7.4.5 and Proposition 7.4.6. For the convenience of the reader, we will give an outline of the proof in Section 7.8 below.

We also consider $\kappa_7, \kappa_8, \kappa_9$ from (7.4.41) and (7.4.44) below, respectively, as model parameters.

Lemma 7.4.8. *Fix* $K \in \mathbb{N}$*. Let* $\kappa_7 > 0$ *and* $\boldsymbol{A} = \sum_{i=1}^N a_i \otimes E_{ii}$ *be a Hermitian matrix such that* (7.4.41) *holds true. Let* $\boldsymbol{H}$ *be a Hermitian Kronecker random matrix as in* (7.4.2) *such that* (7.2.9)*,* (7.2.10) *and* (7.3.27) *are satisfied. We define*

$$\mathbb{H}_{\mathrm{out}}^{(1)} := \Big\{ w \in \mathbb{H}\colon \operatorname{dist}(w, \operatorname{Spec}\boldsymbol{A}) \leq 2\|\mathscr{S}\|^{1/2} + 1,\ \|\boldsymbol{A}\|_2 \leq \kappa_8 \Big\}, \tag{7.4.44a}$$

$$\mathbb{H}_{\mathrm{out}}^{(2)} := \Big\{ w \in \mathbb{H}\colon \operatorname{dist}(w, \operatorname{Spec}\boldsymbol{A}) \geq 2\|\mathscr{S}\|^{1/2} + 1,\ \frac{\|\boldsymbol{A} - w\mathbb{1}\|_2}{\sigma_{\min}(\boldsymbol{A} - w\mathbb{1})} \leq \kappa_9 \Big\}. \tag{7.4.44b}$$

Then there are $p \in \mathbb{N}$ and $P \in \mathbb{N}$ independent of N and the model parameters such that

$$\Big|\frac{1}{N}\sum_{i=1}^{N} \operatorname{Tr} \operatorname{Im}\left(G_{ii}(z) - m_i(z)\right)\Big| \prec \max\Big\{1, \frac{1}{d_\rho^P(z)}\Big\}\Big(\frac{1}{N} + \frac{1}{(N\eta)^2}\Big) \tag{7.4.45}$$

for any $z = E + \mathrm{i}\eta \in \mathbb{H}_{\rm out}^{(1)} \cup \mathbb{H}_{\rm out}^{(2)}$ such that $|E| \le N^{\kappa_7+1}$ and $\eta \ge N^{-1+\gamma}(1 + d_\rho^{-p}(z))$.

We remark that since $\boldsymbol{A}$ is Hermitian, if $\|\boldsymbol{A}\|_2$ is bounded, then the second condition in (7.4.44b) is automatically satisfied (perhaps with a larger κ_9), given the first one. So for $\|\boldsymbol{A}\|_2 \le \kappa_8$, alternatively, we could have defined the sets

$$\begin{aligned} \mathbb{H}_{\rm out}^{(1)} &:= \Big\{w \in \mathbb{H}\colon\ \operatorname{dist}(w, \operatorname{Spec} \boldsymbol{A}) \le 2\|\mathscr{S}\|^{1/2} + 1\Big\}, \\ \mathbb{H}_{\rm out}^{(2)} &:= \Big\{w \in \mathbb{H}\colon\ \operatorname{dist}(w, \operatorname{Spec} \boldsymbol{A}) \ge 2\|\mathscr{S}\|^{1/2} + 1\Big\}. \end{aligned} \tag{7.4.46}$$

If $\|\boldsymbol{A}\|_2$ does not have an N-independent bound, then we could have defined $\mathbb{H}_{\rm out}^{(1)} := \emptyset$ and $\mathbb{H}_{\rm out}^{(2)}$ as in (7.4.42). The estimate (7.4.45) holds as stated with these alternative definitions of $\mathbb{H}_{\rm out}^{(1)}$ and $\mathbb{H}_{\rm out}^{(2)}$.

Definition 7.4.9. (Overwhelming probability) We say that an event $A^{(N)}$ happens *asymptotically with overwhelming probability*, a.w.o.p., if for each $D > 0$ there is $C_D > 0$ such that for all $N \in \mathbb{N}$, we have

$$\mathbb{P}(A^{(N)}) \ge 1 - \frac{C_D}{N^D}.$$

Proof of Theorem 7.4.7. From (7.4.4), we conclude the crude bound

$$\max_{\lambda \in \operatorname{Spec} \boldsymbol{H}} |\lambda|^2 \le \operatorname{Tr}(\boldsymbol{H}^2) = \sum_{i,j=1}^{N} |h_{ij}|^2 \prec (1 + \|\boldsymbol{A}\|_2^2) N. \tag{7.4.47}$$

Therefore, there are a.w.o.p. no eigenvalues of $\boldsymbol{H}$ outside of $[-a, a]$ with $a := (1 + \|\boldsymbol{A}\|_2)\sqrt{N}$.

We introduce the set $A_\delta := \{\omega \in \mathbb{R}\colon\ \operatorname{dist}(\omega, \operatorname{supp} \rho) \ge N^{-\delta}\}$ for $\delta > 0$. The previous argument proves that there are no eigenvalues in $A_\delta \setminus [-a, a]$ for any $\delta > 0$. For the opposite regime, i.e., to show that $A_\delta \cap [-a, a]$ does not contain any eigenvalue of $\boldsymbol{H}$ a.w.o.p. with some small $\delta > 0$, we use the following standard lemma and will include a proof for the reader's convenience at the end of this section.

Lemma 7.4.10. *Let $\boldsymbol{H}$ be an arbitrary Hermitian random matrix and $\boldsymbol{G}(z) := (\boldsymbol{H} - z\mathbb{1})^{-1}$ its resolvent at $z \in \mathbb{H}$. Let $\Phi\colon \mathbb{H} \to \mathbb{R}_+$ be a deterministic (possibly N-dependent) control parameter such that*

$$\frac{1}{N}\mathrm{Im}\ \mathrm{Tr}\,\boldsymbol{G}(\tau + \mathrm{i}\eta_0) \prec \Phi(\tau + \mathrm{i}\eta_0) \tag{7.4.48}$$

for some $\tau \in \mathbb{R}$ and $\eta_0 > 0$.

(i) If $(N\eta_0)^{-1} \geq N^{\varepsilon}\Phi(\tau + \mathrm{i}\eta_0)$ for some $\varepsilon > 0$ then $\mathrm{Spec}(\boldsymbol{H}) \cap [\tau - \eta_0, \tau + \eta_0] = \varnothing$ *a.w.o.p.*

(ii) Let $\mathcal{E} := \{\tau \in [-N^C, N^C]\colon (N\eta_0)^{-1} \geq N^{\varepsilon}\Phi(\tau + \mathrm{i}\eta_0)\}$ for some $C > 0$ and $\varepsilon > 0$. Furthermore, suppose that $\eta_0 \geq N^{-c}$ for some $c > 0$ and (7.4.48) *holds uniformly for all $\tau \in \mathcal{E}$. Then* $\mathrm{Spec}(\boldsymbol{H}) \cap \mathcal{E} = \varnothing$ *a.w.o.p.*

We now finish the proof of Theorem 7.4.7. In fact, by (7.4.41) we have $a \lesssim N^{\kappa_7 + 1/2}$, thus we work in the regime $|E| \leq N^{\kappa_7+1}$. We choose

$$\Phi(z) := \rho(z) + \max\{1, d_\rho^{-P}(z)\}\Big(\frac{1}{N} + \frac{1}{(N\,\mathrm{Im}\,z)^2}\Big) \quad \text{and} \quad \eta_0 := N^{-2/3}.$$

For small enough δ and γ, we can assume that $\eta_0 \geq N^{-1+\gamma}(1 + \mathrm{dist}(\tau + \mathrm{i}\eta_0, \mathrm{supp}\,\rho)^{-p})$ for $\mathrm{dist}(\tau, \mathrm{supp}\,\rho) \geq N^{-\delta}$. Consider first the case when $\|\boldsymbol{A}\|_2 \leq \kappa_4$, then $\mathbb{H}_{\mathrm{out}}^{(1)}$ and $\mathbb{H}_{\mathrm{out}}^{(2)}$ are complements of each other, see the remark at (7.4.46), and then (7.4.48) is satisfied by (7.4.45) for any τ with $|\tau| \leq N^{\kappa_7+1}$. Moreover, owing to (7.3.12), we have

$$\Phi(E + \mathrm{i}\eta_0) \lesssim \frac{N^{2\delta}}{N^{2/3}} + N^{P\delta}\Big(\frac{1}{N} + \frac{1}{N^{2/3}}\Big)$$

for all $E \in A_\delta \cap [-a, a]$. Therefore, by possibly reducing $\delta > 0$ and introducing a sufficiently small $\varepsilon > 0$, we can assume $N^{\varepsilon}\Phi(E + \mathrm{i}\eta_0) \leq N^{-1/3} = (N\eta_0)^{-1}$. Thus, from Lemma 7.4.10 we infer that $\boldsymbol{H}$ does not have any eigenvalues in $A_\delta \cap [-a, a]$ a.w.o.p. Combined with the argument preceding Lemma 7.4.10, which excludes a.w.o.p. eigenvalues of $\boldsymbol{H}$ in $A_\delta \setminus [-a, a]$, this proves (7.4.40) if $\|\boldsymbol{A}\|_2 \leq \kappa_4$. Under the weaker assumption $\|\boldsymbol{A}\|_2 \leq N^{\kappa_7}$ the same argument works but only for $E \in \mathbb{H}_{\mathrm{out}}^{(2)}$ since (7.4.45) was proven only in this regime. □

PROOF OF LEMMA 7.4.10. For the proof of part (i), we compute

$$\frac{1}{N}\mathrm{Im}\ \mathrm{Tr}\,\boldsymbol{G}(\tau+\mathrm{i}\eta)=\frac{1}{N}\sum_i\frac{\eta}{(\lambda_i-\tau)^2+\eta^2}.$$

Estimating the maximum from above by the sum, we obtain from the previous identity and the assumption that

$$\frac{1}{N}\max_i\frac{\eta_0}{(\lambda_i-\tau)^2+\eta_0^2}\prec\Phi\le\frac{N^{-\varepsilon}}{N\eta_0}. \tag{7.4.49}$$

We conclude that $\min_i|\lambda_i-\tau|\ge\eta_0$ a.w.o.p. and hence (i) follows.

The part (ii) is an immediate consequence of (i) and a union bound argument using the Lipschitz-continuity in τ on $\mathcal{E}$ of the left-hand side of (7.4.49) with Lipschitz-constant bounded by $N^{3(C+c)}$ and the boundedness of $\mathcal{E}$, i.e., $\mathcal{E}\subset[-N^C,N^C]$. □

7.5. Fluctuation Averaging: Proof of Proposition 7.4.6

In this section, we prove the Fluctuation Averaging which was stated as Proposition 7.4.6 in the previous section.

PROOF OF PROPOSITION 7.4.6. We fix an even $p\in\mathbb{N}$ and use the abbreviation

$$Z_i\;:=\;c_iQ_i\frac{1}{G_{ii}}m_i\,.$$

We will estimate the p-th moment of $\frac{1}{N}\sum_i Z_i$. For a p-tuple $\boldsymbol{i}=(i_1,\dots,i_p)\in\{1,\dots,N\}^p$ we call a label i_l a *lone label* if it appears only once in $\boldsymbol{i}$. We denote by J_L all tuples $\boldsymbol{i}\in\{1,\dots,N\}^p$ with exactly L lone labels. Then we have

$$\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N Z_i\right|^p\;\le\;\frac{1}{N^p}\sum_{L=0}^p\sum_{\boldsymbol{i}\in J_L}|\mathbb{E}\,Z_{i_1}\dots Z_{i_{p/2}}\overline{Z_{i_{p/2+1}}\dots Z_{i_p}}|\,. \tag{7.5.1}$$

For $\boldsymbol{i}\in J_L$ we estimate

$$|\mathbb{E}\,Z_{i_1}\dots Z_{i_{p/2}}\overline{Z_{i_{p/2+1}}\dots Z_{i_p}}|\;\prec\;\Phi^{p+L}. \tag{7.5.2}$$

Before verifying (7.5.2) we show this bound is sufficient to finish the proof. Indeed, using $|J_L| \le C(p)N^{(L+p)/2}$ and (7.5.2) in (7.5.1) yields

$$\mathbb{E}\Big|\frac{1}{N}\sum_{i=1}^N Z_i\Big|^p \prec \sum_{L=0}^p N^{-(p-L)/2}\Phi^{p+L} \prec \Big(\max\Big\{\frac{1}{\sqrt{N}},\Phi\Big\}\Phi\Big)^p.$$

This implies (7.4.38).

The rest of the proof is dedicated to showing (7.5.2). Since the complex conjugates do not play any role in the following arguments, we omit them in our notation. Furthermore, by symmetry we may assume that $\{i_1,\dots,i_L\}$ are the lone labels in $\boldsymbol{i}$.

We we fix $\ell \in \{0,\dots,L\}$ and $l \in \{1,\dots,p\}$. For any $K \in \mathbb{N}_0$ we call a pair

$$(\boldsymbol{t},\boldsymbol{T}) \qquad \text{with} \qquad \boldsymbol{t} = (t_1,\dots,t_{K-1}), \quad \boldsymbol{T} = (T_0,T_{01},T_1,T_{12},\dots,T_{K-1},T_{K-1K},T_K),$$

an *l-factor* (*at level* ℓ) if for all $k \in \{1,\dots,K-1\}$ and all $k' \in \{1,\dots,K-2\}$ the entries of the pair satisfy

$$\begin{aligned}&t_k \in \{i_1,\dots,i_\ell\}, \qquad T_k, T_{k'k'+1} \subseteq \{i_1,\dots,i_\ell\},\\ &t_{k'} \ne t_{k'+1}, \quad t_k \notin T_k, \quad t_{k'},t_{k'+1} \notin T_{k'k'+1}, \quad t_1 \ne i_l,, \quad t_{K-1} \ne i_l, \quad i_l \notin T_0 \cup T_{K+1}.\end{aligned} \tag{7.5.3}$$

Then we associate to such a pair the expression

$$Z_{\boldsymbol{t},\boldsymbol{T}} := c_{i_l} Q_{i_l}\Big[\frac{1}{G^{T_0}_{i_li_l}}G^{T_{01}}_{i_lt_1}\frac{1}{G^{T_1}_{t_1t_1}}G^{T_{12}}_{t_1t_2}\frac{1}{G^{T_2}_{t_2t_2}}\cdots\frac{1}{G^{T_{K-1}}_{t_{K-1}t_{K-1}}}G^{T_{K-1K}}_{t_{K-1}i_l}\frac{1}{G^{T_K}_{i_li_l}}\Big]m_{i_l}. \tag{7.5.4}$$

In particular, for $K=0$ we have

$$Z_{\emptyset,(T_0)} := c_{i_l}Q_{i_l}\frac{1}{G^{T_0}_{i_li_l}}m_{i_l}, \qquad Z_{\emptyset,(\emptyset)} := Z_{i_l}.$$

We also call

$$d(\boldsymbol{t},\boldsymbol{T}) := K,$$

the *degree* of the l-factor $(\boldsymbol{t},\boldsymbol{T})$.

By induction on ℓ we now prove the identity

$$\mathbb{E}\,Z_{i_1}\dots Z_{i_p} = \sum_{(\underline{\boldsymbol{t}},\underline{\boldsymbol{T}})\in\mathcal{I}_\ell}(\pm)\,\mathbb{E}\,Z_{\boldsymbol{t}_1,\boldsymbol{T}_1}\dots Z_{\boldsymbol{t}_p,\boldsymbol{T}_p}, \tag{7.5.5}$$

where the sign $(\pm)$ indicates that each summand may have a coefficient $+1$ or -1 and the sum is over a set $\mathcal{I}_\ell$ that contains pair of p-tuples $\underline{t} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_p)$ and $\underline{T} = (\boldsymbol{T}_1, \ldots, \boldsymbol{T}_p)$ such that $(\boldsymbol{t}_l, \boldsymbol{T}_l)$ for all $l = 1, \ldots, p$ is an l-factor at level ℓ. Furthermore, for all $\ell \in \{0, \ldots, L\}$ the size of $\mathcal{I}_\ell$ and the maximal degree of the l-factors $(\boldsymbol{t}_l, \boldsymbol{T}_l)$ are bounded by a constant depending only on p and

$$\sum_{i=1}^{p} \max\{1, d(\boldsymbol{t}_l, \boldsymbol{T}_l)\} \geq p + \ell\,, \qquad (\underline{\boldsymbol{t}}, \underline{\boldsymbol{T}}) \in \mathcal{I}_\ell\,. \tag{7.5.6}$$

The bound (7.5.2) follows from (7.5.5) and (7.5.6) for $\ell = L$ because

$$|Z_{\boldsymbol{t},\boldsymbol{T}}| \prec \Phi^{\max\{1, d(\boldsymbol{t},\boldsymbol{T})\}}, \tag{7.5.7}$$

for any l-factor $(\boldsymbol{t}, \boldsymbol{T})$. We postpone the proof of (7.5.7) to the very end of the proof of Proposition 7.4.6.

The start of the induction for the proof of (7.5.5) is trivial since for $\ell = 0$ we can chose the set $\mathcal{I}_\ell$ to contain only one element with $(\boldsymbol{t}_l, \boldsymbol{T}_l) = (\emptyset, (\emptyset))$ for all $l = 1, \ldots, p$. For the induction step, suppose that (7.5.5) and (7.5.6) have been proven for some $\ell \in \{1, \ldots, L-1\}$. Then we expand all l-factors $(\boldsymbol{t}_l, \boldsymbol{T}_l)$ with $l \neq \ell + 1$ within each summand on the right-hand side of (7.5.5) in the lone index $i_{\ell+1}$ by using the formulas

$$G_{ij}^{T} = G_{ij}^{T \cup \{k\}} + G_{ik}^{T} \frac{1}{G_{kk}^{T}} G_{kj}^{T}\,, \qquad i, j \notin \{k\} \cup T\,, \tag{7.5.8a}$$

$$\frac{1}{G_{ii}^{T}} = \frac{1}{G_{ii}^{T \cup \{k\}}} - \frac{1}{G_{ii}^{T}} G_{ik}^{T} \frac{1}{G_{kk}^{T}} G_{ki}^{T} \frac{1}{G_{ii}^{T \cup \{k\}}}\,, \qquad i \notin \{k\} \cup T\,, \tag{7.5.8b}$$

for $k = i_{\ell+1}$. More precisely, for all $l \neq \ell + 1$ we use (7.5.8) on each factor on the right-hand side of (7.5.4) with $(\boldsymbol{t}, \boldsymbol{T}) = (\boldsymbol{t}_l, \boldsymbol{T}_l)$; (7.5.8a) for the off-diagonal and (7.5.8b) for the inverse diagonal resolvent entries. Multiplying out the resulting factors, we write $\mathbb{E}\, Z_{\boldsymbol{t}_1,\boldsymbol{T}_1} \ldots Z_{\boldsymbol{t}_p,\boldsymbol{T}_p}$ as a sum of

$$2^{\sum_{l \neq \ell+1} 2d(\boldsymbol{t}_l, \boldsymbol{T}_l)+1}$$

summands of the form

$$\mathbb{E}\, Z_{\widetilde{\boldsymbol{t}}_1,\widetilde{\boldsymbol{T}}_1} \ldots Z_{\widetilde{\boldsymbol{t}}_p,\widetilde{\boldsymbol{T}}_p}\,, \tag{7.5.9}$$

where for all $l = 1, \ldots, p$ the pair $(\widetilde{\boldsymbol{t}}_l, \widetilde{\boldsymbol{T}}_l)$ is an l-factor at level $\ell + 1$. Note that we did not expand the $\ell + 1$-factor $Z_{\boldsymbol{t}_{\ell+1}, \boldsymbol{T}_{\ell+1}}$. In particular, the only nontrivial conditions for $(\widetilde{\boldsymbol{t}}_l, \widetilde{\boldsymbol{T}}_l)$ to be an l-factor at level $\ell + 1$ (cf. (7.5.3)), namely $t_k \neq t_{k+1}$, $t_1 \neq i_{\ell+1}$ and $t_{K-1} \neq i_{\ell+1}$, are satisfied because $i_{\ell+1}$ does not appear as a lower index on the right-hand side of (7.5.4) when on the left-hand side $(\boldsymbol{t}, \boldsymbol{T}) = (\boldsymbol{t}_l, \boldsymbol{T}_l)$.

Moreover all but one of the summands (7.5.9) satisfy

$$\sum_{i=1}^{p} d(\widetilde{\boldsymbol{t}}_l, \widetilde{\boldsymbol{T}}_l) \;\geq\; p + \ell + 1\,,$$

because the choice of the second summand in both (7.5.8a) and (7.5.8b) increases the number of off-diagonal resolvent elements in the l-factor that is expanded. The only exception is the summand (7.5.9) for which in the expansion in all factors always the first summand of (7.5.8a) and (7.5.8b) is chosen. However, in this case all $Z_{\widetilde{\boldsymbol{t}}_l, \widetilde{\boldsymbol{T}}_l}$ with $l \neq \ell + 1$ are independent of $i_{\ell+1}$ because this lone index has been completely removed from all factors. We conclude that this particular summand vanishes identically. Thus (7.5.6) holds with ℓ replaced by $\ell + 1$ and the induction step is proven.

It remains to verify (7.5.7). For $d(\boldsymbol{t}, \boldsymbol{T}) = 0$ we use that

$$\left| Q_{i_l} \frac{1}{G_{i_l i_l}} m_{i_l} \right| \;\leq\; \left| \frac{1}{G_{i_l i_l}} m_{i_l} - \mathbb{1} \right| \;\prec\; \Phi\,, \qquad \left| \frac{1}{G^{T}_{i_l i_l}} m_{i_l} - \frac{1}{G_{i_l i_l}} m_{i_l} \right| \;\prec\; \Phi^2\,. \tag{7.5.10}$$

The first bound in (7.5.10) simply uses the assumption (7.4.37) while the second bound uses the expansion formulas (7.5.8) and (7.4.37). For $K = d(\boldsymbol{t}, \boldsymbol{T}) > 0$ we realize that K encodes the number of off-diagonal resolvent entries G^{T}_{ij} in (7.5.4). In the factors of (7.5.4) we insert the entries of $\boldsymbol{M}$ so that (7.4.37) becomes usable, i.e., we use

$$\frac{1}{G^{T_k}_{t_k t_k}} G^{T_{kk+1}}_{t_k t_{k+1}} \;=\; \frac{1}{G^{T_k}_{t_k t_k}} m_{t_k} \frac{1}{m_{t_k}} G^{T_{kk+1}}_{t_k t_{k+1}}\,.$$

Then similarly to (7.5.10) we use

$$\left| \frac{1}{m_{t_k}} G_{t_k t_{k+1}} \right| \;\prec\; \Phi\,, \qquad \left| \frac{1}{m_{t_k}} G^{T_{kk+1}}_{t_k t_{k+1}} - \frac{1}{m_{t_k}} G_{t_k t_{k+1}} \right| \;\prec\; \Phi^2\,,$$

where again the first bound follows from (7.4.37) and the second bound from (7.5.8) and (7.4.37). □

7.6. Non-Hermitian Kronecker matrices and proof of Theorem 7.2.4

Since $\operatorname{Spec}(\boldsymbol{X}) \subset \operatorname{Spec}_\varepsilon(\boldsymbol{X})$ (cf. (7.2.8)) for all $\varepsilon > 0$, Theorem 7.2.4 clearly follows from the following lemma.

Lemma 7.6.1 (Pseudospectrum of $\boldsymbol{X}$ contained in self-consistent pseudospectrum)**.** *Under the assumptions of Theorem 7.2.4, we have that for each $\varepsilon \in (0,1]$, $\Delta > 0$ and $D > 0$, there is a constant $C_{\varepsilon,\Delta,D} > 0$ such that*

$$\mathbb{P}(\operatorname{Spec}_\varepsilon(\boldsymbol{X}) \subset \mathbb{D}_{\varepsilon+\Delta}) \geq 1 - \frac{C_{\varepsilon,\Delta,D}}{N^D}. \tag{7.6.1}$$

Proof. Let $\boldsymbol{H}^\zeta$ be defined as in (7.4.1). Note that $\zeta \in \operatorname{Spec}_\varepsilon(\boldsymbol{X})$ if and only if $\operatorname{dist}(0, \operatorname{Spec}(\boldsymbol{H}^\zeta)) \leq \varepsilon$. We set

$$\widetilde{\boldsymbol{A}} := \sum_{i=1}^N \widetilde{a}_i \otimes E_{ii}. \tag{7.6.2}$$

We first establish that $\operatorname{Spec}_\varepsilon(\boldsymbol{X})$ is contained in $D(0,N) := \{w \in \mathbb{C} \colon |w| \leq N\}$ a.w.o.p. Similarly, as in (7.4.47), using an analogue of (7.4.4) for $\boldsymbol{X}$ instead of $\boldsymbol{H}$, we get

$$\max_{\zeta \in \operatorname{Spec}\boldsymbol{X}} |\zeta|^2 \leq \operatorname{Tr}(\boldsymbol{X}^*\boldsymbol{X}) = \sum_{i,j=1}^N \operatorname{Tr}\left((P_{ij}\boldsymbol{X})^*(P_{ij}\boldsymbol{X})\right) \lesssim \sum_{i,j=1}^N |P_{ij}\boldsymbol{X}|^2 \prec (1 + \|\widetilde{\boldsymbol{A}}\|_2^2)N.$$

Thus, all eigenvalues of $\boldsymbol{X}$ have a.w.o.p. moduli smaller than $(1 + \|\widetilde{\boldsymbol{A}}\|_2)\sqrt{N} \leq N$. The above characterization of $\operatorname{Spec}_\varepsilon(\boldsymbol{X})$ and $\varepsilon \leq 1$ yield $\operatorname{Spec}_\varepsilon(\boldsymbol{X}) \subset D(0,N)$ a.w.o.p.

We now fix an $\varepsilon \in (0,1]$ and for the remainder of the proof the comparison relation $\lesssim$ is allowed to depend on ε without indicating that in the notation. In order to show that the complement of $\operatorname{Spec}_\varepsilon(\boldsymbol{X})$ contains $\mathbb{D}^c_{\varepsilon+\Delta} \cap D(0,N)$ a.w.o.p. we will apply Theorem 7.4.7 to $\boldsymbol{H}^\zeta$ for $\zeta \in \mathbb{D}^c_{\varepsilon+\Delta} \cap D(0,N)$. In particular, here we have

$$\boldsymbol{A} \;=\; \boldsymbol{A}^\zeta \;:=\; \sum_i a_i^\zeta \otimes E_{ii}\,,$$

where a_i^ζ is defined as in (7.2.5).

Now, we conclude that $\operatorname{Spec}(\boldsymbol{H}^\zeta) \cap [-\varepsilon - \Delta/2, \varepsilon + \Delta/2] = \varnothing$ a.w.o.p. for each $\zeta \in \mathbb{D}^c_{\varepsilon+\Delta} \cap D(0,N)$. If ζ is bounded, hence $\boldsymbol{A}^\zeta$ is bounded, we can use (7.4.40) and we need to show that $[-\varepsilon - \Delta/2, \varepsilon + \Delta/2] \subset \{\tau \in \mathbb{R} \colon \operatorname{dist}(\tau, \operatorname{supp}\rho^\zeta) \geq N^{-\delta}\}$ but this is straightforward since $\zeta \in \mathbb{D}^c_{\varepsilon+\Delta}$ implies $\operatorname{dist}(0, \operatorname{supp}\rho^\zeta) \geq \varepsilon + \Delta$ by its definition.

For large ζ we use part (ii) of Theorem 7.4.7 and we need to show that $[-\varepsilon-\Delta/2, \varepsilon+\Delta/2]+\mathrm{i}\eta \subset \mathbb{H}_{\mathrm{out}}^{(2)}$ for any small η. Take $z \in \mathbb{H}$ with $|z| \leq \varepsilon+\Delta/2$. If $|\zeta| \geq \|\widetilde{\boldsymbol{A}}\|+2\|\mathscr{S}\|^{1/2}+2$, then $\operatorname{dist}(z, \operatorname{Spec}(\boldsymbol{A}^\zeta)) \geq 2\|\mathscr{S}\|^{1/2}+1$, so the first condition in the definition (7.4.44b) of $\mathbb{H}_{\mathrm{out}}^{(2)}$ is satisfied. The second condition is straightforward since for large ζ and small z, both $\|\boldsymbol{A}^\zeta - z\mathbb{1}\|_2$ and $\sigma_{\min}(\boldsymbol{A}^\zeta - z\mathbb{1})$ are comparable with $|\zeta|$.

Hence, Theorem 7.4.7 is applicable and we conclude that $\operatorname{Spec}(\boldsymbol{H}^\zeta) \cap [-\varepsilon-\Delta/2, \varepsilon+\Delta/2] = \varnothing$ a.w.o.p. for all $\zeta \in \mathbb{D}^c_{\varepsilon+\Delta}$. If $\lambda_1(\zeta) \leq \ldots \leq \lambda_{2LN}(\zeta)$ denote the ordered eigenvalues of $\boldsymbol{H}^\zeta$ then $\lambda_i(\zeta)$ is Lipschitz-continuous in ζ by the Hoffman-Wielandt inequality. Therefore, introducing a grid in ζ and applying a union bound argument yield

$$\sup_{\zeta \in \mathbb{D}^c_{\varepsilon+\Delta} \cap D(0,N)} \operatorname{dist}(0, \operatorname{Spec}(\boldsymbol{H}^\zeta)) \leq \varepsilon \quad \text{a.w.o.p.}$$

Since $\zeta \in \operatorname{Spec}_\varepsilon(\boldsymbol{X})$ if and only if $\operatorname{dist}(0, \operatorname{Spec}(\boldsymbol{H}^\zeta)) \leq \varepsilon$ we obtain $\operatorname{Spec}_\varepsilon(\boldsymbol{X}) \cap \mathbb{D}^c_{\varepsilon+\Delta} \cap D(0,N) = \varnothing$ a.w.o.p. As we proved $\operatorname{Spec}_\varepsilon(\boldsymbol{X}) \cap D(0,N)^c = \varnothing$ a.w.o.p. before this concludes the proof of Lemma 7.6.1. □

7.7. An alternative definition of the self-consistent ε-pseudospectrum

Instead of the self-consistent ε-pseudospectrum $\mathbb{D}_\varepsilon$ introduced in (7.2.7) one may work with the deterministic set $\widetilde{\mathbb{D}}_\varepsilon$ from (7.2.14) when formulating our main result, Theorem 7.2.4. The advantage of the set $\widetilde{\mathbb{D}}_\varepsilon$ is that it only requires solving the Hermitized Dyson equation (7.2.6) for spectral parameters z along the imaginary axis. The following lemma shows that $\mathbb{D}_\varepsilon$ and $\widetilde{\mathbb{D}}_\varepsilon$ are comparable in the sense that for any ε we have $\mathbb{D}_{\varepsilon_1} \subseteq \widetilde{\mathbb{D}}_\varepsilon \subseteq \mathbb{D}_{\varepsilon_2}$ for certain $\varepsilon_1, \varepsilon_2$.

Lemma 7.7.1. *Let $\boldsymbol{m}$ be the solution to the Hermitized Dyson equation* (7.2.6) *and suppose Assumptions 7.2.3 are satisfied. There is a positive constant c, depending only on model parameters, such that for any $\varepsilon \in (0,1)$ we have the inclusions*

$$\widetilde{\mathbb{D}}_\varepsilon \subseteq \mathbb{D}_{\sqrt{\varepsilon}}\,, \qquad \mathbb{D}_{c\varepsilon^{27}} \subseteq \widetilde{\mathbb{D}}_\varepsilon\,,$$

where $\mathbb{D}_\varepsilon$ is the self-consistent ε-pseudospectrum from (7.2.7) *and $\widetilde{\mathbb{D}}_\varepsilon$ is defined in* (7.2.14).

Proof. The inclusion $\widetilde{\mathbb{D}}_\varepsilon \subseteq \mathbb{D}_{\sqrt{\varepsilon}}$ is trivial because m_j^ζ is the Stieltjes transform of v_j^ζ. So we concentrate on the inclusion $\mathbb{D}_{c\varepsilon^{27}} \subseteq \widetilde{\mathbb{D}}_\varepsilon$. We fix $\zeta \in \mathbb{C}\backslash\widetilde{\mathbb{D}}_\varepsilon$ and suppress it from our notation in the following, i.e., $\boldsymbol{m} = \boldsymbol{m}^\zeta, v_j = v_j^\zeta$, etc. Recall that by assumption we have (cf. (7.6.2))

$$\|\widetilde{\boldsymbol{A}}\| \lesssim 1\,.$$

Since any large enough ζ is contained in both sets $\mathbb{C}\setminus\widetilde{\mathbb{D}}_\varepsilon$ and $\mathbb{C}\setminus\mathbb{D}_\varepsilon$ by (7.3.32a) and the upper bound in (7.3.11b), we may assume that $|\zeta| \lesssim 1$. We use the representation of m_i as the Stieltjes transform of v_i and that v_i has bounded support to see

$$|\langle x\,, m_i(z)y\rangle| \,\le\, \frac{1}{2}\int_{\mathbb{R}} \frac{\langle x\,, v_i(\mathrm{d}\tau)x\rangle + \langle y\,, v_i(\mathrm{d}\tau)y\rangle}{|\tau - z|} \,\lesssim\, \frac{1}{\eta}(\langle x\,, \operatorname{Im} m_i(z)x\rangle + \langle y\,, \operatorname{Im} m_i(z)y\rangle)\,,$$

for any $x, y \in \mathbb{C}^K$, where $K = 2L$. In particular

$$|m_i(z)| \,\lesssim\, \frac{|\operatorname{Im} m_i(z)|}{\eta}\,. \tag{7.7.1}$$

Fix an $\eta \in (0,1)$ for which the inequality

$$\frac{1}{\eta}\|\operatorname{Im} \boldsymbol{m}(\mathrm{i}\eta)\| \le \frac{2}{\varepsilon} \tag{7.7.2}$$

holds true. Since $\zeta \in \mathbb{C}\setminus\widetilde{\mathbb{D}}_\varepsilon$ such an η can be chosen arbitrarily small. Then we have

$$\|\boldsymbol{m}(\mathrm{i}\eta)\| \,\lesssim\, \frac{1}{\varepsilon}\,, \qquad \|\boldsymbol{m}(\mathrm{i}\eta)^{-1}\| \,\lesssim\, \frac{1}{\varepsilon}\,, \qquad \eta \,\lesssim\, \operatorname{Im} m_i(\mathrm{i}\eta) \,\lesssim\, \frac{\eta}{\varepsilon}\,. \tag{7.7.3}$$

The first inequality follows from (7.7.1) and (7.7.2), the second inequality from (7.3.11c) and the third from (7.7.2) and the bounded support of v_i. In particular, by the formula (7.3.17) for the norm of $\boldsymbol{\mathcal{F}}$ we have

$$1 - \|\boldsymbol{\mathcal{F}}(\mathrm{i}\eta)\|_{\mathrm{sp}} \,\gtrsim\, \varepsilon^4\,. \tag{7.7.4}$$

To see (7.7.4) we simply follow the calculation in the proof of Lemma 7.3.6 but instead of using the bounds (7.3.11a), (7.3.11c) and (7.3.11b) on $\|\boldsymbol{m}\|$ and $\|\boldsymbol{m}^{-1}\|$ and $\operatorname{Im} m_i$ we use (7.7.3). Similarly we find

$$\|\mathcal{C}_{\boldsymbol{W}}\|\|\mathcal{C}_{\boldsymbol{W}}^{-1}\| \,\lesssim\, \frac{1}{\varepsilon^3}\,, \qquad \|\mathcal{C}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}\|\|\mathcal{C}_{\sqrt{\operatorname{Im}\boldsymbol{M}}}^{-1}\| \,\lesssim\, \frac{1}{\varepsilon}\,.$$

By (7.3.15) we conclude

$$\|\boldsymbol{\mathcal{L}}^{-1}\|_{\mathrm{sp}} \,\lesssim\, \frac{1}{\varepsilon^8}\,.$$

Using (7.3.23) and the bound on $\|\boldsymbol{m}\|$ in (7.7.3) we improve this bound on the $\|\cdot\|_{\mathrm{sp}}$-norm to a bound on the $\|\cdot\|$-norm,

$$\|\boldsymbol{\mathcal{L}}^{-1}\| \lesssim \frac{1}{\varepsilon^{12}}\,.$$

We are therefore in the linear stability regime of the Dyson equation and from the stability equation (cf. (7.3.14)) for the difference $\Delta := \boldsymbol{m}(z) - \boldsymbol{m}(\mathrm{i}\eta)$, i.e., from

$$\mathscr{L}[\Delta] \,=\, (z-\mathrm{i}\eta)\boldsymbol{m}(\mathrm{i}\eta)^2 + \frac{1}{2}\Big(\boldsymbol{m}(\mathrm{i}\eta)\mathscr{S}[\Delta]\Delta + \Delta\mathscr{S}[\Delta]\boldsymbol{m}(\mathrm{i}\eta)\Big)\,, \tag{7.7.5}$$

we infer

$$\|\boldsymbol{m}(z) - \boldsymbol{m}(\mathrm{i}\eta)\| \,\lesssim\, \|\boldsymbol{\mathcal{L}}^{-1}\|\|\boldsymbol{m}\|^2|z-\mathrm{i}\eta| \,\lesssim\, \frac{|z-\mathrm{i}\eta|}{\varepsilon^{14}}\,,$$

for any $z \in \mathbb{H}$ with

$$|z-\mathrm{i}\eta| \,\le\, \frac{C}{\|\boldsymbol{\mathcal{L}}^{-1}\|^2\|\boldsymbol{m}\|^3} \,\lesssim\, \varepsilon^{27}\,,$$

where $C \sim 1$ is a constant depending only on model parameters. Note that in (7.7.5) we symmetrized the quadratic term in Δ which can always be done since every other term of the equation is invariant under taking the Hermitian conjugate. In fact, we see that $\boldsymbol{m}$ can be extended analytically to an ε^{27}-neighborhood of $\mathrm{i}\eta$. Since η can be chosen arbitrarily small we find an analytic extension of $\boldsymbol{m}$ to all $z \in \mathbb{C}$ with $|z| \le c\varepsilon^{27}$ for some constant $c \sim 1$. We denote this extension by the same symbol $\boldsymbol{m} = (m_1, \ldots, m_N)$ as the solution to the Dyson equation. By definition of $\widetilde{\mathbb{D}}_\varepsilon$ we have $\operatorname{Im} m_i(0) = 0$ and it is easy to see by the following argument that for any $z \in \mathbb{R}$ the imaginary part still vanishes as long as we are in the linear stability regime. Thus $\rho^\zeta([-c\varepsilon^{27}, c\varepsilon^{27}]) = 0$: The stability equation (7.7.5) evaluated at $\eta = 0$ and $z \in \mathbb{R}$ is an equation on the space $\{\Delta \in (\mathbb{C}^{K\times K})^N : \Delta_i^* = \Delta_i,\, i = 1, \ldots, N\}$, i.e., for any Δ in this space both sides of the equation remain inside this space. Thus by the implicit function theorem applied within this subspace of $(\mathbb{C}^{K\times K})^N$ we conclude that the solution to (7.7.5) satisfies $\Delta = \Delta^*$, or equivalently $\operatorname{Im}\Delta = 0$, for $z \in \mathbb{R}$ inside the linear stability regime. Since $\rho^\zeta([-c\varepsilon^{27}, c\varepsilon^{27}]) = 0$ we thus obtain $\zeta \in \mathbb{C} \setminus \mathbb{D}_{c\varepsilon^{27}}$ which yields the missing inclusion. $\square$

7.8. Proofs of Theorem 7.2.7 and Lemma 7.4.8

For the reader's convenience, we now state and prove the local law for $\boldsymbol{H}$, Lemma 7.8.1 below. Its first part is designed for all spectral parameters z, where the Dyson equation, (7.3.26), is stable and its solution $\boldsymbol{m}$ is bounded; here the local law holds down to the scale $\eta = \operatorname{Im} z \geq N^{-1+\gamma}$ that is optimal near the self-consistent spectrum. The second part is valid away from the self-consistent spectrum; in this regime the Dyson equation is always stable and the local law holds down to the real line, however the dependence of our estimate on the distance from the spectrum is not optimized. For the proof of Lemma 7.4.8, the second part is sufficient, but we also give the first part for completeness. For simplicity we state the first part under the condition that $\boldsymbol{A} = \sum_i a_i \otimes E_{ii}$ is bounded; in the second part we relax this condition to include the assumptions of Lemma 7.4.8. From now on, we will also consider $\kappa_4, \ldots, \kappa_9$ from (7.4.41), (7.4.44a), (7.4.44b) and (7.8.1) below, respectively, as model parameters.

Lemma 7.8.1 (Local law)**.** *Fix $K \in \mathbb{N}$. Let $\boldsymbol{A} = \sum_{i=1}^N a_i \otimes E_{ii}$ be a deterministic Hermitian matrix. Let $\boldsymbol{H}$ be a Hermitian random matrix as in* (7.4.2) *satisfying Assumptions 7.4.1, i.e.,* (7.2.9)*,* (7.2.10) *and* (7.3.27) *hold true.*

(i) (Stable regime) Let $\gamma, \kappa_4, \kappa_5, \kappa_6 > 0$. Assume that $\|\boldsymbol{A}\|_2 \leq \kappa_4$ and define

$$\mathbb{H}_{\mathrm{stab}} := \Big\{ w \in \mathbb{H} \colon \ \sup_{s\geq 0} \|\boldsymbol{m}(w + \mathrm{i}s)\| \leq \kappa_5, \quad \sup_{s\geq 0} \|\mathscr{L}^{-1}(w+\mathrm{i}s)\|_{\mathrm{sp}} \leq \kappa_6 \quad \text{and } \operatorname{Im} w \geq N^{-1+\gamma} \Big\}. \tag{7.8.1}$$

Then, we have

$$\max_{i,j=1}^N |G_{ij}(z) - m_i(z)\delta_{ij}| \prec \frac{1}{1+\eta}\sqrt{\frac{\|\operatorname{Im} \boldsymbol{m}(z)\|}{N\eta}} + \frac{1}{(1+\eta^2)\sqrt{N}} + \frac{1}{(1+\eta^2)N\eta} \tag{7.8.2}$$

uniformly for $z \in \mathbb{H}_{\mathrm{stab}}$. Moreover, if $c_1, \ldots, c_N \in \mathbb{C}^{K\times K}$ are deterministic and satisfy $\max_{i=1}^N |c_i| \leq 1$ then we have

$$\Big| \frac{1}{N}\sum_{i=1}^N [c_i\,(G_{ii}(z) - m_i(z))] \Big| \prec \frac{1}{1+\eta}\Big(\frac{1}{N\eta} + \frac{1}{N}\Big) \tag{7.8.3}$$

uniformly for $z \in \mathbb{H}_{\mathrm{stab}}$.

(ii) (Away from the spectrum) Let $\kappa_7, \kappa_8, \kappa_9 > 0$ *be fixed. Assume that* (7.4.41) *holds true and* $\mathbb{H}_{\mathrm{out}}^{(1)}$ *and* $\mathbb{H}_{\mathrm{out}}^{(2)}$ *are defined as in* (7.4.44). *Then there are universal constants* $\delta > 0$ *and* $P \in \mathbb{N}$ *such that*

$$\max_{i,j=1}^{N} |G_{ij}(z) - m_i(z)\delta_{ij}| \prec \max\left\{\frac{1}{d_\rho^2(z)}, \frac{1}{d_\rho^P(z)}\right\}\frac{1}{\sqrt{N}} \tag{7.8.4}$$

uniformly for $z \in (\mathbb{H}_{\mathrm{out}}^{(1)} \cap \{w \in \mathbb{H} \colon d_\rho(w) \geq N^{-\delta}\}) \cup \mathbb{H}_{\mathrm{out}}^{(2)}$.

Moreover, if $c_1, \ldots, c_N \in \mathbb{C}^{K\times K}$ *are deterministic and satisfy* $\max_{i=1}^N |c_i| \leq 1$ *then we have*

$$\Big|\frac{1}{N}\sum_{i=1}^{N}\left[c_i\left(G_{ii}(z) - m_i(z)\right)\right]\Big| \prec \max\left\{\frac{1}{d_\rho^2(z)}, \frac{1}{d_\rho^P(z)}\right\}\frac{1}{N} \tag{7.8.5}$$

uniformly for $z \in (\mathbb{H}_{\mathrm{out}}^{(1)} \cap \{w \in \mathbb{H} \colon d_\rho(w) \geq N^{-\delta}\}) \cup \mathbb{H}_{\mathrm{out}}^{(2)}$.

The local laws (7.8.4) and (7.8.5) hold as stated with the alternative definitions of the sets $\mathbb{H}_{\mathrm{out}}^{(1)}$ and $\mathbb{H}_{\mathrm{out}}^{(2)}$ given after Lemma 7.4.8.

Proof of Theorem 7.2.7. Let $\boldsymbol{m}$ be the unique solution of (7.3.26) with positive imaginary part, where $\alpha_\mu := \widetilde{\alpha}_\mu$, $\beta_\nu := 2\widetilde{\beta}_\nu = \widetilde{\beta}_\nu + \widetilde{\gamma}_\nu^*$ and $a_j := \widetilde{a}_j$. Defining ρ_N as in (7.3.34), it is now a standard exercise to obtain (7.2.17) from (7.8.5), since $z \mapsto (NL)^{-1}\operatorname{Tr}((\boldsymbol{H}_N - z\mathbb{1})^{-1})$ is the Stieltjes transform of $\mu_{\boldsymbol{H}_N}$. □

Proof of Lemma 7.8.1. We start with the proof of part (i). For later use, we will present the proof for all spectral parameters z in a slightly larger set than $\mathbb{H}_{\mathrm{stab}}$, namely in the set

$$\mathbb{H}_{\mathrm{stab}}' := \Big\{w \in \mathbb{H} \colon \sup_{s\geq 0}(1 + \|\boldsymbol{A} - w - \mathrm{i}s\|_2)\|\boldsymbol{m}(w + \mathrm{i}s)\| \leq \kappa_5, \\ \sup_{s\geq 0}\|\mathscr{L}^{-1}(w+\mathrm{i}s)\|_{\mathrm{sp}} \leq \kappa_6 \text{ and } \operatorname{Im} w \geq N^{-1+\gamma}\Big\}. \tag{7.8.6}$$

Under the condition $\|\boldsymbol{A}\|_2 \leq \kappa_4$, it is easy to see $\mathbb{H}_{\mathrm{stab}} \subset \mathbb{H}_{\mathrm{stab}}'$ perhaps with somewhat larger κ-parameters. Furthermore, we relax the condition $\|\boldsymbol{A}\|_2 \leq \kappa_4$ to $\|\boldsymbol{A}\|_2 \leq N^{\kappa_7}$ with some positive constant κ_7. We also restrict our attention to the regime $|E| \leq N^{\kappa_7+1}$ since the complementary regime will be covered by the regime (7.4.44b) in part (ii). Let φ and ψ be defined as in part (iii) of Lemma 7.4.4 and recall the definition of ϑ from (7.4.28).

Proof of (7.8.2): We first show that

$$\Lambda(E + \mathrm{i}\eta) \prec \varphi \tag{7.8.7}$$

uniformly for $E + \mathrm{i}\eta \in \mathbb{H}'_{\mathrm{stab}}$ and $|E| \le N^{\kappa_7+1}$.

We start with some auxiliary estimates. By the definition of $\mathbb{H}'_{\mathrm{stab}}$ in (7.8.6) and setting $\boldsymbol{a} := (a_1, \ldots, a_N)$, we have

$$\|\boldsymbol{m}(z)\| \lesssim \frac{1}{1 + \|\boldsymbol{a} - z\mathbb{1}\|} \lesssim 1, \tag{7.8.8}$$

uniformly for $z \in \mathbb{H}'_{\mathrm{stab}}$. We remark that $\|\boldsymbol{a}\| = \|\boldsymbol{A}\|_2$.

We now verify that, uniformly for $z \in \mathbb{H}'_{\mathrm{stab}}$, we have

$$\|\boldsymbol{m}(z)\|\|\boldsymbol{m}^{-1}(z)\| \lesssim 1. \tag{7.8.9}$$

Applying $\|\cdot\|$ to (7.3.26) as well as using (7.3.35) and (7.8.8), we get that

$$\|\boldsymbol{m}^{-1}(z)\| \lesssim \|\boldsymbol{a} - z\mathbb{1}\| + 1 \lesssim 1 + |z| + \|\boldsymbol{a}\| \tag{7.8.10}$$

for $z \in \mathbb{H}'_{\mathrm{stab}}$. Thus, combining the first bounds in (7.8.8) and in (7.8.10) yields (7.8.9).

From the definition of $\mathbb{H}'_{\mathrm{stab}}$ in (7.8.6), using (7.8.8), (7.3.23) and (7.3.37), we obtain

$$\|\mathscr{L}^{-1}\| \lesssim 1, \qquad \|(\mathscr{L}^{-1})^*\| \lesssim 1, \tag{7.8.11}$$

where the adjoint is introduced above (7.4.34).

We will now use part (iii) of Lemma 7.4.4 to prove (7.8.7). To check the condition $\psi(\eta) \le N^{-\delta}$ in that lemma, we use (7.8.8), (7.8.11) and (7.8.9) to obtain $\psi(\eta) \lesssim 1/(N\eta)$. Hence, $\psi(\eta) \le N^{-\gamma/2}$ for $\eta \ge N^{-1+\gamma}$ and we choose $\delta = \gamma/2$ in (7.4.29).

We now estimate φ and ϑ in our setting. From (7.8.9), (7.8.8) and (7.8.11), we conclude that $\varphi \lesssim \|\boldsymbol{m}\|\Psi$, where we introduced the control parameter

$$\Psi := \sqrt{\frac{\|\mathrm{Im}\,\boldsymbol{m}\|}{N\eta}} + \frac{\|\boldsymbol{m}\|}{\sqrt{N}} + \frac{\|\boldsymbol{m}\|}{N\eta}.$$

We note that the factor $\|\boldsymbol{m}\|$ is kept in the bound $\varphi \lesssim \|\boldsymbol{m}\|\Psi$ and the definition of Ψ to control $\|\boldsymbol{m}^{-1}\|$ factors via (7.8.9) later and to track the correct dependence of the right-hand sides of (7.8.2) and (7.8.3) on η. For the second purpose, we will use the following estimate. Combined with (7.3.11a), the bound (7.8.8) yields

$$\|\boldsymbol{m}\| \lesssim \frac{1}{1+d_\rho(z)}. \tag{7.8.12}$$

For ϑ, we claim that

$$\vartheta \gtrsim (1+|z|+\|\boldsymbol{a}\|)^{-1}, \qquad \vartheta \gtrsim \|\boldsymbol{m}\|. \tag{7.8.13}$$

Indeed, for the first bound, we apply (7.3.35), (7.8.8), (7.8.11) and the second bound in (7.8.10) to the definition of ϑ, (7.4.28). Using (7.8.9) instead of (7.8.8) and (7.8.10) yields the second bound.

Now, to prove (7.8.7), we show that $\mathbf{1}(\Lambda \leq \vartheta) = 1$ a.w.o.p. for $\eta \geq N^{-1+\gamma}$ on the left-hand side of (7.4.29). The first step is to establish $\Lambda \leq \vartheta$ for large η. For $\eta \geq \max\{1, |E|, \|\boldsymbol{A}\|_2\}$, we have $\Lambda \prec \eta^{-2}$ by (7.4.26). By (7.8.13), we have $\vartheta \gtrsim \eta^{-1}$ for $\eta \geq \max\{1, |E|, \|\boldsymbol{A}\|_2\}$. Therefore, there is $\kappa > \kappa_7 + 1$ such that $\Lambda(\eta) \leq \vartheta(\eta)$ a.w.o.p. for all $\eta \geq N^\kappa$. Together with (7.4.29), this proves (7.8.7) for $\eta \geq N^\kappa$.

The second step is a stochastic continuity argument to reduce η for the domain of validity of (7.8.7). The estimate (7.4.29) asserts that Λ cannot take on any value between φ and ϑ with very high probability. Since $\eta \mapsto \Lambda(\eta)$ is continuous, Λ remains bounded by φ for all values of η as long as φ is smaller than ϑ. The precise formulation of this procedure is found e.g. in Lemma A.2 of [7] and we leave the straightforward check of its conditions to the reader. The bound (7.8.7) yields (7.8.2) in the regime $|E| \leq N^{\kappa_7+1}$.

Proof of (7.8.3): We apply Lemma 7.4.5 with $\Phi := \|\boldsymbol{m}^{-1}\|\varphi$. The condition (7.4.32) is satisfied by the definition of Φ and (7.8.7). Since $\Phi \lesssim \Psi$ it is easily checked that all terms on the right-hand side of (7.4.33) are bounded by $\|\boldsymbol{m}\| \max\{N^{-1/2}, \Psi\}\Psi$. Therefore, using (7.8.11) and (7.8.12), the averaged local law, (7.4.33), yields

$$\begin{aligned}\Big|\frac{1}{N}\sum_{i=1}^N c_i(G_{ii}-m_i)\Big| &\prec \|\boldsymbol{m}\| \max\Big\{\frac{1}{\sqrt{N}}, \Psi\Big\}\Psi \\ &\lesssim \frac{1}{1+d_\rho(z)}\Big(\frac{\|\operatorname{Im}\boldsymbol{m}(z)\|}{N\eta} + \frac{1}{N} + \frac{1}{1+d_\rho^2(z)}\frac{1}{(N\eta)^2}\Big)\end{aligned} \tag{7.8.14}$$

for any $c_1, \ldots, c_N \in \mathbb{C}^{K\times K}$ such that $\max_i |c_i| \leq 1$. Owing to $\|\operatorname{Im}\boldsymbol{m}\| \lesssim 1$ by (7.8.8), the bound (7.8.3) follows.

We now turn to the proof of (ii) which is divided into two steps. In the first step, we show Lemma 7.4.8. Therefore, we will follow the proof of (7.8.14) with the bounds (7.8.12) and (7.8.11) replaced by their weaker analogues (7.8.15) and (7.8.16) below that deteriorate as $d_\rho(z)$ becomes small. After having completed Lemma 7.4.8, we immediately get Theorem 7.4.7 via the proof given in Section 7.4.2. Finally, in the second step, proceeding similarly as in the proof of (i), the bounds (7.8.4) and (7.8.5) will be obtained from Theorem 7.4.7.

Step 1: Proof of Lemma 7.4.8. We first give the replacements for the bounds (7.8.12) and (7.8.11) that served as inputs for the previous proof of part (i). The replacement for (7.8.12) is a direct consequence of (7.3.11a):

$$\|\boldsymbol{m}\| \leq \frac{1}{d_\rho(z)}. \tag{7.8.15}$$

The replacement of (7.8.11) is the bound

$$\|\mathscr{L}^{-1}\| + \|(\mathscr{L}^{-1})^*\| \lesssim 1 + \frac{1}{d_\rho^{26}(z)}, \tag{7.8.16}$$

which is obtained by distinguishing the regimes $\|\boldsymbol{M}\|_2^2\|\boldsymbol{\mathcal{S}}\| > 1/2$ and $\|\boldsymbol{M}\|_2^2\|\boldsymbol{\mathcal{S}}\| \leq 1/2$. In the first regime, we conclude from (7.3.22) and (7.3.23) that

$$\|\boldsymbol{\mathcal{L}}^{-1}\| + \|(\boldsymbol{\mathcal{L}}^{-1})^*\| \lesssim 1 + \|\boldsymbol{M}\|_2^2 + \frac{\|\boldsymbol{M}\|_2^9\|\boldsymbol{M}^{-1}\|_2^9}{\|\boldsymbol{M}\|_2^4 d_\rho^8(z)} \lesssim 1 + \frac{1}{d_\rho^{26}(z)},$$

where we used the lower bound on $\boldsymbol{M}$ given by the definition of the regime and $\|\boldsymbol{\mathcal{S}}\| \lesssim 1$ as well as the bound $\|\boldsymbol{M}\|_2\|\boldsymbol{M}^{-1}\|_2 \lesssim 1/d_\rho^2(z)$ that is proven as (7.8.17) below. In the second case, we use the simple bound $\|\boldsymbol{\mathcal{L}}^{-1}\| + \|(\boldsymbol{\mathcal{L}}^{-1})^*\| \leq 2/(1 - \|\boldsymbol{M}\|_2^2\|\boldsymbol{\mathcal{S}}\|) \leq 4$. Thus, (7.3.37) yields (7.8.16).

Next, we will check that the following weaker version of (7.8.9) holds

$$\|\boldsymbol{m}(z+\mathrm{i}s)\|\|\boldsymbol{m}^{-1}(z+\mathrm{i}s)\| \lesssim 1 + \frac{1}{d_\rho^2(z+\mathrm{i}s)} \tag{7.8.17}$$

for all $z \in \mathbb{H}_{\mathrm{out}}^{(1)} \cup \mathbb{H}_{\mathrm{out}}^{(2)}$ and $s \geq 0$. This is straightforward for $z \in \mathbb{H}_{\mathrm{out}}^{(1)}$ since in this case $|z|$, $\|\boldsymbol{A}\|_2$ and $\operatorname{supp}\rho$ all remain bounded (see (7.3.32a)), so similarly to (7.8.10) we have $\|\boldsymbol{m}^{-1}(z+\mathrm{i}s)\| \lesssim 1 + s + \|\boldsymbol{m}(z+\mathrm{i}s)\|$. For $|s| \leq C$ (7.8.17) directly follows from (7.8.15), while for large s we have $\|\boldsymbol{m}(z+\mathrm{i}s)\| \lesssim s^{-1}$ and $\|\boldsymbol{m}^{-1}(z+\mathrm{i}s)\| \lesssim s$, so (7.8.17) also holds.

Suppose now that $z \in \mathbb{H}^{(2)}_{\mathrm{out}}$. In this regime z is far away from the spectrum of $\boldsymbol{A}$, so by (7.3.32a) we know that $\operatorname{dist}(z+\mathrm{i}s, \operatorname{Spec}\boldsymbol{A}) \sim \operatorname{dist}(z+\mathrm{i}s, \operatorname{supp}\rho) \geq 1$. This means that

$$\|\boldsymbol{m}(z+\mathrm{i}s)\| \lesssim \frac{1}{\operatorname{dist}(z+\mathrm{i}s, \operatorname{supp}\rho)} \sim \frac{1}{\operatorname{dist}(z+\mathrm{i}s, \operatorname{Spec}\boldsymbol{A})} = \frac{1}{\sigma_{\min}(\boldsymbol{A}-(z+\mathrm{i}s)\mathbb{1})}, \tag{7.8.18}$$

and hence from the Dyson equation

$$\Big\|\frac{1}{\boldsymbol{m}(z+\mathrm{i}s)}\Big\| \leq \|\boldsymbol{A}-(z+\mathrm{i}s)\mathbb{1}\|_2 + \|\boldsymbol{\mathcal{S}}\| \lesssim \|\boldsymbol{A}-(z+\mathrm{i}s)\mathbb{1}\|_2. \tag{7.8.19}$$

Since $\boldsymbol{A}$ is Hermitian, we have the bound

$$\frac{\|\boldsymbol{A}-(z+\mathrm{i}s)\mathbb{1}\|_2}{\sigma_{\min}(\boldsymbol{A}-(z+\mathrm{i}s)\mathbb{1})} \leq \frac{\|\boldsymbol{A}-z\mathbb{1}\|_2}{\sigma_{\min}(\boldsymbol{A}-z\mathbb{1})} \leq \kappa_9 \tag{7.8.20}$$

for any $s \geq 0$, where the first inequality comes from the spectral theorem and the second bound is from the definition of $\mathbb{H}^{(2)}_{\mathrm{out}}$. Therefore $\sigma_{\min}(\boldsymbol{A}-(z+\mathrm{i}s)\mathbb{1}) \sim \|\boldsymbol{A}-(z+\mathrm{i}s)\mathbb{1}\|_2$, and thus (7.8.17) follows from (7.8.18) and (7.8.19).

Now we can complete Step 1 by following the proof of part (i) but using (7.8.15), (7.8.16) and (7.8.17) instead of (7.8.12), (7.8.11) and (7.8.9), respectively. It is easy to see that only these three estimates on $\|\boldsymbol{m}\|$, $\|\boldsymbol{m}\|\|\boldsymbol{m}^{-1}\|$ and $\|\mathscr{L}^{-1}\|$ were used as inputs in this argument. The resulting estimates are weaker by multiplicative factors involving certain power of $1+1/d_\rho(z)$. We thus obtain a version of (7.8.14) for $\eta \geq N^{-1+\gamma}(1+d_\rho^{-p}(z))$ with $(1+d_\rho(z))^{-1}$ replaced by $\max\{1, d_\rho^{-P}(z)\}$ for some explicit $p, P \in \mathbb{N}$. Thus, applying (7.3.11b) to estimate $\operatorname{Im}\boldsymbol{m}$ in (7.8.14) instead of $\|\operatorname{Im}\boldsymbol{m}\| \lesssim 1$ and possibly increasing P yields (7.4.45). $\square$

Step 2: Continuing the proof of part (ii) of Lemma 7.8.1, we draw two consequences from Theorem 7.4.7 and the fact that $\boldsymbol{G}$ is the Stieltjes transform of a positive semidefinite matrix-valued measure $V_{\boldsymbol{G}}$ supported on $\operatorname{Spec}\boldsymbol{H}$ with $V_{\boldsymbol{G}}(\operatorname{Spec}\boldsymbol{H}) = \mathbb{1}$. Let $\delta > 0$ be chosen as in Theorem 7.4.7. Since the spectrum of $\boldsymbol{H}$ is contained in

$\{\omega \in \mathbb{R}\colon\ \mathrm{dist}(\omega, \operatorname{supp}\rho) \leq N^{-\delta}\}$ a.w.o.p. by Theorem 7.4.7, we have

$$\|\boldsymbol{G}\|_2 \lesssim \frac{1}{d_\rho(z)}, \qquad \operatorname{Im}\boldsymbol{G} \lesssim \frac{\eta}{d_\rho^2(z)}\mathbb{1}$$

a.w.o.p. for all $z \in \mathbb{H}$ satisfying $d_\rho(z) \geq N^{-\delta/2}$. Therefore, (7.4.30) implies for all $z \in \mathbb{H}$ satisfying $d_\rho(z) \geq N^{-\delta/2}$ that

$$\Lambda_{\mathrm{hs}} + \Lambda_{\mathrm{w}} \prec \frac{1}{d_\rho(z)\sqrt{N}}. \tag{7.8.21}$$

Since $\boldsymbol{M}$ is the Stieltjes transform of $V_{\boldsymbol{M}}$ defined in (7.3.33) and $V_{\boldsymbol{M}}(\mathbb{R}) = \mathbb{1}$ and $\boldsymbol{G}$ is the Stieltjes transform of $V_{\boldsymbol{G}}$ we conclude that there is $\kappa > 0$ such that

$$\Lambda \lesssim \|\boldsymbol{G} - \boldsymbol{M}\|_2 \lesssim |z|^{-2} \tag{7.8.22}$$

a.w.o.p. uniformly for all $z \in \mathbb{H}$ satisfying $|z| \geq N^\kappa$. Here, we used that $\operatorname{supp} V_{\boldsymbol{M}} \subset \operatorname{supp}\rho$ and hence $\mathrm{diam}(\operatorname{supp} V_{\boldsymbol{M}}) \lesssim N^{\kappa_7+1}$ by (7.4.41) and (7.3.32a) as well as $\mathrm{diam}(\operatorname{supp} V_{\boldsymbol{G}}) \leq \mathrm{diam}(\operatorname{Spec}\boldsymbol{H}) \lesssim N^{\kappa_7+1}$ a.w.o.p. by Theorem 7.4.7.

Hence, owing to (7.8.13) and (7.8.22), by possibly increasing $\kappa > 0$, we can assume that $\Lambda \leq \vartheta$ a.w.o.p. for all $z \in \mathbb{H}_{\mathrm{out}}^{(1)} \cup \mathbb{H}_{\mathrm{out}}^{(2)}$ satisfying $|z| \geq N^\kappa$. Thus, to estimate $\|\boldsymbol{g} - \boldsymbol{m}\|$ we start from (7.4.27) and use (7.8.16), (7.8.15), (7.8.21) and (7.8.9) to obtain an explicit $P \in \mathbb{N}$ such that $\|\boldsymbol{g} - \boldsymbol{m}\| \prec \|\boldsymbol{m}\| \max\{d_\rho^{-1}(z), d_\rho^{-P}(z)\}N^{-1/2}$ a.w.o.p. For the offdiagonal terms of $\boldsymbol{G}$, we apply (7.8.21) to (7.4.13). This yields

$$\Lambda \prec \|\boldsymbol{m}\| \max\left\{\frac{1}{d_\rho(z)}, \frac{1}{d_\rho^P(z)}\right\}\frac{1}{\sqrt{N}} \tag{7.8.23}$$

for $z \in \mathbb{H}_{\mathrm{out}}^{(1)} \cup \mathbb{H}_{\mathrm{out}}^{(2)}$ satisfying $|z| \geq N^\kappa$. Employing the stochastic continuity argument from Lemma A.2 in [7] as before, we obtain (7.8.23) for all $z \in \mathbb{H}_{\mathrm{out}}^{(1)} \cup \mathbb{H}_{\mathrm{out}}^{(2)}$ satisfying $d_\rho(z) \geq N^{-\delta/2}$. We use (7.8.15) in (7.8.23), replace P by $P+1$ and δ by $\delta/2$. Thus, we have proven (7.8.4) for all $z \in \mathbb{H}_{\mathrm{out}}^{(1)} \cup \mathbb{H}_{\mathrm{out}}^{(2)}$ satisfying $d_\rho(z) \geq N^{-\delta}$. Notice that this argument covers the case $|E| \geq N^{\kappa_7+1}$ as well that was left open in Step 1.

For the proof of (7.8.5), we set $\Phi := (d_\rho(z)\sqrt{N})^{-1}$ and apply Lemma 7.4.5. Its assumption $\Lambda \prec \Phi/\|\boldsymbol{m}^{-1}\|$ is satisfied by (7.8.23) and (7.8.9). Using (7.8.16), (7.8.15), (7.8.9) and (7.8.21), this proves (7.8.5) and hence concludes the proof of Lemma 7.8.1. □

CHAPTER 8

The Dyson equation with linear self-energy: spectral bands, edges and cusps

The current chapter contains the preprint [15] which is joint work with László Erdős and Torben Krüger. We study the unique solution m of the Dyson equation

$$-m(z)^{-1} = z\mathbb{1} - a + S[m(z)]$$

on a von Neumann algebra $\mathcal{A}$ with the constraint $\operatorname{Im} m \geq 0$. Here, z lies in the complex upper half-plane, a is a self-adjoint element of $\mathcal{A}$ and S is a positivity-preserving linear operator on $\mathcal{A}$. We show that m is the Stieltjes transform of a compactly supported $\mathcal{A}$-valued measure on $\mathbb{R}$. Under suitable assumptions, we establish that this measure has a uniformly 1/3-Hölder continuous density with respect to the Lebesgue measure, which is supported on finitely many intervals, called bands. In fact, the density is analytic inside the bands with a square-root growth at the edges and internal cubic root cusps whenever the gap between two bands vanishes. The shape of these singularities is universal and no other singularity may occur. We give a precise asymptotic description of m near the singular points. These asymptotics play a key role in Chapter 9 below, where the Tracy-Widom universality for the edge eigenvalue statistics for correlated random matrices is proven. We also show that the spectral mass of the bands is topologically rigid under deformations and we conclude that these masses are quantized in some important cases.

8.1. Introduction

An important task in random matrix theory is to determine the eigenvalue distribution of a random matrix as its size tends to infinity. Similarly, in free probability theory, the scalar-valued distribution of operator-valued semicircular elements is of particular interest. In both cases, the distribution can be obtained from a *Dyson equation*

$$-m(z)^{-1} = z\mathbb{1} - a + S[m(z)] \tag{8.1.1}$$

on some von Neumann algebra $\mathcal{A}$ with a unit $\mathbb{1}$ and a tracial state $\langle\,\cdot\,\rangle$. Here, $z \in \mathbb{H} := \{w \in \mathbb{C} \colon \operatorname{Im} w > 0\}$, $a = a^* \in \mathcal{A}$ and $S \colon \mathcal{A} \to \mathcal{A}$ is a positivity-preserving linear operator. There is a unique solution $m \colon \mathbb{H} \to \mathcal{A}$ of (8.1.1) under the assumption that $\operatorname{Im} m(z) := (m(z) - m(z)^*)/(2\mathrm{i})$ is a strictly positive element of $\mathcal{A}$ for all $z \in \mathbb{H}$ [96]. For suitably chosen a and S as well as $\mathcal{A}$, this solution characterizes the distributions in the applications mentioned above. In fact, in both cases, the distribution will be the measure ρ on $\mathbb{R}$ whose Stieltjes transform is given by $z \mapsto \langle m(z)\rangle$. The measure ρ is called the *self-consistent density of states* and its support is the *self-consistent spectrum*. This terminology stems from the physics literature on the Dyson equation, where z is often called *spectral parameter* and S is the *self-energy operator*. The linearity of S is a distinctive feature of our setup.

We first explain the connection between the eigenvalue density of a large random matrix and the Dyson equation in more detail. Let $H \in \mathbb{C}^{n\times n}$ be a $\mathbb{C}^{n\times n}$-valued random variable, $n \in \mathbb{N}$, such that $H = H^*$. A central objective is now the analysis of the *empirical spectral measure* $\mu_H := n^{-1}\sum_{i=1}^n \delta_{\lambda_i}$, or its expectation, the *density of states*, for large n, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of H. An easy computation shows that $n^{-1}\operatorname{Tr}(H-z)^{-1}$ is the Stieltjes transform of μ_H at $z \in \mathbb{H}$. Therefore, the resolvent $(H-z)^{-1}$ is commonly studied to obtain information about μ_H. In fact, for many random matrix ensembles, it turns out that the resolvent $(H-z)^{-1}$ is well approximated for large n by the solution $m(z)$ of the Dyson equation (8.1.1). Here, we choose $\mathcal{A} = \mathbb{C}^{n\times n}$ equipped with the operator norm induced by the Euclidean distance on $\mathbb{C}^n$ and the normalized trace $\langle\,\cdot\,\rangle = n^{-1}\operatorname{Tr}(\,\cdot\,)$ as tracial state as well as

$$a := \mathbb{E}H, \qquad S[x] := \mathbb{E}[(H-a)x(H-a)], \quad x \in \mathbb{C}^{n\times n}. \tag{8.1.2}$$

If $(H-z)^{-1}$ is well approximated by $m(z)$ for large n then μ_H will be well approximated by the deterministic measure ρ, whose Stieltjes transform is given by $z \mapsto \langle m(z)\rangle$. The importance of the Dyson equation (8.1.1) for random matrix theory has been realized by many authors on various levels of generality [20, 34, 84, 99, 131, 156], see also the monographs [82, 119] and the more recent works [6, 7, 56, 94, 101] as well as Chapters 4, 5 and 7.

Secondly, we relate the Dyson equation to free probability theory by noticing that the Cauchy transform of a shifted operator-valued semicircular element is given by m. More precisely, let $\mathcal{B}$ be a unital C^*-algebra, $\mathcal{A} \subset \mathcal{B}$ be a C^*-subalgebra with the same unit $\mathbb{1}$ and $E\colon \mathcal{B} \to \mathcal{A}$ is a conditional expectation (we refer to Chapter 9 in [115] for notions from free probability theory). Pick an $a = a^* \in \mathcal{A}$ and an operator-valued semicircular element $s = s^* \in \mathcal{B}$ then $G(z) := E[(z - s - a)^{-1}]$ is the *Cauchy-transform* of $s + a$. In this case, $m(z) = -G(z)$ satisfies (8.1.1) with $S[x] := E[sxs]$ for all $x \in \mathcal{A}$ [154]. If $\mathcal{A}$ is a von Neumann algebra with a tracial state, then our results yield information about the scalar-valued distribution $\rho = \rho_{s+a}$ of $s + a$ with respect to this state. The study of qualitative regularity properties for this distribution has a long history in free probability. For example, the question of whether ρ has atoms or not is intimately related to noncommutative identity testing (see [79, 110] and references therein) and the notions of free entropy and Fischer information (see [151, 152] and the survey [153]). We also refer to the recent preprint [111], where the distribution of rational functions in noncommutative random variables is studied with the help of linearization ideas from [86, 87] and [95]. Under strong assumptions, our results provide extremely detailed information about the regularity properties of ρ, thus complementing these more general insights. In particular, we show that ρ_s is absolutely continuous with respect to the Lebesgue measure away from zero for any operator-valued semicircular element s. For other applications of the Dyson equation (8.1.1) in free probability theory, we refer to [96, 137, 154, 155] and the recent monograph [115].

In this paper, we analyze the regularity properties of the self-consistent density of states ρ in detail. More precisely, under suitable assumptions on S, we show that the boundedness of m already implies that ρ has a 1/3-Hölder continuous density $\rho(\tau)$ with respect to the Lebesgue measure. We provide a broad class of models for which the boundedness of m is ensured. Furthermore, the set where the density is positive, $\{\tau : \rho(\tau) > 0\}$, splits into finitely many connected components, called *bands*. The density is real-analytic inside the bands with a square root growth behavior at the edges. If two bands touch, however, a cubic root cusp emerges. These are the only possible types of singularities. In fact, $m(z)$ is the Stieltjes transform of a positive operator-valued measure v and we establish the properties mentioned above for v as well. We also provide a novel formula for the masses that ρ assigns to the bands. We use it to infer a certain quantization of the band masses that we call *band rigidity*, because it is invariant under small perturbations of the data a and S of the Dyson equation. In particular, we extend a quantization result from [86] and [132] to cover limits of Kronecker random matrices. We remark that in the context of random matrices the analogous phenomenon was coined as "exact separation of eigenvalues" in [23].

In the commutative setup, the band structure and singularity behavior of the density have been obtained in [4, 5], where a detailed analysis of the regularity of ρ was initiated. In the special noncommutative situation $\mathcal{A} = \mathbb{C}^{n\times n}$ and $\langle \cdot \rangle = n^{-1}\operatorname{Tr}(\cdot)$, it has been shown that ρ is Hölder-continuous and real-analytic wherever it is positive [6]. The main novelty of the current work is to give an effective regularity analysis for the general noncommutative case, including a precise description of all singularities, i.e., edges and cusps. One of the main applications is the proof of the eigenvalue rigidity on optimal scale and the Tracy-Widom universality of the local spectral statistics near the spectral edges for random matrices with general correlation structure (cf. Chapter 9 below).

The key strategy behind the current paper as well as its predecessors [4, 5, 6] is a refined stability analysis of the Dyson equation (8.1.1) against small perturbations. It turns out that the equation is stable in the bulk regime, i.e., where $\rho(\operatorname{Re} z)$ is separated away from zero, but is unstable near the points, where the density vanishes. Even the stability in the bulk requires an unconventional idea; it relies on rewriting the stability operator, i.e., the derivative of the Dyson equation with respect to m, through the use of a positivity-preserving symmetric map, called the *saturated self-energy operator*, F. We then extract information on the spectral gap of F by a Perron-Frobenius argument using the positivity of $\operatorname{Im} m$ [4, 5]. In the noncommutative setup this transformation was based on a novel *balanced polar decomposition* formula [6]. In the small density regime, in particular near the edges, the stability deteriorates due to an *unstable direction*, which is related to the Perron-Frobenius eigenvector of F. The analysis boils down to a scalar quantity, Θ, the overlap between the solution and the unstable direction. For the commutative case in [4, 5], it is shown that Θ approximately satisfies a cubic equation. The structural property of this cubic equation is its *stability*, i.e., that the coefficients of the cubic and quadratic terms do not simultaneously vanish. This guarantees that higher order terms are negligible and the order of any singularity is either cubic root or square root.

Now we synthesize both analyses in the previous works to study the small density regime in the most general setup. The major obstacle is the noncommutativity that already substantially complicated the bulk analysis in [6] but there the saturated self-energy operator, F, governed all estimates. However, near the edges the unstable direction is identified via the top eigenvector of a non-symmetric operator that coincides with the symmetric F only in the commutative case. Thus we need to perform a non-symmetric perturbation expansion that requires precise control on the resolvent of the non-self-adjoint stability operator in the entire complex plane. We still work with a cubic equation for Θ, but the analysis of its coefficients is considerably more involved. Along all estimates, the noncommutativity is a permanent enemy; in some cases it can be treated perturbatively, but for the most critical parts new non-perturbative proofs are needed. Most critically, the stability of the cubic equation is proven with a new method.

Another novelty of the current paper, in addition to handling the noncommutativity and lack of symmetry, is that we present the cubic analysis in a conceptually clean way that will be used in future works. Our analysis strongly suggests that our cubic equation for Θ is the key to any detailed singularity analysis of Dyson-type equations and its remarkable structure is responsible for the universal behavior of the singularities in the density.

8.2. Main results

Let $\mathcal{A}$ be a finite von Neumann algebra with unit $\mathbb{1}$ and norm $\|\cdot\|$. We recall that a von Neumann algebra $\mathcal{A}$ is called *finite* if there is a state $\langle\,\cdot\,\rangle\colon \mathcal{A} \to \mathbb{C}$ which is (i) *tracial*, i.e., $\langle xy\rangle = \langle yx\rangle$ for all $x, y \in \mathcal{A}$, (ii) *faithful*, i.e., $\langle x^*x\rangle = 0$ for some $x \in \mathcal{A}$ implies $x = 0$, and (iii) *normal*, i.e., continuous with respect to the weak* topology. In the following,

$\langle \cdot \rangle$ will always denote such state. The tracial state defines a scalar product $\mathcal{A} \times \mathcal{A} \to \mathbb{C}$ through

$$\langle x\,, y\rangle \;:=\; \langle x^* y\rangle \tag{8.2.1}$$

for $x, y \in \mathcal{A}$. The induced norm is denoted by $\|x\|_2 := \langle x\,, x\rangle^{1/2}$ for $x \in \mathcal{A}$. Clearly, $\|x\|_2 \leq \|x\|$ for all $x \in \mathcal{A}$. We follow the convention that small letters are elements of $\mathcal{A}$ while capital letters denote linear operators on $\mathcal{A}$. The spectrum of $x \in \mathcal{A}$ is denoted by $\operatorname{Spec} x$, i.e., $\operatorname{Spec} x = \mathbb{C} \setminus \{z \in \mathbb{C} \colon (x - z)^{-1} \in \mathcal{A}\}$.

For an operator $T \colon \mathcal{A} \to \mathcal{A}$, we will work with three norms. We denoted these norms by $\|T\|$, $\|T\|_2$ and $\|T\|_{2\to\|\cdot\|}$ if T is considered as an operator $(\mathcal{A}, \|\cdot\|) \to (\mathcal{A}, \|\cdot\|)$, $(\mathcal{A}, \|\cdot\|_2) \to (\mathcal{A}, \|\cdot\|_2)$ or $(\mathcal{A}, \|\cdot\|_2) \to (\mathcal{A}, \|\cdot\|)$, respectively.

We denote by $\mathcal{A}_{\mathrm{sa}}$ the self-adjoint elements of $\mathcal{A}$, by $\mathcal{A}_+$ the cone of positive definite elements of $\mathcal{A}$, i.e.,

$$\mathcal{A}_{\mathrm{sa}} \;:=\; \{x \in \mathcal{A} : x^* = x\}, \qquad \mathcal{A}_+ \;:=\; \{x \in \mathcal{A}_{\mathrm{sa}} \colon x > 0\},$$

and by $\overline{\mathcal{A}}_+$, the $\|\cdot\|$-closure of $\mathcal{A}_+$, the cone of positive semidefinite elements (or positive elements). We now introduce two classes of linear operators on $\mathcal{A}$ that preserve the cone $\overline{\mathcal{A}}_+$. Such operators are called *positivity-preserving* (or *positive maps*). We define

$$\Sigma \;:=\; \Big\{ S \colon \mathcal{A} \to \mathcal{A} : S \text{ is linear, symmetric wrt. (8.2.1) and } S[\overline{\mathcal{A}}_+] \subset \overline{\mathcal{A}}_+ \Big\}, \tag{8.2.2a}$$

$$\Sigma_{\mathsf{flat}} \;:=\; \Big\{ S \in \Sigma : \varepsilon \mathbb{1} \leq \inf_{x \in \mathcal{A}_+} \frac{S[x]}{\langle x \rangle} \leq \sup_{x \in \mathcal{A}_+} \frac{S[x]}{\langle x \rangle} \leq \varepsilon^{-1} \mathbb{1} \text{ for some } \varepsilon > 0 \Big\}. \tag{8.2.2b}$$

Moreover, if $S \colon \mathcal{A} \to \mathcal{A}$ is a positivity-preserving operator, then S is bounded, i.e., $\|S\|$ is finite (see e.g. Proposition 2.1 in [120]).

Let $a \in \mathcal{A}_{\mathrm{sa}}$ be a self-adjoint element and $S \in \Sigma$. For the *data pair* (a, S), we consider the associated *Dyson equation*

$$-\, m(z)^{-1} \;=\; z\mathbb{1} - a + S[m(z)]\,, \tag{8.2.3}$$

with spectral parameter $z \in \mathbb{H} := \{w \in \mathbb{C} \colon \operatorname{Im} w > 0\}$, for a function $m \colon \mathbb{H} \to \mathcal{A}$ such that its imaginary part is positive definite,

$$\operatorname{Im} m(z) \;=\; \frac{1}{2\mathrm{i}}(m(z) - m(z)^*) \;\in\; \mathcal{A}_+\,.$$

There always exists a unique solution m to the Dyson equation (8.2.3) satisfying $\operatorname{Im} m(z) \in \mathcal{A}_+$ [96]. Moreover, this solution is holomorphic in z [96]. For Dyson equations in the context of renormalization theory, a is called the *bare matrix* and S the *self-energy (operator)*. In applications to free probability theory, S is usually denoted by η and called the *covariance mapping* or *covariance matrix* [115].

We now introduce positive operator-valued measures with values in $\overline{\mathcal{A}}_+$. If v maps Borel sets on $\mathbb{R}$ to elements of $\overline{\mathcal{A}}_+$ such that $\langle x\,, v(\,\cdot\,)x\rangle$ is a positive measure for all $x \in \mathcal{A}$ then we say that v is a *measure on* $\mathbb{R}$ *with values in* $\overline{\mathcal{A}}_+$ or an $\overline{\mathcal{A}}_+$*-valued measure on* $\mathbb{R}$.

First, we list a few propositions that are necessary to state our main theorem. They will be proven in Section 8.3, Section 8.4.2 and Section 8.4.3, respectively.

Proposition 8.2.1 (Stieltjes transform representation)**.** *Let* $(a, S) \in \mathcal{A}_{\mathrm{sa}} \times \Sigma$ *be a data pair and* m *the solution to the associated Dyson equation. Then there exists a measure* v *on* $\mathbb{R}$ *with values in* $\overline{\mathcal{A}}_+$ *such that* $v(\mathbb{R}) = \mathbb{1}$ *and*

$$m(z) \;=\; \int_{\mathbb{R}} \frac{v(\mathrm{d}\tau)}{\tau - z} \tag{8.2.4}$$

for all $z \in \mathbb{H}$*. The support of* v *and the spectrum of* a *satisfy the following inclusions*

$$\operatorname{supp} v \subset \operatorname{Spec} a + [-2\|S\|^{1/2}, 2\|S\|^{1/2}], \tag{8.2.5a}$$

$$\operatorname{Spec} a \subset \operatorname{supp} v + [-\|S\|^{1/2}, \|S\|^{1/2}]. \tag{8.2.5b}$$

Furthermore, if $a = 0$ *then, for any* $z \in \mathbb{H}$*,* $m(z)$ *satisfies the bound*

$$\|m(z)\|_2 \;\leq\; \frac{2}{|z|}\,. \tag{8.2.6}$$

Our goal is to obtain regularity results for the measure v. We first present some regularity results on the self-consistent density of states introduced in the following definition.

Definition 8.2.2 (Density of states)**.** Let $(a, S) \in \mathcal{A}_{\mathrm{sa}} \times \Sigma$ be a data pair, m the solution to the associated Dyson equation, (8.2.3), and v the $\overline{\mathcal{A}}_+$-valued measure of Proposition 8.2.1. The positive measure $\rho = \langle v \rangle$ on $\mathbb{R}$ is called the *self-consistent density of states* or short *density of states*.

We have $\operatorname{supp}\rho = \operatorname{supp} v$ due to the faithfulness of $\langle\,\cdot\,\rangle$. Moreover, the Stieltjes transform of ρ is given by $\langle m\rangle$ since, by (8.2.3), for any $z \in \mathbb{H}$, we have

$$\langle m(z)\rangle \,=\, \int_{\mathbb{R}} \frac{\rho(\mathrm{d}\tau)}{\tau - z}.$$

Proposition 8.2.3 (Regularity of density of states)**.** *Let (a, S) be a data pair with $S \in \Sigma_{\mathrm{flat}}$ and $\rho_{a,S}$ the corresponding density of states. Then $\rho_{a,S}$ has a uniformly Hölder-continuous, compactly supported density with respect to the Lebesgue measure,*

$$\rho_{a,S}(\mathrm{d}\tau) \,=\, \rho_{a,S}(\tau)\mathrm{d}\tau\,.$$

Furthermore, there exists a universal constant $c > 0$ such that the function $\rho\colon \mathcal{A}_{\mathrm{sa}} \times \Sigma_{\mathrm{flat}} \times \mathbb{R} \to [0,\infty), (a, S, \tau) \mapsto \rho_{a,S}(\tau)$ is locally Hölder-continuous with Hölder exponent c and analytic whenever it is positive, i.e., for any $(a, S, \tau) \in \mathcal{A}_{\mathrm{sa}} \times \Sigma_{\mathrm{flat}} \times \mathbb{R}$ such that $\rho_{a,S}(\tau) > 0$ the function ρ is analytic in a neighbourhood of (a, S, τ). Here, $\mathcal{A}_{\mathrm{sa}}$ and Σ_{flat} are equipped with the metrics induced by $\|\cdot\|$ on $\mathcal{A}$ and its operator norm on $\mathcal{A} \to \mathcal{A}$, respectively.

The following proposition is stated under a boundedness assumption on m (see (8.2.7) below). In the random matrix context, in Section 8.9, we provide a sufficient condition for this assumption to hold purely expressed in terms of a and S for a large class of random matrix models.

Proposition 8.2.4 (Regularity of m)**.** *Let (a, S) be a data pair with $S \in \Sigma_{\mathrm{flat}}$ and m the solution to the associated Dyson equation. Suppose that for a nonempty open interval $I \subset \mathbb{R}$ we have*

$$\limsup_{\eta\downarrow 0}\, \sup_{\tau\in I} \|m(\tau + \mathrm{i}\eta)\| \,<\, \infty\,. \tag{8.2.7}$$

Then m has a $1/3$-Hölder continuous extension (also denoted by m) to any closed interval $I' \subset I$, i.e.,

$$\sup_{z_1,z_2\in I'\times \mathrm{i}[0,\infty)} \frac{\|m(z_1) - m(z_2)\|}{|z_1 - z_2|^{1/3}} \,<\, \infty\,. \tag{8.2.8}$$

Moreover, m is real-analytic in I wherever ρ is positive.

The purpose of the interval I in Proposition 8.2.4 (see also Theorem 8.2.5 below) is to demonstrate the local nature of these statements and their proofs; if m is bounded on

I in the sense of (8.2.7) then we will prove regularity of m and later its behaviour close to singularities on a genuine subinterval $I' \subset I$. At first reading, the reader may ignore this subtlety and assume $I' = I = \mathbb{R}$.

In Proposition 8.4.7 below, we provide a quantitative version of (8.2.8) under slightly weaker conditions than those of Proposition 8.2.4. The bound in this quantitative version only depends on a few basic parameters of the model.

For the following main theorem, we remark that if m has a continuous extension to an interval $I \subset \mathbb{R}$ then the restriction of the measure v from (8.2.4) to I has a density with respect to the Lebesgue measure, i.e., for each Borel set $A \subset I$, we have

$$v(A) = \frac{1}{\pi}\int_A \operatorname{Im} m(\tau)\mathrm{d}\tau. \tag{8.2.9}$$

The existence of a continuous extension can be guaranteed by (8.2.7) in Proposition 8.2.4.

Theorem 8.2.5 (Im m close to its singularities)**.** *Let (a, S) be a data pair with $S \in \Sigma_{\text{flat}}$ and m the solution to the associated Dyson equation. Suppose m has a continuous extension to a nonempty open interval $I \subset \mathbb{R}$. Then any $\tau_0 \in \operatorname{supp}\rho \cap I$ with $\rho(\tau_0) = 0$ belongs to exactly one of the following cases:*

Edge: *The point τ_0 is a left/right edge of the density of states, i.e., there is some $\varepsilon > 0$ such that* $\operatorname{Im} m(\tau_0 \mp \omega) = 0$ *for $\omega \in [0, \varepsilon]$ and for some $v_0 \in \mathcal{A}_+$ we have*

$$\operatorname{Im} m(\tau_0 \pm \omega) \;=\; v_0\omega^{1/2} + \mathcal{O}(\omega)\,, \qquad \omega \downarrow 0\,.$$

Cusp: *The point τ_0 lies in the interior of* $\operatorname{supp}\rho$ *and for some $v_0 \in \mathcal{A}_+$ we have*

$$\operatorname{Im} m(\tau_0 + \omega) \;=\; v_0\,|\omega|^{1/3} + \mathcal{O}(|\omega|^{2/3})\,, \qquad \omega \to 0\,.$$

Moreover, $\operatorname{supp}\rho \cap I = \operatorname{supp} v \cap I$ *is a finite union of closed intervals with nonempty interior.*

Theorem 8.2.5 is a simplified version of our more detailed and quantitative Theorem 8.7.1 below. We can treat all small local minima of ρ on $\operatorname{supp}\rho \cap I$ – not only those ones, where ρ vanishes – and provide precise expansions corresponding to those in Theorem 8.2.5 which are valid in some neighbourhood of τ_0. Moreover, the coefficients v_0 in Theorem 8.2.5 are bounded from above and below in terms of the basic parameters of the model. By applying $\langle \cdot \rangle$ to the results of Theorem 8.2.5 and Theorem 8.7.1, we also obtain an expansion of the self-consistent density of states ρ near small local minima in Theorem 8.7.2 below.

Finally, we present our quantization result.

Proposition 8.2.6 (Band mass formula)**.** *Let $(a, S) \in \mathcal{A}_{\mathrm{sa}} \times \Sigma$ be a data pair and m the solution to the associated Dyson equation,* (8.2.3)*. We assume that there is a constant $C > 0$ such that $S[x] \leq C\langle x\rangle \mathbb{1}$ for all $x \in \overline{\mathcal{A}}_+$. Then we have*

(i) For each $\tau \in \mathbb{R}\backslash \operatorname{supp}\rho$, there is $m(\tau) \in \mathcal{A}_{\mathrm{sa}}$ such that $\lim_{\eta\downarrow 0} \|m(\tau+\mathrm{i}\eta)-m(\tau)\| = 0$. Moreover, $m(\tau)$ determines the mass of $(-\infty, \tau)$ and (τ, ∞) with respect to ρ in the sense that

$$\rho((-\infty,\tau)) = \langle \mathbf{1}_{(-\infty,0)}(m(\tau))\rangle, \tag{8.2.10}$$

where $\mathbf{1}_{(-\infty,0)}$ denotes the characteristic function of the interval $(-\infty, 0)$.

(ii) If $\pi \colon \mathcal{A} \to \mathbb{C}^{n\times n}$ is a faithful representation such that $\langle x\rangle = n^{-1}\operatorname{Tr}(\pi(x))$ for all $x \in \mathcal{A}$ and $J \subset \operatorname{supp}\rho$ is a connected component of $\operatorname{supp}\rho$ then we have

$$n\rho(J) \in \{1, \ldots, n\}.$$

In particular, $\operatorname{supp}\rho$ has at most n connected components.

We will prove Proposition 8.2.6 in Section 8.8 below. A result similar to part (ii) has been obtained by a different method in [86], see also [132]. In fact, we will use the band mass formula, (8.2.10), in Corollary 8.9.4 below to strengthen the quantization result in (ii) for a large class of random matrices (Kronecker matrices, see Section 8.9). In Section 8.10, we study the stability of the Dyson equation, (8.2.3), under small general pertubations of the data pair (a, S).

8.2.1. Examples. We now present some examples that show the different types of singularities described by Theorem 8.2.5. These examples are obtained by considering the Dyson equation, (8.2.3), on $\mathbb{C}^{n\times n}$ with $\langle \cdot \rangle = n^{-1}\operatorname{Tr}$ for

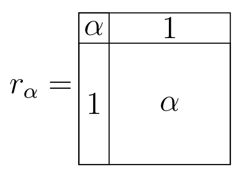
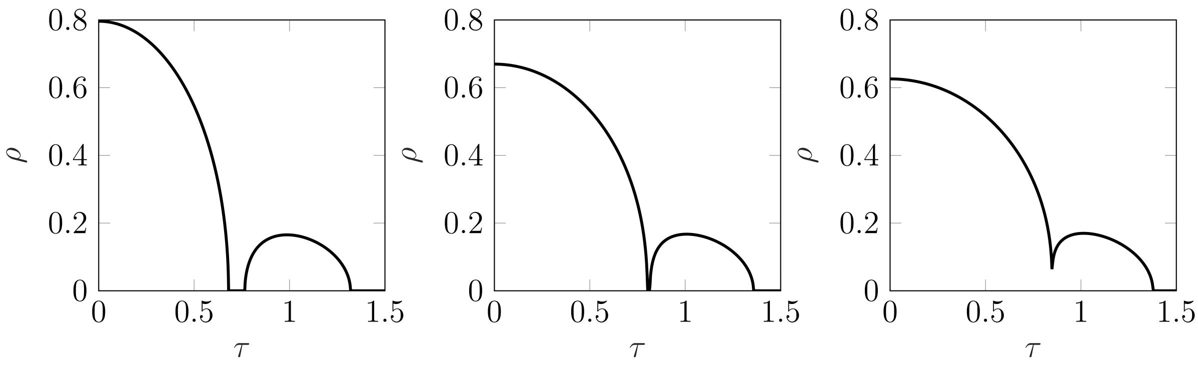


FIGURE 8.1. Structure of $r_\alpha \in \mathbb{C}^{n\times n}$.

large n and choosing $a = 0$ as well as $S = S_\alpha$, where

$$S_\alpha[x] := \frac{1}{n} \operatorname{diag}(r_\alpha \operatorname{diag}(x))$$

for any $x \in \mathbb{C}^{n\times n}$. Here, for $x \in \mathbb{C}^{n\times n}$, $\operatorname{diag}(x)$ denotes the vector of diagonal entries, $r_\alpha \in \mathbb{C}^{n\times n}$ is the symmetric block matrix from Figure 8.1 with $\alpha \in (0, \infty)$. All elements in each block are the indicated constants. Moreover, we write $\operatorname{diag}(v)$ with $v \in \mathbb{C}^n$ to denote



(A) $\alpha = 0.14$ (B) $\alpha = 0.2$ (C) $\alpha = 0.23$

FIGURE 8.2. Examples of the self-consistent density of states ρ from (8.2.11) for $\delta = 0.1$ and several values of α.

the diagonal matrix in $\mathbb{C}^{n\times n}$ with v on its diagonal. In fact, this example can also be realized on $\mathbb{C}^2$ with entrywise multiplication. Here, we choose $\langle (x_1, x_2)\rangle = \delta x_1 + (1-\delta)x_2$, where δ is the relative block size of the small block in the definition of r_α. In this setup on $\mathbb{C}^2$, the Dyson equation can be written as

$$-\begin{pmatrix} m_1^{-1} \\ m_2^{-1} \end{pmatrix} = z \begin{pmatrix} 1 \\ 1 \end{pmatrix} + R_\alpha \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \qquad R_\alpha = \begin{pmatrix} \alpha\delta & 1-\delta \\ \delta & \alpha(1-\delta) \end{pmatrix} \tag{8.2.11}$$

for $(m_1, m_2) \in \mathbb{C}^2$. We remark that R_α is symmetric with respect to the scalar product (8.2.1) induced by $\langle \cdot \rangle$. Figure 8.2 contains the graphs of some self-consistent densities of states ρ obtained from (8.2.11) for $\delta = 0.1$ and different values of α. As the self-consistent density of states is symmetric around zero in these cases, only the part of the density on $[0, \infty)$ is shown. The density in Figure 8.2 (a) has a small internal gap with square root edges on both sides of this gap. Figure 8.2 (b) contains a cusp which is transformed, by increasing α, into an internal nonzero local minimum in Figure 8.2 (c). This nonzero local minimum is covered by Theorem 8.7.1 (d) below.

8.2.2. Main ideas of the proofs. In this subsection, we informally summarize several key ideas in the proofs of Proposition 8.2.4 and Theorem 8.2.5.

Hölder-continuity of m. To simplify the notation, we assume in this outline that $\|m(z)\| \lesssim 1$ for all $z \in \mathbb{H}$, i.e., we assume (8.2.7) with $I = \mathbb{R}$. We first show that $\operatorname{Im} m(z)$ is 1/3-Hölder continuous and then conclude the same regularity for $m = m(z)$. To that end, we now control $\partial_z \operatorname{Im} m(z)$ by differentiating the Dyson equation, (8.2.3), with respect to z. This yields

$$2\mathrm{i}\partial_z \operatorname{Im} m = (\mathrm{Id} - C_m S)^{-1}[m^2].$$

Here, Id denotes the identity map on $\mathcal{A}$ and $C_m\colon \mathcal{A} \to \mathcal{A}$ is defined by $C_m[x] := mxm$ for any $x \in \mathcal{A}$.

In order to control the norm of the stability operator $(\mathrm{Id} - C_m S)^{-1}$, we rewrite it in a more symmetric form. We find an invertible V with $\|V\|, \|V^{-1}\| \lesssim 1$, a unitary operator U and a self-adjoint operator T acting on $\mathcal{A}$ such that

$$\mathrm{Id} - C_m S = V^{-1}(U - T)V.$$

The Rotation-Inversion Lemma from [5] (see Lemma 8.4.4 below) is designed to control $(U - T)^{-1}$ for a unitary operator U and a self-adjoint operator T with $\|T\|_2 \leq 1$. Applying this lemma in our setup yields $\|(\mathrm{Id} - C_m S)^{-1}\| \lesssim \|\operatorname{Im} m\|^{-2}$.

Since $\|m\| \lesssim 1$, we thus obtain

$$\|\partial_z \operatorname{Im} m\| \lesssim \|\operatorname{Im} m\|^{-2}. \tag{8.2.12}$$

This bound implies that $(\operatorname{Im} m)^3\colon \mathbb{H} \to \mathcal{A}_+$ is uniformly Lipschitz-continuous. Hence, we can extend $\operatorname{Im} m$ to a 1/3-Hölder continuous function on $\mathbb{R} \cup \mathbb{H}$ and we obtain

$$m(z) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\operatorname{Im} m(\tau) \mathrm{d}\tau}{\tau - z}.$$

This also implies that m is uniformly 1/3-Hölder continuous on $\mathbb{R} \cup \mathbb{H}$. Furthermore, $m(\tau)$ and $\operatorname{Im} m(\tau)$ are real-analytic in τ around $\tau_0 \in \mathbb{R}$, wherever $\rho(\tau_0)$ is positive.

Behaviour of $\operatorname{Im} m$ **where it is not analytic.** Owing to (8.2.12), some unstable behaviour of the Dyson equation is expected close to points $\tau_0 \in \mathbb{R}$, where $\operatorname{Im} m(\tau_0)$ is zero or small. In order to analyze this behaviour of $\operatorname{Im} m(\tau)$, we compute $\Delta := m(\tau_0 + \omega) - m(\tau_0)$ from the Dyson equation, (8.2.3). Since m has a continuous extension to $\mathbb{R}$, (8.2.3) holds true for $z \in \mathbb{R}$ as well. We evaluate (8.2.3) at $z = \tau_0$ and $z = \tau_0 + \omega$ and obtain the quadratic $\mathcal{A}$-valued equation

$$B[\Delta] = mS[\Delta]\Delta + \omega m\Delta + \omega m^2, \qquad B := \mathrm{Id} - C_m S. \tag{8.2.13}$$

The blow-up of the stability operator B^{-1} close to τ_0 requires analyzing the contributions of Δ in the unstable direction of B^{-1} separately. In fact, B possesses precisely one unstable direction denoted by b since we will show that $\|T\|_2$ is a non-degenerate eigenvalue of T. We decompose Δ into $\Delta = \Theta b + r$, where Θ is the scalar contribution of Δ in the direction b and r lies in the spectral subspace of B complementary to b.

We view τ_0 as fixed and consider $\omega \ll 1$ as the main variable. Projecting (8.2.13) onto b and its complement yield the scalar-valued cubic equation

$$\psi\Theta(\omega)^3 + \sigma\Theta(\omega)^2 + \pi\omega = \mathcal{O}(|\omega||\Theta(\omega)| + |\Theta(\omega)|^4) \tag{8.2.14}$$

with two parameters $\psi \geq 0$ and $\sigma \in \mathbb{R}$. In fact, the 1/3-Hölder continuity of m implies $\Theta = \mathcal{O}(|\omega|^{1/3})$ and, hence, the right-hand side of (8.2.14) is indeed of lower order than the terms on the left-hand side. Analyzing (8.2.14) instead of (8.2.13) is a more tractable problem since we have reduced a quadratic $\mathcal{A}$-valued equation, (8.2.13), to the scalar-valued cubic equation, (8.2.14).

The essential feature of the cubic equation (8.2.14) is its stability. By this, we mean that there exists a constant $c > 0$ such that

$$\psi + \sigma^2 \geq c.$$

This bound will follow from the structure of the Dyson equation and prevents any singularities of higher order than $\omega^{1/2}$ or $\omega^{1/3}$. Obtaining more detailed information about Θ from (8.2.14) requires applying Cardano's formula with an error term. Therefore, we switch to *normal coordinates*, $(\omega, \Theta(\omega)) \to (\lambda, \Omega(\lambda))$, in (8.2.14). We will study four normal forms, one quadratic $\Omega(\lambda)^2 + \Lambda(\lambda) = 0$, and three cubics, $\Omega(\lambda)^3 + \Lambda(\lambda) = 0$ and $\Omega(\lambda)^3 \pm 3\Omega(\lambda) + 2\Lambda(\lambda) = 0$, where $\Lambda(\lambda)$ is a perturbation of the identity map $\lambda \mapsto \lambda$. The first case corresponds to the square root singularity of the isolated edge, the second is the cusp. The last two cases describe the situation of *almost cusps*, see later.

The correct branches in Cardano's formula are identified with the help of four *selection principles* for the solution $\Omega(\lambda)$ corresponding to Θ of the cubic equation in normal form (see **SP1** to **SP4'** at the beginning of Section 8.7.2 below). These selection principles are special properties of Ω which originate from the continuity of m, $\operatorname{Im} m \geq 0$ and the Stieltjes transform representation, (8.2.4), of m. Once the correct branch is chosen, we obtain the precise behaviour of $\operatorname{Im} m$ around τ_0, where $\tau_0 \in \operatorname{supp} \rho$ satisfies $\rho(\tau_0) = 0$ or even $\rho(\tau_0) \ll 1$, from Cardano's formula and careful estimates of r in the decomposition $\Delta = \Theta b + r$ (see Theorem 8.7.1 below).

8.3. The solution of the Dyson equation

In this section, we first introduce some notations used in the proof of Proposition 8.2.1, then prove the proposition and finally give a few further properties of m.

For $x, y \in \mathcal{A}$, we introduce the bounded operator $C_{x,y} \colon \mathcal{A} \to \mathcal{A}$ defined through $C_{x,y}[h] := xhy$ for $h \in \mathcal{A}$. We set $C_x := C_{x,x}$. For $x, y \in \mathcal{A}$, the operator $C_{x,y}$ satisfies the simple relations

$$C_{x,y}^* = C_{x^*,y^*}, \qquad C_{x,y}^{-1} = C_{x^{-1},y^{-1}},$$

where $C_{x,y}^*$ is the adjoint with respect to the scalar product defined in (8.2.1). Here, the second identity holds if x and y are invertible in $\mathcal{A}$. In fact, $C_{x,y}$ is invertible if and only if x and y are invertible in $\mathcal{A}$.

In the following, we will often use the functional calculus for normal elements of $\mathcal{A}$. As we will explain now, our setup allows for a direct way to represent $\mathcal{A}$ as a subalgebra of the bounded operators on a Hilbert space. Therefore, one can think of the functional calculus being performed on this Hilbert space. The Hilbert space is the completion of $\mathcal{A}$ equipped with the scalar product defined in (8.2.1) and denoted by L^2. In order to represent $\mathcal{A}$ as subalgebra of the bounded operators $B(L^2)$ on L^2, we denote by ℓ_x for $x \in \mathcal{A}$ the left-multiplication on L^2 by x, i.e., $\ell_x \colon L^2 \to L^2$, $\ell_x(y) = xy$ for $y \in L^2$. The inclusion $\mathcal{A} \subset L^2$ and the Cauchy-Schwarz inequality yield the well-definedness of ℓ_x and $\ell_x \in B(L^2)$, the bounded linear operators on L^2. In fact,

$$\mathcal{A} \to B(L^2), \qquad x \mapsto \ell_x$$

defines a faithful representation of $\mathcal{A}$ as a von Neumann algebra in $B(L^2)$ [138, Theorem 2.22].

We now introduce the *balanced polar decomposition* of m. If $w = w(z) \in \mathcal{A}$, $q = q(z) \in \mathcal{A}$ and $u = u(z) \in \mathcal{A}$ are defined through

$$w := (\operatorname{Im} m)^{-1/2}(\operatorname{Re} m)(\operatorname{Im} m)^{-1/2} + \mathrm{i}\mathbb{1}, \qquad q := |w|^{1/2}(\operatorname{Im} m)^{1/2}, \qquad u := \frac{w}{|w|} \tag{8.3.1}$$

via the spectral calculus of the self-adjoint operator $(\operatorname{Im} m)^{-1/2}(\operatorname{Re} m)(\operatorname{Im} m)^{-1/2}$ then we have

$$m(z) = \operatorname{Re} m(z) + \mathrm{i}\operatorname{Im} m(z) = q^* u q. \tag{8.3.2}$$

Here, u is unitary and commutes with w. The decomposition $m = q^* u q$ was already introduced and also called balanced polar decomposition in [6] in the special setting of matrix algebras. The operators $|w|^{1/2}$, q and u correspond to $\mathbf{W}$, $\mathbf{W}\sqrt{\operatorname{Im} \mathbf{M}}$ and $\mathbf{U}^*$ in the notation of [6], respectively. With the definitions in (8.3.1), (8.2.3) reads as

$$-u^* = q(z-a)q^* + F[u], \tag{8.3.3}$$

where we introduced the *saturated self-energy operator*

$$F := C_{q,q^*} S C_{q^*,q}. \tag{8.3.4}$$

It is positivity-preserving as well as symmetric, $F = F^*$, and corresponds to the saturated self-energy operator $\mathcal{F}$ in [6].

Proof of Proposition 8.2.1. The existence of v will be a consequence of the following lemma which will be proven in Section 8.11 below.

Lemma 8.3.1. *Let $\mathcal{A}$ be a von Neumann algebra with unit $\mathbb{1}$ and a tracial, faithful, normal state $\langle\,\rangle \colon \mathcal{A} \to \mathbb{C}$. If $h \colon \mathbb{H} \to \mathcal{A}$ is a holomorphic function satisfying* $\operatorname{Im} h(z) \in \mathcal{A}_+$

for all $z \in \mathbb{H}$ *and*

$$\lim_{\eta\to\infty} \mathrm{i}\eta h(\mathrm{i}\eta) = -\mathbb{1} \tag{8.3.5}$$

then there exists a unique measure $v\colon \mathcal{B} \to \mathcal{A}$ *on the Borel sets* $\mathcal{B}$ *of* $\mathbb{R}$ *with values in* $\overline{\mathcal{A}}_+$ *such that*

$$h(z) = \int_{\mathbb{R}} \frac{v(\mathrm{d}\tau)}{\tau - z} \tag{8.3.6}$$

for all $z \in \mathbb{H}$ *and* $v(\mathbb{R}) = \mathbb{1}$*.*

In order to apply Lemma 8.3.1, we have to verify (8.3.5) for $h = m$. To that end, we take the imaginary part of (8.2.3) and use $\operatorname{Im} m \geq 0$ as well as $S \in \Sigma$ to conclude

$$-\operatorname{Im} m^{-1}(z) = \operatorname{Im} z\mathbb{1} + S[\operatorname{Im} m] \geq \operatorname{Im} z\mathbb{1}.$$

Hence, $\|m(z)\| \leq (\operatorname{Im} z)^{-1}$ as for any $x \in \mathcal{A}$ we have $\|x\| \leq 1$ if x is invertible and $\operatorname{Im} x^{-1} \geq \mathbb{1}$. Therefore, evaluating (8.2.3) at $z = \mathrm{i}\eta$, $\eta > 0$, and multiplying the result by m from the left yield

$$\mathrm{i}\eta m(\mathrm{i}\eta) = -\mathbb{1} + m(\mathrm{i}\eta)a - m(\mathrm{i}\eta)S[m(\mathrm{i}\eta)] \to -\mathbb{1}$$

for $\eta \to \infty$ as S is bounded. Hence, Lemma 8.3.1 implies the existence of v, i.e., the Stieltjes transform representation of m in (8.2.4).

This representation has the following well-known bounds as a direct consequence (e.g. [4, 6] or Chapter 7).

Lemma 8.3.2. *Let* v *be the measure in Proposition 8.2.1 and* $\rho = \langle v \rangle$*. Then, for any* $z \in \mathbb{H}$*, we have*

$$\|m(z)\| \leq \frac{1}{\operatorname{dist}(z, \operatorname{supp}\rho)}, \qquad \operatorname{Im} m(z) \leq \frac{\operatorname{Im} z}{\operatorname{dist}(z, \operatorname{supp}\rho)^2}\mathbb{1}. \tag{8.3.7}$$

□

For the proofs of (8.2.5a) and (8.2.5b), we refer to the proofs of Proposition 2.1 in [6] and (7.3.4) in Chapter 7 in the matrix setup, the same argument works for our general setup as well.

We now prove (8.2.6) and hence assume $a = 0$. Taking the imaginary part of the Dyson equation, (8.3.3), yields

$$\operatorname{Im} u \;=\; (\operatorname{Im} z)qq^* + F[\operatorname{Im} u] \;\geq\; \max\{(\operatorname{Im} z)qq^*, F[\operatorname{Im} u]\}\,.$$

Thus, $\operatorname{Im} u \geq (\operatorname{Im} z)\|(qq^*)^{-1}\|^{-1}\mathbb{1}$. We remark that qq^* is invertible since $\operatorname{Im} m(z) > 0$ for $z \in \mathbb{H}$. Therefore, the following Lemma 8.3.3 with $h = \operatorname{Im} u/\|\operatorname{Im} u\|_2$ implies $\|F\|_2 \leq 1$.

Lemma 8.3.3. *Let $T\colon \mathcal{A} \to \mathcal{A}$ be a positivity-preserving operator which is symmetric with respect to* (8.2.1)*. If there are $h \in \mathcal{A}$ and $\varepsilon > 0$ such that $h \geq \varepsilon\mathbb{1}$ and $Th \leq h$ then $\|T\|_2 \leq 1$.*

Proof. The argument in the proof of Lemma 4.6 in [4] also yields this lemma in our current setup. □

We rewrite the Dyson equation (8.3.3) in the form

$$qq^* \;=\; -\frac{1}{z}(u^* + F[u])\,. \tag{8.3.8}$$

We take the $\|\cdot\|_2$-norm on both sides of (8.3.8) and use that $\|u\|_2 = 1$ (since it is unitary) and $\|F\|_2 \leq 1$ to find

$$\|qq^*\|_2 \;\leq\; \frac{2}{|z|}\,. \tag{8.3.9}$$

Then we take the $\|\cdot\|_2$-norm of m and use the balanced polar decomposition $m = q^*uq$ again,

$$\|m\|_2^2 \;=\; \langle m^*m\rangle \;=\; \langle q^*u^*qq^*uq\rangle \;=\; \langle qq^*, C_{u^*,u}[qq^*]\rangle \;\leq\; \|qq^*\|_2^2\,,$$

where the operator $C_{u^*,u}$ is unitary with respect to the scalar product (8.2.1). With (8.3.9) we conclude (8.2.6).

From now on until the end of Section 8.4.2, we will always assume that S is *flat*, i.e., $S \in \Sigma_{\text{flat}}$ (cf. (8.2.2b)). In fact, all of our estimates will be uniform in all data pairs (a, S) that satisfy

$$c_1\langle x\rangle\mathbb{1} \leq S[x] \leq c_2\langle x\rangle\mathbb{1}, \qquad \|a\| \leq c_3 \tag{8.3.10}$$

for all $x \in \mathcal{A}_+$ with the some fixed constants $c_1, c_2, c_3 > 0$. Therefore, the constants c_1, c_2, c_3 from (8.3.10) are called *model parameters* and we introduce the following convention.

Convention 8.3.4 (Comparison relation)**.** *Let $x, y \in \mathcal{A}_{\mathrm{sa}}$. We write $x \lesssim y$ if there is $c > 0$ depending only on the model parameters c_1, c_2, c_3 from* (8.3.10) *such that $cy - x$ is positive definite, i.e., $cy - x \in \overline{\mathcal{A}_+}$. We define $x \gtrsim y$ and $x \sim y$ accordingly. We also use this notation for scalars x, y. Moreover, we write $x = y + \mathcal{O}(\alpha)$ for $x, y \in \mathcal{A}$ and $\alpha > 0$ if $\|x - y\| \lesssim \alpha$.*

We remark that we will choose a different set of model parameters later and redefine $\sim$ accordingly (cf. Convention 8.4.6).

Proposition 8.3.5 (Properties of the solution)**.** *Let (a, S) be a data pair satisfying* (8.3.10) *and m be the solution to the associated Dyson equation,* (8.2.3). *We have*

$$\|m(z)\|_2 \lesssim 1, \tag{8.3.11}$$

$$\|m(z)\| \lesssim \frac{1}{\langle \operatorname{Im} m(z)\rangle + \operatorname{dist}(z, \operatorname{supp} \rho)}, \tag{8.3.12}$$

$$\|m(z)^{-1}\| \lesssim 1 + |z|, \tag{8.3.13}$$

$$\langle \operatorname{Im} m(z)\rangle \mathbb{1} \lesssim \operatorname{Im} m(z) \lesssim (1 + |z|^2)\|m(z)\|^2 \langle \operatorname{Im} m(z)\rangle \mathbb{1} \tag{8.3.14}$$

uniformly for $z \in \mathbb{H}$.

These bounds are immediate consequences of the flatness of S exactly as in the proof of Proposition 4.2 in [6] using $\operatorname{supp} \rho = \operatorname{supp} v$ by the faithfulness of $\langle \cdot \rangle$. We omit the details.

Note that (8.3.13) implies a lower bound $\|m(z)\| \gtrsim (1 + |z|)^{-1}$ since $\|m\|\|m^{-1}\| \geq 1$.

8.4. Regularity of the solution and the density of states

In this section, we will prove Proposition 8.2.3 and Proposition 8.2.4. Their proofs are based on a bound on the stability operator $(\mathrm{Id} - C_m S)^{-1}$ of the Dyson equation, (8.2.3), which will be given in Proposition 8.4.1 below.

8.4.1. Linear stability of the Dyson equation. For the formulation of the following proposition, we introduce the harmonic extension of the density of states ρ defined in Definition 8.2.2 to $\mathbb{H}$. The harmonic extension at $z \in \mathbb{H}$ is denoted by $\rho(z)$ and given by

$$\rho(z) := \frac{1}{\pi}\langle \operatorname{Im} m(z) \rangle.$$

Proposition 8.4.1 (Linear Stability). *There is a universal constant $C > 0$ such that, for the solution m to* (8.2.3) *associated to any $a \in \mathcal{A}_{\mathrm{sa}}$ and $S \in \Sigma$ satisfying* (8.3.10)*, we have*

$$\|(\mathrm{Id} - C_{m(z)}S)^{-1}\|_2 \lesssim 1 + \frac{1}{(\rho(z) + \operatorname{dist}(z, \operatorname{supp}\rho))^C} \tag{8.4.1}$$

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

Before proving Proposition 8.4.1, we will explain how the linear stability yields the Hölder-continuity and analyticity of ρ in Proposition 8.2.3. Indeed, assuming that m depends differentiably on (z, a, S), we can compute the directional derivative $\nabla_{(\delta,d,D)}$ at (z, a, S) of both sides in (8.2.3). The result of this computation is

$$(\mathrm{Id} - C_m S)[\nabla_{(\delta,d,D)} m] \;=\; m(\delta - d + D[m])m.$$

Using the bound in Proposition 8.4.1 and $\rho(z) = \pi^{-1}\langle \operatorname{Im} m(z)\rangle$, we conclude from (8.3.12) that

$$|\nabla_{(\delta,d,D)}\rho| \;\leq\; \frac{1}{\rho^C}(|\delta| + \|d\| + \|D\|) \tag{8.4.2}$$

with a possibly larger C. Therefore, it is clear that the control on $(\mathrm{Id} - C_m S)^{-1}$ will be the key input in the proof of Proposition 8.2.3.

In order to prove Proposition 8.4.1, we will use the representation

$$\mathrm{Id} - C_m S = C_{q^*,q} C_u (C_u^* - F) C_{q^*,q}^{-1}, \tag{8.4.3}$$

where q, u and F were defined in (8.3.1) and (8.3.4), respectively. This representation has the advantage that C_u^* is unitary and F is symmetric. Hence, it is much easier to obtain some spectral properties for $C_u^* - F$ compared to $\mathrm{Id} - C_m S$. Now, we will first analyze q and F in the following two lemmas and then use this knowledge to verify Proposition 8.4.1.

Lemma 8.4.2. *If* (8.3.10) *holds true then we have*

$$\|q(z)\| \lesssim (1+|z|)^{1/2}\|m(z)\|, \qquad \|q(z)^{-1}\| \lesssim (1+|z|)\|m(z)\|^{1/2}$$

uniformly for $z \in \mathbb{H}$.

Proof. For $q = q(z)$, we will show below that

$$\frac{A^{1/2}}{B^{1/2}}\|m(z)^{-1}\|^{-1}\mathbb{1} \le q^*q \le \frac{B^{1/2}}{A^{1/2}}\|m(z)\|\mathbb{1} \tag{8.4.4}$$

if $A\mathbb{1} \le \operatorname{Im} m(z) \le B\mathbb{1}$ for some $A, B \in (0,\infty)$ and $z \in \mathbb{H}$. Choosing A and B according to (8.3.14), using the C^*-property of $\|\cdot\|$, $\|q^*q\| = \|q\|^2$, and (8.3.13), we immediately obtain Lemma 8.4.2.

For the proof of (8.4.4), we set $g := \operatorname{Re} m$ and $h := \operatorname{Im} m$. Using the monotonicity of the square root, we compute

$$\begin{aligned} q^*q &= h^{1/2}\Big(\mathbb{1} + h^{-1/2}gh^{-1}gh^{-1/2}\Big)^{1/2}h^{1/2} \\ &\le A^{-1/2}h^{1/2}\Big(h^{-1/2}(h^2+g^2)h^{-1/2}\Big)^{1/2}h^{1/2} \\ &\le \|m\|A^{-1/2}h^{1/2}. \end{aligned}$$

Here, we employed $h^{-1} \le A^{-1}\mathbb{1}$ as well as $\mathbb{1} \le A^{-1}h$ in the first step and $(\operatorname{Re} m)^2 + (\operatorname{Im} m)^2 = (m^*m + mm^*)/2 \le \|m\|^2$ in the second step. Thus, $h \le B\mathbb{1}$ yields the upper bound in (8.4.4). Similar estimates using $\mathbb{1} \ge B^{-1}h$ and $\|m^{-1}\|^{-2} \le (m^*m + mm^*)/2$ prove the lower bound in (8.4.4) which completes the proof of the lemma. □

Lemma 8.4.3 (Properties of F)**.** *If the bounds in* (8.3.10) *are satisfied then* $\|F\|_2$ *is a simple eigenvalue of* $F\colon \mathcal{A} \to \mathcal{A}$ *defined in* (8.3.4)*. Moreover, there is a unique eigenvector* $f \in \mathcal{A}_+$ *such that* $F[f] = \|F\|_2 f$ *and* $\|f\|_2 = 1$*. This eigenvector satisfies*

$$1 - \|F\|_2 = (\operatorname{Im} z)\frac{\langle f\,, qq^*\rangle}{\langle f\,, \operatorname{Im} u\rangle}. \tag{8.4.5}$$

In particular, $\|F\|_2 \le 1$*. Furthermore, the following properties hold true uniformly for* $z \in \mathbb{H}$ *satisfying* $|z| \le 3(1 + \|a\| + \|S\|^{1/2})$ *and* $\|F(z)\|_2 \ge 1/2$*:*

(i) The eigenvector f *has upper and lower bounds*

$$\|m\|^{-4}\mathbb{1} \lesssim f \lesssim \|m\|^4\mathbb{1}. \tag{8.4.6}$$

(ii) The operator F has a spectral gap $\vartheta \in (0,1]$ satisfying $\vartheta \gtrsim \|m\|^{-28}$ and

$$\mathrm{Spec}(F/\|F\|_2) \subset [-1+\vartheta, 1-\vartheta] \cup \{1\}. \tag{8.4.7}$$

Proof. The definition of F in (8.3.4), (8.3.10) and Lemma 8.4.2 imply

$$(1+|z|)^{-4}\|m(z)\|^{-2}\langle a\rangle \mathbb{1} \lesssim F[a] \lesssim (1+|z|)^2\|m(z)\|^4\langle a\rangle \mathbb{1} \tag{8.4.8}$$

for all $a \in \mathcal{A}_+$ and all $z \in \mathbb{H}$. We will use Lemma 8.12.1 (ii) from Section 8.12 below. The condition (8.12.1) with $T = F$ is satisfied by (8.4.8) with constants depending on $\|m\|$ and $|z|$. Hence, Lemma 8.12.1 (ii) implies the existence and uniqueness of the eigenvector f. We compute the scalar product of f with the imaginary part of (8.3.3). Since F is symmetric, this immediately yields (8.4.5).

We now assume that $z \in \mathbb{H}$ satisfies $|z| \leq 3(1+\|a\|+\|S\|^{1/2})$ and $\|F(z)\|_2 \geq 1/2$. Then $|z| \lesssim 1$ and, by using this in (8.4.8), we thus obtain (8.4.6) and (8.4.7) from Lemma 8.12.1 (ii) since $\|m\| \gtrsim 1$ by (8.3.13). □

The following proof of Proposition 8.4.1 proceeds similarly to the one of Proposition 4.4 in [6].

Proof of Proposition 8.4.1. We will distinguish several cases. If $|z| \geq 3(1+\kappa)$ with $\kappa := \|a\| + 2\|S\|^{1/2}$ then we conclude from (8.2.4) and $\operatorname{supp}\rho \subset [-\kappa,\kappa]$ by (8.2.5a) that $\|m(z)\| \leq (|z|-\kappa)^{-1}$. Thus,

$$\|C_{m(z)}S\|_2 \leq \frac{\|S\|_2}{(|z|-\kappa)^2} \leq \frac{\|S\|}{4(1+\kappa)^2} \leq \frac{1}{4}.$$

Here, we used $\|S\|_2 \leq \|S\|$ since S is symmetric and $\kappa \geq \|S\|^{1/2}$. This shows (8.4.1) for large $|z|$.

Next, we assume $|z| \leq 3(1+\kappa)$. In this regime, we use the alternative representation of $\mathrm{Id} - C_mS$ in (8.4.3) and the spectral properties of F from Lemma 8.4.3. Indeed, from (8.4.3) and Lemma 8.4.2, we conclude

$$\|(\mathrm{Id}-C_mS)^{-1}\|_2 \lesssim \|m\|^3\|(C_u^*-F)^{-1}\|_2 \lesssim \frac{1}{(\rho(z)+\mathrm{dist}(z,\operatorname{supp}\rho))^3}\|(C_u^*-F)^{-1}\|_2 \tag{8.4.9}$$

as $u \in \mathcal{A}$ is unitary. Here, we used (8.3.12) in the last step. If $\|F(z)\|_2 \leq 1/2$ then this immediately yields (8.4.1) as $\|C_u\|_2 = 1$. We now assume $\|F(z)\|_2 \geq 1/2$. In this case, we will use the following lemma.

Lemma 8.4.4 (Rotation-Inversion Lemma)**.** *Let U be a unitary operator on L^2 and T a symmetric operator on L^2. We assume that there is a constant $\theta > 0$ such that*

$$\operatorname{Spec} T \subset [-\|T\|_2 + \theta, \|T\|_2 - \theta] \cup \{\|T\|_2\}$$

with a non-degenerate eigenvalue $\|T\|_2 \leq 1$. Then there is a universal constant $C > 0$ such that

$$\|(U-T)^{-1}\|_2 \leq \frac{C}{\theta |1 - \|T\|_2 \langle t\,, U[t]\rangle|},$$

where $t \in L^2$ is the normalized, $\|t\|_2 = 1$, eigenvector of T corresponding to $\|T\|_2$.

The proof of this lemma is identical to the proof of Lemma 5.6 in [5], where a result of this type was first applied in the context of vector Dyson equations.

We start from the estimate (8.4.9), use the Rotation-Inversion Lemma, Lemma 8.4.4, with $U = C_u^*$ and $T = F$ as well as (8.4.7) and (8.3.12) and obtain

$$\|(\mathrm{Id} - C_m S)^{-1}\|_2 \lesssim \frac{(\rho(z) + \operatorname{dist}(z, \operatorname{supp}\rho))^{-31}}{|1 - \|F\|_2 \langle f\,, C_u^*[f]\rangle|} \leq \frac{(\rho(z) + \operatorname{dist}(z, \operatorname{supp}\rho))^{-31}}{\max\{1 - \|F\|_2, |1 - \langle f C_u^*[f]\rangle|\}}.$$

In order to complete the proof of (8.4.1), we now show that

$$\max\{1 - \|F\|_2, |1 - \langle f C_u^*[f]\rangle|\} \gtrsim (\rho(z) + \operatorname{dist}(z, \operatorname{supp}\rho))^C \tag{8.4.10}$$

for some universal constant $C > 0$. We first prove auxiliary upper and lower bounds on $\operatorname{Im} u = (q^*)^{-1}(\operatorname{Im} m) q^{-1}$. We have

$$\rho(z)(\rho(z) + \operatorname{dist}(z, \operatorname{supp}\rho))^2 \mathbb{1} \lesssim \operatorname{Im} u \lesssim \frac{\operatorname{Im} z \|m\|}{\operatorname{dist}(z, \operatorname{supp}\rho)^2} \mathbb{1}. \tag{8.4.11}$$

For the lower bound, we used the lower bound in (8.3.14), Lemma 8.4.2 and (8.3.12). The upper bound is a direct consequence of (8.3.7) as well as Lemma 8.4.2. Since $\langle f\,, qq^*\rangle \geq \|(qq^*)^{-1}\|^{-1}\langle f\rangle \gtrsim \|m\|\langle f\rangle$ by Lemma 8.4.2, the relation (8.4.5) and the upper bound in (8.4.11) yield

$$1 - \|F\|_2 \gtrsim \operatorname{dist}(z, \operatorname{supp}\rho)^2.$$

As $1 - \langle f C_{\mathrm{Re}\, u}[f] \rangle \geq 0$ and $\langle f^2 \rangle = 1$, we obtain from the lower bound in (8.4.11) that

$$|1 - \langle f C_u^*[f] \rangle| \geq \mathrm{Re}\,[1 - \langle f C_u^*[f] \rangle] = 1 - \langle f C_{\mathrm{Re}\, u}[f] \rangle + \langle f C_{\mathrm{Im}\, u}[f] \rangle \\ \gtrsim \rho(z)^2 (\rho(z) + \mathrm{dist}(z, \mathrm{supp}\, \rho))^4. \tag{8.4.12}$$

This completes the proof of (8.4.10) and hence of Proposition 8.4.1. □

8.4.2. Proof of Proposition 8.2.3. The following proof of Proposition 8.2.3 is similar to the one of Proposition 2.2 in [6].

Proof of Proposition 8.2.3. We first show that $\rho\colon \mathbb{H} \to (0, \infty)$ has a uniformly Hölder-continuous extension to $\overline{\mathbb{H}}$, which we will also denote by ρ. This extension restricted to $\mathbb{R}$ will be the density of the measure ρ from Definition 8.2.2. Since $\mathrm{Id} - C_m S$ is invertible for each $z \in \mathbb{H}$ by (8.4.1), the implicit function theorem allows us to differentiate (8.2.3) with respect to z. This yields

$$(\mathrm{Id} - C_m S)[\partial_z m] = m^2. \tag{8.4.13}$$

Since $z \mapsto \langle m(z) \rangle$ is holomorphic on $\mathbb{H}$ as remarked below (8.2.3), we have $2\pi\mathrm{i}\partial_z \rho(z) = 2\mathrm{i}\partial_z \mathrm{Im}\, \langle m(z) \rangle = \partial_z \langle m(z) \rangle$. Thus, we obtain from (8.4.13) that

$$|\partial_z \rho| \lesssim \|\partial_z m\|_2 \leq \|(\mathrm{Id} - C_m S)^{-1}\|_2 \|m\|^2 \lesssim \rho^{-(C+2)} \tag{8.4.14}$$

Here, we used (8.4.1), $\rho(z) \lesssim \|m(z)\|_2 \lesssim 1$ by (8.3.11) and (8.3.12) in the last step. Hence, ρ^{C+3} is a uniformly Lipschitz-continuous function on $\mathbb{H}$. Therefore, ρ defines uniquely a uniformly $1/(C+3)$-Hölder continuous function on $\mathbb{R}$ which is a density of the measure ρ from Definition 8.2.2 with respect to the Lebesgue measure on $\mathbb{R}$.

Next, we show the Hölder-continuity with respect to a and S. As before in (8.4.2), we compute the derivatives and use (8.3.12) and (8.4.1) to obtain

$$|\nabla_{(d,D)} \rho_{(a,S)}(z)| \lesssim |\langle \nabla_{(d,D)} m \rangle| \lesssim \frac{\|d\| + \|D\|}{\rho^{C+3}}.$$

Since the constants in (8.4.1) and (8.3.12) depend on the constants in (8.3.10), we conclude that ρ is also a locally $1/(C+4)$-Hölder continuous function of a and S.

We are left with showing that ρ is real-analytic in a neighbourhood of (τ_0, a, S) if $\rho_{a,S}(\tau_0) > 0$. Since $\rho(\tau_0) > 0$, we can extend m to τ_0 by (8.4.14). Moreover, $m(\tau_0)$ is invertible as $\operatorname{Im} m(\tau_0) > 0$ and, thus, solves (8.2.3) with $z = \tau_0$. Since (8.2.3) depends analytically on $z = \tau$, a and S in a small neighbourhood of (τ_0, a, S), the solution m and thus ρ will depend analytically on (τ, a, S) in this neighbourhood by the implicit function theorem. This completes the proof of Proposition 8.2.3. □

8.4.3. Proof of Proposition 8.2.4. For $I \subset \mathbb{R}$ and $\eta_* > 0$, we define

$$\mathbb{H}_{I,\eta_*} := \{z \in \mathbb{H} \colon \operatorname{Re} z \in I,\ \operatorname{Im} z \in (0, \eta_*]\} \tag{8.4.15}$$

and its closure $\overline{\mathbb{H}}_{I,\eta_*}$.

Assumptions 8.4.5. Let m be the solution of (8.2.3) for $a = a^* \in \mathcal{A}$ satisfying $\|a\| \leq k_1$ with a positive constant k_1 and $S \in \Sigma$ satisfying $\|S\|_{2\to\|\cdot\|} \leq k_2$ for some positive constant k_2. For an interval $I \subset \mathbb{R}$ and some $\eta_* \in (0, 1]$, we assume that

(i) There are positive constants k_3, k_4 and k_5 such that

$$\|m(z)\| \leq k_3, \tag{8.4.16}$$

$$k_4 \langle \operatorname{Im} m(z) \rangle \mathbb{1} \leq \operatorname{Im} m(z) \leq k_5 \langle \operatorname{Im} m(z) \rangle \mathbb{1}, \tag{8.4.17}$$

uniformly for all $z \in \mathbb{H}_{I,\eta_*}$.

(ii) The operator $F := C_{q,q^*} S C_{q^*,q}$ has a simple eigenvalue $\|F\|_2$ with eigenvector $f \in \mathcal{A}_+$ that satisfies (8.4.5) for all $z \in \mathbb{H}_{I,\eta_*}$. Moreover, (8.4.7) holds true and there are positive constants k_6, k_7 and k_8 such that

$$k_6 \mathbb{1} \leq f \leq k_7 \mathbb{1}, \qquad \vartheta \geq k_8. \tag{8.4.18}$$

uniformly for all $z \in \mathbb{H}_{I,\eta_*}$.

We remark that $S \in \Sigma_{\text{flat}}$ is not necessarily required in Assumptions 8.4.5. In fact, we will show in Lemma 8.4.8 below that $S \in \Sigma_{\text{flat}}$ and (8.4.16) imply all other conditions in Assumptions 8.4.5.

Convention 8.4.6 (Model parameters, Comparison relation)**.** *For the remainder of the Section 8.4 as well as Section 8.5 and Section 8.6, we will only consider $k_1, \dots, k_8$ as* model parameters *and understand the comparison relation $\sim$ from Convention 8.3.4 with respect to this set of model parameters.*

We remark that all of our estimates will be uniform in $\eta_* \in (0,1]$. Therefore, η_* is not considered a model parameter. At the end of this section, we will directly conclude Proposition 8.2.4 from the following proposition.

Proposition 8.4.7 (Regularity of m). *Let Assumptions 8.4.5 hold true on an interval $I \subset \mathbb{R}$ for some $\eta_* \in (0,1]$.*

Then, for any $\theta \in (0,1]$, m can be uniquely extended to $I_\theta := \{\tau \in I \colon \operatorname{dist}(\tau, \partial I) \geq \theta\}$ such that it is uniformly $1/3$-Hölder continuous, indeed,

$$\|m(z_1) - m(z_2)\| \lesssim \theta^{-4/3}|z_1 - z_2|^{1/3} \tag{8.4.19}$$

for all $z_1, z_2 \in I_\theta \times \mathrm{i}[0,\infty)$. Moreover, if $\rho(\tau_0) > 0$, $\tau_0 \in I$, then m is real-analytic in a neighbourhood of τ_0 and

$$\|\partial_\tau m(\tau_0)\| \lesssim \rho(\tau_0)^{-2}. \tag{8.4.20}$$

We remark that the bound in (8.4.20) will be extended to higher derivatives in Lemma 8.5.7 below.

In the following lemma, we establish a very helpful consequence of (i) in Assumptions 8.4.5. Moreover, part (ii) of the following lemma shows that all conditions in Assumptions 8.4.5 are satisfied if we assume (8.4.16) and the flatness of S.

Lemma 8.4.8. *Let m be the solution to* (8.2.3) *for some data pair $(a, S) \in \mathcal{A}_{\mathrm{sa}} \times \Sigma$. We have*

(i) Let $\|a\| \lesssim 1$, $\|S\| \lesssim 1$ and $U \subset \mathbb{H}$ such that $\sup\{|z| \colon z \in U\} \lesssim 1$. If (8.4.16) *and* (8.4.17) *hold true uniformly for $z \in U$ then, uniformly for $z \in U$, we have*

$$\|q\|, \|q^{-1}\| \sim 1, \qquad \operatorname{Im} u \sim \langle \operatorname{Im} u \rangle \mathbb{1} \sim \rho \mathbb{1}. \tag{8.4.21}$$

(ii) Let $I \subset [-C, C]$ for some $C \sim 1$ and (8.4.16) *hold true uniformly for all $z \in \mathbb{H}_{I,\eta_*}$. If $S \in \Sigma_{\mathrm{flat}}$ and $\|a\| \lesssim 1$ then $\|S\|_{2\to\|\cdot\|} \lesssim 1$,* (8.4.17) *holds true uniformly for all $z \in \mathbb{H}_{I,\eta_*}$ and part (ii) of Assumptions 8.4.5 is satisfied.*

(iii) If Assumptions 8.4.5 hold true then, uniformly for $z \in \mathbb{H}_{I,\eta_*}$, we have

$$\|(\mathrm{Id} - C_{m(z)}S)^{-1}\|_2 + \|(\mathrm{Id} - C_{m(z)}S)^{-1}\| \lesssim \rho(z)^{-2}. \tag{8.4.22}$$

Proof of Lemma 8.4.8. For the proof of (i), we use $\|a\| \lesssim 1$, $\|S\| \lesssim 1$ and (8.2.3) to show $\|m(z)^{-1}\| \lesssim 1$ uniformly for all $z \in U$. Thus, following the proof of Lemma 8.4.2 immediately yields the estimates on q and q^{-1} in (8.4.21) due to (8.4.16) and (8.4.17). Thus, as $\|q\|, \|q^{-1}\| \sim 1$, we obtain the missing relations in (8.4.21) from (8.4.17) since

$$\operatorname{Im} u = (q^*)^{-1}(\operatorname{Im} m)q^{-1} \sim \operatorname{Im} m \sim \langle \operatorname{Im} m\rangle \sim \langle \operatorname{Im} u\rangle.$$

We now show (ii). By Lemma 8.12.2 (i), the upper bound in the definition of flatness, (8.3.10), implies $\|S\|_{2\to\|\cdot\|} \lesssim 1$. Owing to (8.4.16) and (8.3.13), we have $\|m(z)\| \sim 1$ for all $z \in \mathbb{H}_{I,\eta_*}$. Hence, (8.4.17) follows from (8.3.14) since $|z| \le C+1$ for $z \in \mathbb{H}_{I,\eta_*}$. Moreover, (ii) in Assumptions 8.4.5 is a consequence of Lemma 8.4.3.

To prove (8.4.22), we follow the proof of Proposition 8.4.1 and replace the use of (8.3.12) as well as (8.4.6) and (8.4.7) from Lemma 8.4.3 by (8.4.16) and (8.4.18), respectively. This yields

$$\|(\mathrm{Id} - C_m S)^{-1}\|_2 \lesssim 1 + |1 - \|F\|_2\langle f C_u^*[f]\rangle|^{-1} \lesssim |1 - \|F\|_2\langle f C_u^*[f]\rangle|^{-1}, \tag{8.4.23}$$

where we used in the last step that (8.4.16) implies $\rho(z) \lesssim 1$ on $\mathbb{H}_{I,\eta_*}$. Since $\operatorname{Im} u \sim \rho$ by (8.4.21) and $\|F\|_2 \le 1$ by (8.4.5) that holds under Assumptions 8.4.5 (ii), we conclude

$$|1 - \|F\|_2\langle f C_u^*[f]\rangle|^{-1} \lesssim |1 - \langle f C_u^*[f]\rangle|^{-1} \lesssim \rho^{-2}$$

as in (8.4.12) in the proof of Proposition 8.4.1. This shows $\|(\mathrm{Id} - C_m S)^{-1}\|_2 \lesssim \rho(z)^{-2}$. Using $\|S\|_{2\to\|\cdot\|} \lesssim 1$ and Lemma 8.12.2 (ii), we obtain the missing $\|\cdot\|$-bound in (8.4.22). This completes the proof of Lemma 8.4.8. □

Proof of Proposition 8.4.7. Similarly to the proof of Proposition 8.2.3, we obtain

$$\|\partial_z \operatorname{Im} m(z)\| \lesssim \|\partial_z m(z)\| \le \|(\mathrm{Id} - C_m S)^{-1}\|\|m(z)\|^2 \lesssim \rho(z)^{-2} \sim \|\operatorname{Im} m(z)\|^{-2} \tag{8.4.24}$$

for $z \in \mathbb{H}_{I,\eta_*}$ from (8.4.16), (8.4.22) and (8.4.17). By the submultiplicativity of $\|\cdot\|$, $(\operatorname{Im} m(z))^3 \colon \mathbb{H}_{I,\eta_*} \to (\mathcal{A}, \|\cdot\|)$ is a uniformly Lipschitz-continuous function. Hence, $\operatorname{Im} m(z)$ is uniformly 1/3-Hölder continuous on $\mathbb{H}_{I,\eta_*}$ (see e.g. Theorem X.1.1 in [35]) and, thus, has a uniformly 1/3-Hölder continuous extension to $\overline{\mathbb{H}}_{I,\eta_*}$. We conclude that the measure v restricted to I has a density with respect to the Lebesgue measure on I, i.e., (8.2.9) holds true for all measurable $A \subset I$. Now, (8.11.3) in Lemma 8.11.1 implies the uniform 1/3-Hölder continuity of m on $I_\theta \times \mathrm{i}(0,\infty)$. In particular, m can be uniquely extended to a uniformly 1/3-Hölder continuous function on $I_\theta \times \mathrm{i}[0,\infty)$ such that (8.4.19) holds true.

To prove the analyticity of m, we refer to the proof of the analyticity of ρ in Proposition 8.2.3. The bound (8.4.20) can be read off from (8.4.24). This completes the proof of the proposition. □

Proof of Proposition 8.2.4. By (8.2.7), there are $C_0 > 0$ and $\eta_* \in (0,1]$ such that $\|m(\tau + \mathrm{i}\eta)\| \leq C_0$ for all $\tau \in I$ and $\eta \in (0,\eta_*]$. Hence, by Lemma 8.4.8 (ii), the flatness of S implies Assumptions 8.4.5 on $I \cap [-C,C]$ for $C := 3(1 + \|a\| + \|S\|^{1/2})$, i.e., $C \sim 1$. Therefore, Proposition 8.4.7 yields Proposition 8.2.4 on $I \cap [-C,C]$.

Owing to (8.3.7) and $\operatorname{supp} v = \operatorname{supp} \rho$, we have $\operatorname{dist}(\tau, \operatorname{supp} v) \geq 1$ for $\tau \in I$ satisfying $\tau \notin [-C+1, C-1]$. Hence, for these τ, the Hölder-continuity follows immediately from (8.11.4) in Lemma 8.11.1. By (8.2.5a), we have $\operatorname{Im} m(\tau) = 0$ for $\tau \in I$ satisfying $\tau \notin [-C,C]$. Therefore, the statement about the analyticity is trivial outside of $[-C,C]$. This completes the proof of Proposition 8.2.4. □

8.5. Spectral properties of the stability operator for small self-consistent density of states

In this section, we study the stability operator B^{-1}, where $B = B(z) := \mathrm{Id} - C_{m(z)}S$, when $\rho = \rho(z)$ is small and Assumptions 8.4.5 hold true. Note that we do not require S to be flat, i.e., to satisfy (8.3.10). We will view B as a perturbation of the operator B_0, which we introduce now. We define

$$\begin{aligned} s &:= \operatorname{sign} \operatorname{Re} u, \qquad\qquad B_0 := C_{q^*,q}(\mathrm{Id} - C_s F)C_{q^*,q}^{-1}, \\ E &:= (C_{q^*sq} - C_m)S = C_{q^*,q}(C_s - C_u)FC_{q^*,q}^{-1}, \end{aligned} \tag{8.5.1}$$

with u and q defined in (8.3.1). Note $B_0 = \mathrm{Id} - C_{q^*sq}S$, i.e., in the definition of B, u in $m = q^*uq$ is replaced by s. Thus, we have $B = B_0 + E$. Under Assumptions 8.4.5, (8.4.21) holds true which we will often use in the following. Since $\mathbb{1} - |\operatorname{Re} u| = \mathbb{1} - \sqrt{\mathbb{1} - (\operatorname{Im} u)^2} \leq$

$(\mathrm{Im}\, u)^2 \lesssim \rho^2$, we also obtain

$$\mathrm{Re}\, u = s + \mathcal{O}(\rho^2), \qquad \mathrm{Im}\, u \;=\; \mathcal{O}(\rho)\,, \qquad \mathrm{Re}\, m = q^* s q + \mathcal{O}(\rho^2) \tag{8.5.2}$$

and with $C_s - C_u = \mathcal{O}(\|s-u\|) = \mathcal{O}(\rho)$ we get

$$E \;=\; \mathcal{O}(\rho)\,. \tag{8.5.3}$$

Here, we use the notation $R = T + \mathcal{O}(\alpha)$ for operators T and R on $\mathcal{A}$ and $\alpha > 0$ if $\|R-T\| \lesssim \alpha$. We introduce

$$f_u := \rho^{-1}\mathrm{Im}\, u. \tag{8.5.4}$$

By the functional calculus for the normal operator u, $\mathrm{Re}\, u$, s and f_u commute. Hence, $C_s[f_u] = f_u$. From the imaginary part of (8.3.3) and (8.4.21), we conclude that

$$(\mathrm{Id} - F)[f_u] = \rho^{-1}\mathrm{Im}\, zqq^* = \mathcal{O}(\rho^{-1}\mathrm{Im}\, z). \tag{8.5.5}$$

In the following, for $z \in \mathbb{C}$ and $\varepsilon > 0$, we denote by $D_\varepsilon(z) := \{w \in \mathbb{C} \colon |z-w| < \varepsilon\}$ the disk in $\mathbb{C}$ of radius ε around z.

Lemma 8.5.1 (Spectral properties of stability operator)**.** *Let* $T \in \{\mathrm{Id} - F,\ \mathrm{Id} - C_s F,\ B_0,\ B,\ \mathrm{Id} - C_{m^*,m}S\}$*. If Assumptions 8.4.5 are satisfied on an interval* $I \subset \mathbb{R}$ *for some* $\eta_* \in (0,1]$*, then there are* $\rho_* \sim 1$ *and* $\varepsilon \sim 1$ *such that*

$$\|(T - \omega\,\mathrm{Id})^{-1}\|_2 + \|(T - \omega\,\mathrm{Id})^{-1}\| + \|(T^* - \omega\,\mathrm{Id})^{-1}\| \lesssim 1 \tag{8.5.6}$$

uniformly for all $z \in \mathbb{H}_{I,\eta_*}$ *satisfying* $\rho(z) + \rho(z)^{-1}\mathrm{Im}\, z \le \rho_*$ *and for all* $\omega \in \mathbb{C}$ *with* $\omega \notin D_\varepsilon(0) \cup D_{1-2\varepsilon}(1)$*. Furthermore, there is a single simple (algebraic multiplicity* 1*) eigenvalue* λ *in the disk around* 0*, i.e.,*

$$\begin{aligned} \mathrm{Spec}(T) \cap D_\varepsilon(0) \;=\; \{\lambda\} \qquad &\textit{and} \qquad \mathrm{rank}\, P_T \;=\; 1\,, \\ \textit{where} \quad P_T \;:=\; -&\frac{1}{2\pi \mathrm{i}} \int_{\partial D_\varepsilon(0)} (T - \omega \mathrm{Id})^{-1} \mathrm{d}\omega\,. \end{aligned} \tag{8.5.7}$$

If Assumptions 8.4.5 are satisfied on I for some $\eta_* \in (0,1]$ then we have

$$f_u = \rho^{-1}\mathrm{Im}\, u \sim 1. \tag{8.5.8}$$

uniformly for $z \in \mathbb{H}_{I,\eta_*}$ due to (8.4.21). This fact will often be used in the following without mentioning it.

PROOF. First, we introduce the bounded operators $V_t \colon \mathcal{A} \to \mathcal{A}$ for $t \in [0,1]$ interpolating between Id and C_s by

$$V_t := (1-t)\mathrm{Id} + t\,C_s\,.$$

We will perform the proof one by one for the choices $T = \mathrm{Id} - F, \mathrm{Id} - V_t F, B_0, B, \mathrm{Id} - C_{m^*,m}S$ in that order. The operator $\mathrm{Id} - F$ has a spectral gap above the single eigenvalue around 0, so for this choice the statements are easy. Then we perform two approximations. First, we interpolate between $\mathrm{Id} - F$ and $\mathrm{Id} - C_s F$ via $\mathrm{Id} - V_t F$. This gives Lemma 8.5.1 for $T = B_0$. Then we use perturbation theory to get the results for $T = B = B_0 + \mathcal{O}(\rho)$ and for $T = \mathrm{Id} - C_{m^*,m}S = B_0 + \mathcal{O}(\rho)$. Note that for all these choices of T the bound $\|\mathrm{Id} - T\|_{2\to\|\cdot\|} \lesssim 1$ holds due to $\|S\|_{2\to\|\cdot\|} \lesssim 1$, (8.4.16) and (8.4.21). Hence, the invertibility of $T - \omega\,\mathrm{Id}$ as an operator on $\mathcal{A}$ and on L^2 are therefore closely related by Lemma 8.12.2 (ii). In particular, it suffices to show (8.5.7) and the $\|\cdot\|_2$-norm bound

$$\|(T - \omega\,\mathrm{Id})^{-1}\|_2 \lesssim 1\,, \tag{8.5.9}$$

for $\omega \notin D_\varepsilon(0) \cup D_{1-2\varepsilon}(1)$ in (8.5.6) to establish the lemma. For $T = \mathrm{Id} - F$ both of these assertion are true due to Lemma 8.4.3. In particular, we find

$$f = \|f_u\|_2^{-1} f_u + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)\,, \tag{8.5.10}$$

where f is the single top eigenvector of F, $Ff = \|F\|_2 f$ (see Lemma 8.4.3). The proof of (8.5.10) follows from (8.5.5) and $\|F\|_2 = 1 + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)$ (cf. (8.4.5)) by straightforward perturbation theory of the simple isolated eigenvalue $\|F\|_2$.

Now we consider the choice $T = T_t = \mathrm{Id} - V_t F$. Once (8.5.9), and with it (8.5.6), is established for T_t, the statement about the single isolated eigenvalue (8.5.7) follows. Indeed, assuming (8.5.6) for $T = T_t$, we obtain that T_t and, hence, the rank of P_{T_t} is a continuous function of t on $[0,1]$. Hence, the rank of P_{T_t} is constant along this interpolation. On the other hand, $\mathrm{rank}\, P_{T_0} = 1$ by Lemma 8.4.3. Therefore, for each $t \in [0,1]$, $\mathrm{Spec}(T_t) \cap D_\varepsilon(0)$ consists of precisely one simple eigenvalue. We are thus left with establishing (8.5.9) for T_t. As $\|V_t\|_2 \leq 1$ and $\|F\|_2 \leq 1$ the bound (8.5.9) is certainly satisfied for $|\omega| \geq 3$. Thus, we now assume $|\omega| \leq 3$. In order to conclude (8.5.9), we now show a lower bound on $\|((1-\omega)\mathrm{Id} - V_t F)[x]\|_2$ for all normalized, $\|x\|_2 = 1$, elements $x \in \mathcal{A}$. We decompose $x \in L^2$ as $x = \alpha f + y$, where $y \perp f$ and $\alpha \in \mathbb{C}$. Then

$$\|((1-\omega)\mathrm{Id} - V_t F)[x]\|_2^2 = |\alpha|^2|\omega|^2 + \|((1-\omega)\mathrm{Id} - V_t F)[y]\|_2^2 + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)\,, \quad (8.5.11)$$

because of $\|F\|_2 = 1 + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)$, $V_t[f_u] = f_u$ together with (8.5.10), and because the mixed terms are negligible due to

$$\langle f\,, V_t F[y]\rangle \,=\, \langle F V_t f\,, y\rangle \,=\, \mathcal{O}(\|y\|_2\, \rho^{-1}\mathrm{Im}\, z)\,.$$

Using the spectral gap $\vartheta \sim 1$ of F from (8.4.7) and $y \perp f$ we infer (8.5.9) from (8.5.11) by estimating

$$\|((1-\omega)\mathrm{Id} - V_t F)[y]\|_2^2 \,\geq\, \mathrm{dist}(\omega, D_{1-\vartheta}(1))\|y\|_2^2 \,\geq\, (\vartheta - \varepsilon)(1 - |\alpha|^2),$$

optimizing in α and choosing $\delta \leq \vartheta/2$. This shows the lemma for $T = \mathrm{Id} - V_t F$.

Since B_0 is related by the similarity transform (8.5.1) to $\mathrm{Id} - V_1 F = \mathrm{Id} - C_s F$ and $\|q\|\|q^{-1}\| \lesssim 1$ (cf. (8.4.21)), the operator B_0 inherits the properties listed in the lemma from $\mathrm{Id} - C_s F$. Finally, we can perform analytic perturbation theory for the simple isolated eigenvalue in $D_\varepsilon(0)$ of B_0 to verify the lemma for $T = B = B_0 + E$ with $E = \mathcal{O}(\rho)$ (cf. (8.5.3)) and $T = \mathrm{Id} - C_{m^*,m} S = B_0 + E_*$ with $E_* = \mathcal{O}(\rho)$ if ρ_* is sufficiently small. Here, we introduced

$$E_* := (C_{q^* s q} - C_{m^*,m})S = C_{q^*,q}(C_s - C_{u^*,u})F C_{q^*,q}^{-1}$$

and used $C_s - C_{u^*,u} = \mathcal{O}(\|s - u\|) = \mathcal{O}(\rho)$ due to (8.5.2). □

If $z \in \mathbb{H}_{I,\eta_*}$ satisfies $\rho(z) + \rho(z)^{-1}\mathrm{Im}\, z \leq \rho_*$ for $\rho_* \sim 1$ from Lemma 8.5.1 then we denote by $P_{s,F}$ the spectral projection corresponding to the isolated eigenvalue of $\mathrm{Id} - C_s F$, i.e., $P_{s,F}$ equals P_T in (8.5.7) with $T = \mathrm{Id} - C_s F$. We also set $Q_{s,F} := \mathrm{Id} - P_{s,F}$. Moreover, for such z, we define ψ and σ by

$$\psi(z) := \langle s f_u^2\,, (\mathrm{Id} + F)(\mathrm{Id} - C_s F)^{-1} Q_{s,F}[s f_u^2]\rangle, \qquad \sigma(z) := \langle s f_u^3\rangle. \quad (8.5.12)$$

Corollary 8.5.2. *Let $z \in \mathbb{H}_{I,\eta_*}$ satisfy $\rho(z) + \rho(z)^{-1}\mathrm{Im}\, z \le \rho_*$ for $\rho_* \sim 1$ from Lemma 8.5.1. Let (β_0, b_0, l_0) and (β, b, l) be the triple of eigenvalue, right and left eigenvector for the operators B_0 and B corresponding to the isolated eigenvalue in $D_\varepsilon(0)$ from Lemma 8.5.1, respectively. Then with a properly chosen normalization of the eigenvectors we have*

$$b_0 = C_{q^*,q}[f_u] + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z), \qquad l_0 = C_{q,q^*}^{-1}[f_u] + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z), \tag{8.5.13a}$$

$$\beta_0 = \frac{\mathrm{Im}\, z}{\rho}\frac{\pi}{\langle f_u^2\rangle} + \mathcal{O}(\rho^{-2}(\mathrm{Im}\, z)^2) = \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)\,, \tag{8.5.13b}$$

as well as

$$b \;=\; b_0 + 2\mathrm{i}\rho C_{q^*,q}(\mathrm{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] + \mathcal{O}(\rho^2 + \mathrm{Im}\, z)\,, \tag{8.5.14a}$$

$$l \;=\; l_0 - 2\mathrm{i}\rho C_{q,q^*}^{-1}(\mathrm{Id} - FC_s)^{-1}Q_{s,F}^*F[sf_u^2] + \mathcal{O}(\rho^2 + \mathrm{Im}\, z)\,, \tag{8.5.14b}$$

$$\beta\langle l\,, b\rangle \;=\; \pi\rho^{-1}\mathrm{Im}\, z - 2\mathrm{i}\rho\sigma + 2\rho^2\Big(\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^3 + \mathrm{Im}\, z + \rho^{-2}(\mathrm{Im}\, z)^2)\,. \tag{8.5.14c}$$

Furthermore, let P_0 and P be the spectral projections corresponding to the isolated eigenvalue of B_0 and B, respectively. Then with $Q_0 := \mathrm{Id} - P_0$ and $Q := \mathrm{Id} - P$ we have

$$\|B^{-1}Q\| + \|B^{-1}Q\|_2 + \|B_0^{-1}Q_0\| \lesssim 1. \tag{8.5.15}$$

Moreover, we have

$$\|b\| \lesssim 1, \qquad \|l\| \lesssim 1. \tag{8.5.16}$$

For later use, we record some identities here. From (8.5.10) in the proof of Lemma 8.5.1 with $C_s[f_u] = f_u$, we obtain the first relation in

$$\begin{aligned} P_{s,F} &= \frac{\langle f_u\,, \cdot\,\rangle}{\langle f_u^2\rangle} f_u + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z), \\ P_{s,F}^* &= P_{s,F} + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z), \qquad Q_{s,F}^* = Q_{s,F} + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z). \end{aligned} \tag{8.5.17}$$

This first relation together with $f_u = f_u^*$ implies the second and third one. Moreover, the definitions of B_0 and Q_0 yield

$$B_0^{-1}Q_0 = C_{q^*,q}(\mathrm{Id} - C_sF)^{-1}Q_{s,F}C_{q^*,q}^{-1}. \tag{8.5.18}$$

By a direct computation starting from the definitions of f_u in (8.5.4) and u in (8.3.1), we obtain

$$\langle f_u q q^* \rangle = \rho^{-1} \langle \mathrm{Im}\, m \rangle = \pi. \tag{8.5.19}$$

PROOF. Using (8.5.5) and $C_s[f_u] = f_u$, we see that

$$B_0^* C_{q,q^*}^{-1}[f_u] \;=\; \rho^{-1}(\mathrm{Im}\, z)\mathbb{1}\,, \qquad B_0 C_{q^*,q}[f_u] \;=\; \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)\,. \tag{8.5.20}$$

We set $b_0 \coloneqq P_0 C_{q^*,q}[f_u]$ and $l_0 \coloneqq P_0^* C_{q,q^*}^{-1}[f_u]$ which amounts to a normalization as β_0 is a nondegenerate eigenvalue. The representations of b_0 and l_0 in (8.5.13a) follow by simple perturbation theory because β_0 is a nondegenerate isolated eigenvalue. The expression for β_0 in (8.5.13b) is seen by taking the scalar product with b_0 in the first identity of (8.5.20) as well as using (8.5.13a) and (8.5.19).

The expansions (8.5.14) follow by analytic perturbation theory. Indeed, $b = b_0 + b_1 + \mathcal{O}(\rho^2)$ and $l = l_0 + l_1 + \mathcal{O}(\rho^2)$ with $b_1 \coloneqq -(B_0 - \beta_0 \mathrm{Id})^{-1} Q_0 E[b_0]$ and $l_1 \coloneqq -(B_0^* - \bar{\beta}_0 \mathrm{Id})^{-1} Q_0^* E^*[l_0]$ (cf. Lemma 8.13.1 with E satisfying (8.5.3)). Here the invertibility of $B_0 - \beta_0\mathrm{Id}$ on the range of Q_0 is seen from the second part of Lemma 8.5.1 with $T = B_0$. In fact,

$$(B_0 - \beta_0 \mathrm{Id})^{-1} Q_0 = B_0^{-1} Q_0 + \mathcal{O}(\beta_0). \tag{8.5.21}$$

Furthermore, we use (8.5.13a) and obtain the first equalities below:

$$\begin{aligned} E[b_0] &= C_{q^*,q}(C_s - C_u) F[f_u] + \mathcal{O}(\mathrm{Im}\, z) \\ &= -2\mathrm{i}\rho C_{q^*,q}[s f_u^2] + 2\rho^2 C_{q^*,q}[f_u^3] + \mathcal{O}(\rho^3 + \mathrm{Im}\, z), \end{aligned} \tag{8.5.22a}$$

$$\begin{aligned} E^*[l_0] &= C_{q,q^*}^{-1} F (C_s - C_u^*)[f_u] + \mathcal{O}(\mathrm{Im}\, z) \\ &= 2\mathrm{i}\rho C_{q,q^*}^{-1} F[s f_u^2] + 2\rho^2 C_{q,q^*}^{-1} F[f_u^3] + \mathcal{O}(\rho^3 + \mathrm{Im}\, z). \end{aligned} \tag{8.5.22b}$$

For the second equality in (8.5.22a), we used (8.5.5), $\|C_s - C_u\| = \mathcal{O}(\rho)$ and $(C_s - C_u)[f_u] = 2(\mathrm{Im}\, u - \mathrm{i}\mathrm{Re}\, u)(\mathrm{Im}\, u) f_u = -2\mathrm{i}\rho s f_u^2 + 2\rho^2 f_u^3 + \mathcal{O}(\rho^3)$ due to (8.5.2). For the second equality in (8.5.22b), we applied $(C_s - C_u^*)[f_u] = 2\mathrm{i}\rho s f_u^2 + 2\rho^2 f_u^3 + \mathcal{O}(\rho^3)$.

For the proof of (8.5.14c), we start from (8.13.3), use $E = \mathcal{O}(\rho)$ and obtain

$$\beta \langle l\,, b \rangle = \beta_0 \langle l_0\,, b_0 \rangle + \langle l_0\,, E[b_0] \rangle - \langle l_0\,, E B_0 (B_0 - \beta_0 \mathrm{Id})^{-2} Q_0 E[b_0] \rangle + \mathcal{O}(\rho^3). \tag{8.5.23}$$

Each of the terms on the right-hand side is computed individually. For the first term, we use $\langle l_0 , b_0\rangle = \langle f_u^2\rangle + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)$ due to (8.5.13a) and thus obtain from (8.5.13b) that

$$\beta_0\langle l_0 , b_0\rangle = \pi\rho^{-1}\mathrm{Im}\, z + \mathcal{O}(\rho^{-2}(\mathrm{Im}\, z)^2).$$

Using (8.5.13a) and (8.5.22) yields for the second term

$$\begin{aligned}\langle l_0 , E[b_0]\rangle &= -2\mathrm{i}\rho\langle s f_u^3\rangle + 2\rho^2\langle f_u^4\rangle + \mathcal{O}(\rho^3 + \mathrm{Im}\, z)\\ &= -2\mathrm{i}\rho\sigma + 2\rho^2\Big(\frac{\sigma^2}{\langle f_u^2\rangle} + \langle s f_u^2 , Q_{s,F}[s f_u^2]\rangle\Big) + \mathcal{O}(\rho^3 + \mathrm{Im}\, z),\end{aligned}$$

where we used $\mathrm{Id} = P_{s,F} + Q_{s,F}$ and $\langle s f_u^2 , P_{s,F}[s f_u^2]\rangle = \sigma^2/\langle f_u^2\rangle + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)$ by (8.5.17) in the last step.

For the third term, we use (8.5.13b) and $E = \mathcal{O}(\rho)$ which yields

$$\begin{aligned}\langle l_0 , E B_0(B_0 - \beta_0\mathrm{Id})^{-2}Q_0E[b_0]\rangle &= \langle E^*[l_0] , (B_0 - \beta_0\mathrm{Id})^{-1}Q_0E[b_0]\rangle + \mathcal{O}(\beta_0\|E\|^2)\\ &= \langle E^*[l_0] , B_0^{-1}Q_0E[b_0]\rangle + \mathcal{O}(\rho\mathrm{Im}\, z)\\ &= -4\rho^2\langle s f_u^2 , F(\mathrm{Id} - C_sF)^{-1}Q_{s,F}[s f_u^2]\rangle + \mathcal{O}(\rho\mathrm{Im}\, z + \rho^3).\end{aligned}$$

Here, we used (8.5.21) in the second step and (8.5.22) as well as (8.5.18) in the last step. Collecting the results for the three terms in (8.5.23) and using $C_s = C_s^*$ as well as $C_s[s f_u^2] = s f_u^2$ yield (8.5.14c).

The bounds in (8.5.15) and (8.5.16) follow directly from the analytic functional calculus and Lemma 8.5.1. □

Corollary 8.5.3 (Improved bound on B^{-1})**.** *Let Assumptions 8.4.5 hold true on an interval $I \subset \mathbb{R}$ for some $\eta_* \in (0, 1]$. Then, uniformly for all $z \in \mathbb{H}_{I,\eta_*}$, we have*

$$\|B^{-1}(z)\|_2 + \|B^{-1}(z)\| \lesssim \frac{1}{\rho(z)(\rho(z) + |\sigma(z)|) + \rho(z)^{-1}\mathrm{Im}\, z}. \tag{8.5.24}$$

Proof. If $\rho \geq \rho_*$ for some $\rho_* \sim 1$ then (8.5.24) have been shown in (8.4.22) as $|\sigma| \lesssim 1$. Therefore, we prove (8.5.24) for $\rho \leq \rho_*$ and a sufficiently small $\rho_* \sim 1$. By $\|S\|_{2\to\|\cdot\|} \lesssim 1$ and Lemma 8.12.2 (ii), it suffices to show the bound for $\|\cdot\|_2$. We follow the proof of (8.4.22) until (8.4.23). Hence, for the improved bound, we have to show that

$$|1 - \|F\|_2\langle f C_u^*[f]\rangle| \gtrsim \rho(\rho + |\sigma|) + \rho^{-1}\mathrm{Im}\, z. \tag{8.5.25}$$

We have $|1 - \|F\|_2 \langle f C_u^*[f]\rangle| \gtrsim \max\{1 - \|F\|_2, |1 - \langle f C_u^*[f]\rangle|\} \gtrsim \rho^{-1}\mathrm{Im}\, z + |1 - \langle f C_u^*[f]\rangle|$ by (8.4.5). We continue

$$|1-\langle f C_u^*[f]\rangle| = |1-\langle f u^* f u^*\rangle| \gtrsim \langle f \mathrm{Im}\, u f \mathrm{Im}\, u\rangle + |\langle f \mathrm{Im}\, u f \mathrm{Re}\, u\rangle| \gtrsim \rho^2 + \rho|\sigma| + \mathcal{O}(\rho^3 + \mathrm{Im}\, z).$$

Here, we used $1 \geq \langle f \mathrm{Re}\, u f \mathrm{Re}\, u\rangle$ due to $\|f\|_2 = 1$, (8.4.21) as well as $\langle f \mathrm{Im}\, u f \mathrm{Re}\, u\rangle = \rho \|f_u\|_2^{-2} \langle f_u^3 s\rangle + \mathcal{O}(\rho^3 + \mathrm{Im}\, z)$ by (8.5.10) and (8.5.2). By possibly shrinking $\rho_* \sim 1$, we thus obtain (8.5.25). This completes the proof of (8.5.24). ☐

The remainder of this section is devoted to several results about the behaviour of $\rho(z)$, $\sigma(z)$ and $\psi(z)$ close to the real axis. They will be applied in the next section. We now prepare these results by extending q, u, f_u and s to the real axis.

Lemma 8.5.4 (Extensions of q, u, f_u and s)**.** *Let $I \subset \mathbb{R}$ be an interval, $\theta \in (0,1]$ and Assumptions 8.4.5 hold true on I for some $\eta_* \in (0,1]$. We set $I_\theta := \{\tau \in I\colon \mathrm{dist}(\tau, \partial I) \geq \theta\}$. Then we have*

(i) The functions q, u and f_u have unique uniformly $1/3$-Hölder continuous extensions to $\overline{\mathbb{H}}_{I_\theta,\eta_*}$.

(ii) The function $z \mapsto \rho(z)^{-1}\mathrm{Im}\, z$ has a unique uniformly $1/3$-Hölder continuous extension to $\overline{\mathbb{H}}_{I_\theta,\eta_*}$. In particular, we have

$$\lim_{z \to \tau_0} \rho(z)^{-1}\mathrm{Im}\, z = 0 \tag{8.5.26}$$

for all $\tau_0 \in \mathrm{supp}\,\rho \cap I_\theta$. Moreover, for $z \in \overline{\mathbb{H}}_{I_\theta,\eta_*}$, we have

$$\mathrm{dist}(z, \mathrm{supp}\,\rho) \gtrsim 1 \qquad \Longleftrightarrow \qquad \rho(z)^{-1}\mathrm{Im}\, z \gtrsim 1.$$

(iii) There is a threshold $\rho_* \sim 1$ such that $s = \mathrm{sign}(\mathrm{Re}\, u)$ has a unique uniformly $1/3$-Hölder continuous extension to $\{w \in \overline{\mathbb{H}}_{I_\theta,\eta_*} : \rho(w) \leq \rho_*\}$.

Proof. For the proof of (i), we will show below that

$$f_m(z) := \rho(z)^{-1}\mathrm{Im}\, m(z)$$

is uniformly 1/3-Hölder continuous on $\mathbb{H}_{I_\theta,\eta_*}$. Indeed, this suffices to obtain the Hölder-continuity of q and u since their definitions in (8.3.1) can be rewritten as

$$\begin{aligned} q &= |h^{-1/2}gh^{-1/2} + \mathrm{i}\mathbb{1}|^{1/2}h^{1/2} = \Big(\rho(z)^2\mathbb{1} + f_m^{-1/2}gf_m^{-1}gf_m^{-1/2}\Big)^{1/4} f_m^{1/2}, \\ u &= \frac{\rho(z)w}{|\rho(z)w|} = \frac{\mathrm{i}\rho(z)\mathbb{1} + f_m^{-1/2}gf_m^{-1/2}}{|\mathrm{i}\rho(z)\mathbb{1} + f_m^{-1/2}gf_m^{-1/2}|}, \end{aligned} \tag{8.5.27}$$

where $g = \operatorname{Re} m$, $h = \operatorname{Im} m$, w is defined in (8.3.1) and $z \in \mathbb{H}$ is arbitrary. Since $|\rho(z)w| \sim 1$ and $f_m \sim 1$ on $\mathbb{H}_{I_\theta,\eta_*}$ by (8.4.21) as well as (8.4.17) and m, hence ρ and $\operatorname{Re} m$ are Hölder-continuous on $I_\theta \times \mathrm{i}[0,\infty)$ (Proposition 8.4.7), it thus suffices to show that f_m is uniformly Hölder-continuous to conclude from (8.5.27) that q and u are Hölder-continuous. As $f_u = \rho^{-1}\operatorname{Im} u = (q^*)^{-1} f_m q^{-1}$, the Hölder-continuity of f_m, the Hölder-continuity of q and the upper and lower bounds on q from (8.4.21) imply that f_u can be extended to a 1/3-Hölder continuous function on $\overline{\mathbb{H}}_{I_\theta,\eta_*}$.

Therefore, we now complete the proof of (i) by showing the 1/3-Hölder continuity of f_m. To that end, we distinguish three subsets of $\mathbb{H}_{I_\theta,\eta_*}$.

Case 1: On the set $\{z \in \mathbb{H}_{I_\theta,\eta_*} : \rho(z) \geq \rho_*\}$ for any $\rho_* \sim 1$, the uniform 1/3-Hölder continuity of f_m follows from $\rho(z) \gtrsim 1$ and the 1/3-Hölder continuity of m from Proposition 8.4.7.

Case 2: In order to analyze f_m on the set $\{z \in \mathbb{H}_{I_\theta,\eta_*} : \rho(z) \leq \rho_*\}$ for some $\rho_* \sim 1$ to be chosen later, we take the imaginary part of the Dyson equation, (8.2.3), at $z \in \mathbb{H}$ and obtain

$$B_*[\operatorname{Im} m] = (\operatorname{Im} z) m^* m, \qquad B_* := \mathrm{Id} - C_{m^*,m}S, \tag{8.5.28}$$

where $m = m(z)$. We follow the proof of (8.5.24) in Corollary 8.5.3 and use

$$\mathrm{Id} - C_{m^*,m}S = C_{q^*,q}C_{u^*,u}(C_{u,u^*} - F)C_{q^*,q}^{-1}$$

instead of (8.4.3) to see the invertibility of B_* for each $z \in \mathbb{H}_{I,\eta_*}$ and

$$\|B_*^{-1}(z)\|_2 + \|B_*^{-1}(z)\| \lesssim \frac{1}{\rho(z)(\rho(z) + |\sigma(z)|) + \rho(z)^{-1}\operatorname{Im} z} \tag{8.5.29}$$

for all $z \in \mathbb{H}_{I,\eta_*}$. Since B_* is invertible for any $z \in \mathbb{H}_{I,\eta_*}$, we conclude from (8.5.28) that

$$f_m(z) = \pi \frac{\operatorname{Im} m(z)}{\langle \operatorname{Im} m(z) \rangle} = \pi \frac{B_*^{-1}[m^*m]}{\langle B_*^{-1}[m^*m] \rangle} \tag{8.5.30}$$

for all $z \in \mathbb{H}_{I_\theta,\eta_*}$.

On the set $\{z \in \mathbb{H}_{I_\theta,\eta_*} : \rho(z)^{-1}\mathrm{Im}\, z \geq \rho_*\}$ for any $\rho_* \sim 1$, $B_*^{-1}[m^*m]$ is uniformly 1/3-Hölder continuous due to (8.5.29) and the 1/3-Hölder continuity of m. Moreover, from (8.4.5) and $\mathrm{Im}\, u \sim \rho\mathbb{1}$, we see that $1 - \|F\|_2 \sim 1$ if $\rho(z)^{-1}\mathrm{Im}\, z \gtrsim 1$. Hence, by Lemma 8.12.3 in Appendix 8.12 below, $(\mathrm{Id} - C_{u^*,u}F)^{-1}$ is positivity-preserving and satisfies

$$(\mathrm{Id} - C_{u^*,u}F)^{-1}[xx^*] \geq xx^* \tag{8.5.31}$$

for any $x \in \mathcal{A}$. We conclude that $B_*^{-1} = C_{q^*,q}(\mathrm{Id} - C_{u^*,u}F)^{-1}C_{q^*,q}^{-1}$ is positivity-preserving. Together with (8.4.21), (8.5.31) implies $\langle B_*^{-1}[m^*m]\rangle \gtrsim 1$ as $\|m(z)^{-1}\| \lesssim 1$ by $\|a\| \lesssim 1$, $\|S\| \lesssim 1$ and (8.2.3). Thus, (8.5.30) yields the uniform 1/3-Hölder continuity of f_m on $\{z \in \mathbb{H}_{I_\theta,\eta_*} : \rho(z)^{-1}\mathrm{Im}\, z \geq \rho_*\}$ for any $\rho_* \sim 1$.

Case 3: We now show that f_m is Hölder-continuous on $\{z \in \mathbb{H}_{I_\theta,\eta_*} : \rho(z)+\rho(z)^{-1}\mathrm{Im}\, z \leq \rho_*\}$ for some sufficiently small $\rho_* \sim 1$. In fact, Lemma 8.5.1 applied to $T = B_*$ yields the existence of a unique eigenvalue β_* of B_* of smallest modulus. Inspecting the proof of Corollary 8.5.2 for B reveals that this proof only used $B = B_0 + \mathcal{O}(\rho)$ about B. Therefore, the same argument works if B is replaced by B_* since $B_* = B_0 + \mathcal{O}(\rho)$ (compare the proof of Lemma 8.5.1). We thus find a right eigenvector b_* and a left eigenvector l_* of B_* associated to β_*, i.e.,

$$B_*[b_*] = \beta_* b_*, \qquad (B_*)^*[l_*] = \overline{\beta_*} l_*,$$

which satisfy

$$b_* = b_0 + \mathcal{O}(\rho) = q^* f_u q + \mathcal{O}(\rho + \rho^{-1}\mathrm{Im}\, z), \tag{8.5.32a}$$

$$l_* = l_0 + \mathcal{O}(\rho) = q^{-1} f_u (q^*)^{-1} + \mathcal{O}(\rho + \rho^{-1}\mathrm{Im}\, z), \tag{8.5.32b}$$

$$\beta_* \langle l_*, b_*\rangle = \pi\rho^{-1}\mathrm{Im}\, z + \mathcal{O}(\rho + \rho^{-2}(\mathrm{Im}\, z)^2). \tag{8.5.32c}$$

Moreover, we have

$$\|B_*^{-1}Q_*\| + \|B_*^{-1}Q_*\|_2 \lesssim 1, \tag{8.5.33}$$

where Q_* denotes the spectral projection of B_* to the complement of the spectral subspace of β_*.

Therefore, as $\beta_* \neq 0$ (cf. (8.5.29)) if $\operatorname{Im} z > 0$, we obtain

$$\operatorname{Im} m = (\operatorname{Im} z) B_*^{-1}[m^* m] = (\operatorname{Im} z)\Big(\beta_*^{-1}\frac{\langle l_*\,, m^* m\rangle}{\langle l_*\,, b_*\rangle} b_* + B_*^{-1}Q_*[m^* m]\Big).$$

Consequently, as $\operatorname{Im} m > 0$, we have

$$\frac{\operatorname{Im} m}{\langle \operatorname{Im} m\rangle} = \frac{\langle l_*\,, m^* m\rangle b_* + \beta_* \langle l_*\,, b_*\rangle B_*^{-1} Q_*[m^* m]}{\langle l_*\,, m^* m\rangle\langle b_*\rangle + \beta_*\langle l_*\,, b_*\rangle \langle B_*^{-1}Q_*[m^* m]\rangle}, \tag{8.5.34}$$

which together with (8.5.30) shows that f_m is uniformly 1/3-Hölder continuous on $\{z \in \mathbb{H}_{I_\theta,\eta_*} : \rho(z) + \rho(z)^{-1}\operatorname{Im} z \leq \rho_*\}$. Here, we used that B_* and, thus, β_*, l_*, b_* and $B_*^{-1}Q_*$ are 1/3-Hölder continuous and the denominator in (8.5.34) is $\gtrsim 1$ due to

$$\begin{aligned}\langle l_*\,, m^* m\rangle &= \langle q^{-1} f_u (q^*)^{-1} q^* u^* q q^* u q\rangle + \mathcal{O}(\rho + \rho^{-1}\operatorname{Im} z)\\ &= \rho^{-1}\operatorname{Im}\langle q^* u u u^* q\rangle + \mathcal{O}(\rho + \rho^{-1}\operatorname{Im} z) = \pi + \mathcal{O}(\rho + \rho^{-1}\operatorname{Im} z)\end{aligned}$$

by (8.5.32a) and (8.5.32b) as well as $\langle b_*\rangle = \pi + \mathcal{O}(\rho + \rho^{-1}\operatorname{Im} z)$ by (8.5.19). Here, we also used (8.5.32c) and (8.5.33). This completes the proof of (i).

For the proof of (ii), we multiply (8.5.28) by $\rho(z)^{-1}(m^* m)^{-1}$ which yields

$$\rho(z)^{-1}\operatorname{Im} z = (m^* m)^{-1} B_*[f_m].$$

Owing to $m^* m \geq \|m^{-1}\|^{-2} \gtrsim 1$ as well as the 1/3-Hölder continuity of m, B_* and f_m, we obtain the same regularity for $z \mapsto \rho(z)^{-1}\operatorname{Im} z$. Since $\lim_{\eta\downarrow 0} \rho(\tau + \mathrm{i}\eta)^{-1}\eta = 0$ for $\tau \in \operatorname{supp}\rho \cap I_\theta$ satisfying $\rho(\tau) > 0$, the continuity of $\rho(z)^{-1}\operatorname{Im} z$ directly implies (8.5.26). If $\operatorname{dist}(z, \operatorname{supp}\rho) \gtrsim 1$ then $\rho(z)^{-1}\operatorname{Im} z \gtrsim 1$ as $\rho(z) \leq \operatorname{Im} z / \operatorname{dist}(z, \operatorname{supp}\rho)^2$ which can be seen by applying $\langle\,\cdot\,\rangle$ to the second bound in (8.3.7). Conversely, if $\operatorname{dist}(z, \operatorname{supp}\rho) \lesssim 1$ then the Hölder-continuity of $\rho(z)^{-1}\operatorname{Im} z$ and (8.5.26) imply $\rho(z)^{-1}\operatorname{Im} z \lesssim 1$.

We now turn to the proof of (iii). Owing to the first relation in (8.5.2), there is $\rho_* \sim 1$ such that $|\operatorname{Re} u| \geq \frac{1}{2}\mathbb{1}$ if $z \in \mathbb{H}_{I_\theta,\eta_*}$ satisfies $\rho(z) \leq \rho_*$. Therefore, we find a smooth function $\varphi\colon \mathbb{R} \to [-1, 1]$ such that $\varphi(t) = 1$ for all $t \in [1/2, \infty)$, $\varphi(t) = -1$ for all $t \in (-\infty, -1/2]$ and $s(z) = \operatorname{sign}(\operatorname{Re} u(z)) = \varphi(\operatorname{Re} u(z))$ for all $z \in \mathbb{H}_{I_\theta,\eta_*}$ satisfying $\rho(z) \leq \rho_*$. Since φ is smooth, we conclude that φ is an *operator Lipschitz function* [8, Theorem 1.6.1], i.e., $\|\varphi(x) - \varphi(y)\| \leq C\|x - y\|$ for all self-adjoint $x, y \in \mathcal{A}$. Hence, we conclude

$$\|s(z_1) - s(z_2)\| = \|\varphi(\operatorname{Re} u(z_1)) - \varphi(\operatorname{Re} u(z_2))\| \lesssim \|z_1 - z_2\|^{1/3},$$

where we used that φ is operator Lipschitz and u is $1/3$-Hölder continuous in the last step. This completes the proof of Lemma 8.5.4. □

Lemma 8.5.5 (Properties of ψ and σ)**.** *Let $I \subset \mathbb{R}$ be an interval and $\theta \in (0,1]$. If m satisfies Assumptions 8.4.5 on I for some $\eta_* \in (0,1]$ then there is a threshold $\rho_* \sim 1$ such that, with*

$$\mathbb{H}_{\mathrm{small}} := \{w \in \mathbb{H}_{I_\theta,\eta_*} : \rho(w) + \rho(w)^{-1}\operatorname{Im} w \le \rho_*\},$$

we have

(i) The functions σ and ψ defined in (8.5.12) have unique uniformly $1/3$-Hölder continuous extensions to $\{z \in \overline{\mathbb{H}}_{I_\theta,\eta_*} : \rho(z) \le \rho_*\}$ and $\overline{\mathbb{H}}_{\mathrm{small}}$, respectively.

(ii) Uniformly for all $z \in \overline{\mathbb{H}}_{\mathrm{small}}$, we have

$$\psi(z) + \sigma(z)^2 \sim 1. \tag{8.5.35}$$

Proof. For the proof of (i), we choose $\rho_* \sim 1$ so small that all parts of Lemma 8.5.4 are applicable. Thus, Lemma 8.5.4 and $\sigma = \langle s f_u^3 \rangle$ yield (i) for σ. Similarly, since q is now defined on $\overline{\mathbb{H}}_{I_\theta,\eta_*}$, we can define F via (8.3.4) on this set as well. Moreover, owing to the uniform $1/3$-Hölder continuity of q from Lemma 8.5.4, F is uniformly $1/3$-Hölder continuous on $\overline{\mathbb{H}}_{I_\theta,\eta_*}$. Hence, using Lemma 8.5.1 for $T = \mathrm{Id} - C_s F$, the Hölder-continuity of s and f_u, the function ψ has a unique $1/3$-Hölder continuous extension to $\overline{\mathbb{H}}_{\mathrm{small}}$. This completes the proof of (i) for ψ.

We now turn to the proof of (ii). In fact, we will show (8.5.35) only on $\{w \in \mathbb{H}_{I_\theta,\eta_*} : \rho(w) + \rho(w)^{-1}\operatorname{Im} w \le \rho_*\}$, where $\rho_* \sim 1$ is chosen small enough such that Lemma 8.5.1 is applicable. By the continuity of σ and ψ, the bound (8.5.35) immediately extends to $\mathbb{D}_{\rho_*,\theta}$. Instead of (8.5.35), we will prove that

$$\langle x\,, (\mathrm{Id} + F)(\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x] \rangle + \langle f_u\,, x \rangle^2 \sim \|x\|_2^2 \tag{8.5.36}$$

for all $x \in \mathcal{A}$ satisfying $C_s[x] = x$ and $x = x^*$. Since these conditions are satisfied by $x = s f_u^2$, (8.5.36) immediately implies (8.5.35). In fact, the upper bound in (8.5.36) follows from $\|(\mathrm{Id} - C_s F)^{-1} Q_{s,F}\|_2 \lesssim 1$ by Lemma 8.5.1, $\|F\|_2 \leq 1$ and $f_u \sim 1$ due to (8.5.8).

From $C_s[x] = x$, we conclude

$$\begin{aligned}\langle x\,, (\mathrm{Id} + F)(\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x]\rangle &= \langle x\,, (\mathrm{Id} + C_s F)(\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x]\rangle \\ &= \langle x\,, ((C_s F - \mathrm{Id}) + 2\mathrm{Id})(\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x]\rangle \\ &= \langle x\,, (-\mathrm{Id} + 2(\mathrm{Id} - C_s F)^{-1}) Q_{s,F}[x]\rangle.\end{aligned} \tag{8.5.37}$$

Using (8.5.17) and $C_s[f_u] = f_u$, we see that

$$C_s P_{s,F}[x] = P_{s,F}[x] + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z), \qquad C_s Q_{s,F}[x] = Q_{s,F}[x] + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z) \tag{8.5.38}$$

for $x \in \mathcal{A}$ satisfying $C_s[x] = x$.
When applied to (8.5.37), the expansion (8.5.38) and $(\mathrm{Id} - F C_s)^{-1} = C_s(\mathrm{Id} - C_s F)^{-1} C_s$ yield

$$\begin{aligned}&\langle x\,, (\mathrm{Id} + F)(\mathrm{Id} - C_s F)^{-1}) Q_{s,F}[x]\rangle \\ &\quad= \langle Q_{s,F}[x]\,, (-\mathrm{Id} + (\mathrm{Id} - C_s F)^{-1} + (\mathrm{Id} - F C_s)^{-1}) Q_{s,F}[x]\rangle + \mathcal{O}(\|x\|_2^2 \rho^{-1}\mathrm{Im}\, z) \\ &\quad= \langle Q_{s,F}[x]\,, (\mathrm{Id} - F C_s)^{-1}(\mathrm{Id} - F^2)(\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x]\rangle + \mathcal{O}(\|x\|_2^2 \rho^{-1}\mathrm{Im}\, z) \\ &\quad= \langle (\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x]\,, Q_f(\mathrm{Id} - F^2) Q_f (\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x]\rangle + \mathcal{O}(\|x\|_2^2 \rho^{-1}\mathrm{Im}\, z) \\ &\quad\gtrsim \|Q_f(\mathrm{Id} - C_s F)^{-1} Q_{s,F}[x]\|_2^2 + \mathcal{O}(\|x\|_2^2 \rho^{-1}\mathrm{Im}\, z) \\ &\quad\gtrsim \|Q_{s,F}[x]\|_2^2 + \mathcal{O}(\|x\|_2^2 \rho^{-1}\mathrm{Im}\, z).\end{aligned} \tag{8.5.39}$$

Here, in the first step, we also used the second and third relation in (8.5.17). In the third step, we then defined the orthogonal projections $P_f := \langle f\,, \cdot\rangle f$ and $Q_f := \mathrm{Id} - P_f$, where $Ff = \|F\|_2 f$ (cf. Assumptions 8.4.5 (ii)), and inserted Q_f using

$$P_f Q_{s,F} = \mathcal{O}(\rho^{-1}\mathrm{Im}\, z) \tag{8.5.40}$$

which follows from (8.5.10) and (8.5.17). We also used that $Q_{s,F}$ commutes with $(\mathrm{Id} - C_s F)^{-1}$. The fourth step is a consequence of (8.4.7) and (8.4.18). In the last step, we employed $Q_f Q_{s,F} = Q_{s,F} + \mathcal{O}(\rho^{-1}\mathrm{Im}\, z)$ by (8.5.40) and $\|\mathrm{Id} - C_s F\|_2 \leq 2$.

By (8.5.17), we have $\|P_{s,F}[x]\|_2^2 = \langle f_u, x\rangle^2 + \mathcal{O}(\|x\|_2^2 \rho^{-1} \mathrm{Im}\, z)$ if $x = x^*$. Combining this observation with (8.5.39) proves (8.5.36) up to terms of order $\mathcal{O}(\|x\|_2^2\rho^{-1}\mathrm{Im}\, z)$. Hence, possibly shrinking $\rho_* \sim 1$ and requiring $\rho(z)^{-1}\mathrm{Im}\, z \leq \rho_*$ complete the proof of the lemma. □

Remark 8.5.6 (Auxiliary quantities as functions of m)**.** Inspecting the proofs of Lemma 8.5.4 and Lemma 8.5.5 reveals that q, u, f_u and s as well as σ and ψ are Lipschitz-continuous functions of m. More precisely, we define

$$\mathcal{M} := \Big\{ m \in \mathcal{A} \colon m \text{ satisfies (8.2.3) for some data pair } (a, S) \text{ and some } z \in \mathbb{H} \\ \text{such that } |z| \leq k_9,\ \mathrm{Im}\, m \in \mathcal{A}_+ \text{ and } m, a, S \text{ satisfy Assumptions 8.4.5 at } z \Big\}.$$

for some $k_9 > 0$. Then we have

(i) The functions q, u and f_u are uniformly Lipschitz-continuous functions of m on $\mathcal{M}$.

(ii) There is $\rho_* \sim 1$ such that the functions s and σ are uniformly Lipschitz-continuous as functions of m on $\{m \in \mathcal{M} \colon \langle \mathrm{Im}\, m\rangle \leq \pi\rho_*\}$.

(iii) There is $\rho_* \sim 1$ such that the function ψ is uniformly Lipschitz-continuous as function of m on $\{m \in \mathcal{M} \colon \langle \mathrm{Im}\, m\rangle + \pi^2 \langle \mathrm{Im}\, m\rangle^{-1} \mathrm{Im}\, z \leq \pi\rho_*$, where $z \in \mathbb{H}$ is the spectral parameter in (8.2.3)$\}$.

Here, we also consider k_9 in the definition of $\mathcal{M}$ a model parameter in addition to those introduced in Convention 8.4.6.

The careful analysis of the operator B and its inverse allows for the precise bounds on the derivatives of m in the following lemma.

Lemma 8.5.7 (Derivatives of m)**.** *Let $I \subset \mathbb{R}$ be an open interval and $\theta \in (0, 1]$. If Assumptions 8.4.5 hold true on I for some $\eta_* \in (0, 1]$ then there is $C \sim 1$ such that*

$$\|\partial_z^k m(\tau)\| \lesssim \frac{C^k}{\rho(\tau)^{3k-1}}$$

uniformly for all $\tau \in I_\theta$ satisfying $\rho(\tau) > 0$ and all $k \in \mathbb{N}$ satisfying $k \geq 1$.

PROOF. To indicate the mechanism, we first prove that

$$\|\partial_z m(\tau)\| \lesssim \rho(\tau)^{-2}, \qquad \|\partial_z^2 m(\tau)\| \lesssim \rho(\tau)^{-5}, \qquad \|\partial_z^3 m(\tau)\| \lesssim \rho(\tau)^{-8} \tag{8.5.41}$$

for all $\tau \in I_\theta$ satisfying $\rho(\tau) > 0$.

Since $\rho(\tau) > 0$, m is real analytic around τ by Proposition 8.4.7 and we can differentiate the Dyson equation, (8.2.3), with respect to z and evaluate at $z = \tau$. Differentiating (8.2.3) iteratively yields

$$\begin{aligned} &B[\partial_z m] = m^2, \qquad\qquad B[\partial_z^2 m] = 2(\partial_z m)m^{-1}(\partial_z m), \\ &B[\partial_z^3 m] = -6(\partial_z m)m^{-1}(\partial_z m)m^{-1}(\partial_z m) + 3(\partial_z^2 m)m^{-1}(\partial_z m) + 3(\partial_z m)m^{-1}(\partial_z^2 m) \end{aligned} \tag{8.5.42}$$

where $B = \mathrm{Id} - C_m S$ and $m := m(\tau)$. Since $\rho(\tau) > 0$, B is invertible by (8.5.24), (8.5.26) and the 1/3-Hölder continuity of m by Proposition 8.4.7.

We set $\rho := \rho(\tau)$. If $\rho > \rho_*$ for some $\rho_* \sim 1$ then (8.5.41) follows trivially from (8.5.42), $\|B^{-1}\| \lesssim 1$ by (8.5.24) and $\|m\| + \|m^{-1}\| \lesssim 1$.

We now prove (8.5.41) for $\rho \leq \rho_*$ and some sufficiently small $\rho_* \sim 1$. Under this assumption, Lemma 8.5.1 and Corollary 8.5.2 are applicable. In the remainder of this proof, the eigenvalue β, the eigenvectors l and b as well as the spectral projections P and Q are understood to be evaluated at τ. We will now estimate the image of B^{-1} applied to the right-hand sides of (8.5.42) in order to prove (8.5.41).

Inserting $P + Q = \mathrm{Id}$ on the right-hand side of the first identity in (8.5.42), inverting B and using

$$P = \frac{\langle l \,, \cdot \rangle}{\langle l \,, b\rangle} b$$

as well as $B^{-1}[b] = \beta^{-1} b$ yield

$$\partial_z m = \frac{\langle l \,, m^2\rangle}{\beta \langle l \,, b\rangle} b + B^{-1} Q[m^2]. \tag{8.5.43}$$

We will now estimate $\langle l \,, m^2\rangle$ and $\beta\langle l \,, b\rangle$. From $m = q^* s q + \mathcal{O}(\rho)$ by (8.5.2), (8.5.13a), (8.5.14b) and (8.5.26), we obtain

$$\langle l \,, m^2\rangle = \langle f_u s q q^* s\rangle + \mathcal{O}(\rho) = \pi + \mathcal{O}(\rho), \tag{8.5.44}$$

where we used $sf_us = f_us^2 = f_u$ and (8.5.19) in the last step.

From (8.5.14c) and (8.5.26), we conclude

$$\beta\langle l\,, b\rangle = -2\mathrm{i}\rho\sigma + \rho^2\Big(\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^3). \tag{8.5.45}$$

Here and in the remainder of the proof, σ, ψ, f_u, q and s are understood to be evaluated at τ.

Since σ and ψ are real, we conclude $|\beta\langle l\,, b\rangle| \sim \rho(\rho + |\sigma|)$ for $\rho_* \sim 1$ sufficiently small. As $\|B^{-1}Q\| \lesssim 1$ and $\|b\| \lesssim 1$, we thus obtain $\|\partial_z m\| \lesssim \rho^{-2}$ from (8.5.43).

Using (8.5.42), $\|\partial_z m\| \lesssim \rho^{-2}$ and $\|B^{-1}\| \lesssim \rho^{-2}$ yield

$$\partial_z^2 m = 2\frac{\langle l\,, m^2\rangle^2 \langle l\,, bm^{-1}b\rangle}{(\beta\langle l\,, b\rangle)^3} b + \mathcal{O}(\rho^{-4}) = \mathcal{O}(\rho^{-5}).$$

Here, in the last step, we used $\|b\| \lesssim 1$ and $|\langle l\,, bm^{-1}b\rangle| \lesssim |\sigma| + \rho$ due to the expansion

$$\langle l\,, bm^{-1}b\rangle = \langle q^{-1}f_u(q^*)^{-1}q^*f_u q q^{-1}s(q^*)^{-1}q^*f_u q\rangle + \mathcal{O}(\rho) = \sigma + \mathcal{O}(\rho) \tag{8.5.46}$$

as well as $|\beta\langle l\,, b\rangle| \sim \rho(\rho + |\sigma|)$ and $\langle l\,, m^2\rangle = \mathcal{O}(1)$. The proof of (8.5.46) is a consequence of (8.5.13a), (8.5.14a), (8.5.14b), (8.5.26), $m^{-1} = q^{-1}s(q^*)^{-1} + \mathcal{O}(\rho)$ by (8.5.2) as well as $q \sim 1$.

Similarly, owing to (8.5.42), we obtain

$$\partial_z^3 m = 12\frac{\langle l\,, m^2\rangle^3 \langle l\,, bm^{-1}b\rangle^2}{(\beta\langle l\,, b\rangle)^5} b + \mathcal{O}(\rho^{-8}) = \mathcal{O}(\rho^{-8}).$$

We now estimate $\partial_z^k m(z)$ for $k > 3$. To that end, we will fix a parameter $\alpha > 1$ and prove that there are $\rho_* \sim 1$, $C_1 \sim_\alpha 1$ and $C_2 \sim_\alpha 1$ such that, for $k \in \mathbb{N}$, we have

$$m^{(k)} := \partial_z^k m = \beta_k b + q_k, \tag{8.5.47}$$

where $m = m(\tau)$ for $\tau \in I_\theta$ satisfying $\rho := \rho(\tau) \leq \rho_*$ and $\beta_k \in \mathbb{C}$ and $q_k \in \operatorname{ran}Q$ satisfy

$$|\beta_k| \leq \frac{k!C_1C_2^{k-1}}{k^\alpha}\rho^{-3k+1}, \qquad \|q_k\| \leq \frac{k!C_1C_2^{k-1}}{k^\alpha}\rho^{-3k+2}. \tag{8.5.48}$$

Here, $\sim_\alpha$ indicates that the constants in the definition of the comparison relation $\sim$ will depend on α.

Before we prove (8.5.47) below, we note two auxiliary statements. First, as $\partial_z m^{-1} = -m^{-1}(\partial_z m)m^{-1}$ it is easy to check the following version of the usual Leibniz-rule:

$$\partial_z^k m^{-1} = \sum_{n=1}^{k} \sum_{\substack{a_1+\ldots+a_n=k \\ 1\le a_i\le k}} \frac{k!}{a_1!\ldots a_n!}(-1)^n\, m^{-1}m^{(a_1)}m^{-1}m^{(a_2)}\ldots m^{-1}m^{(a_n)}m^{-1} \tag{8.5.49}$$

for any $k \in \mathbb{N}$. Here, in the sum over $a_1 + \ldots + a_n = k$, the order of $a_1, \ldots, a_n$ has to be taken into account since m^{-1} and $m^{(a)}$ do not commute in general.

We prove (8.5.49) by induction on k: The case $k = 1$ is trivial. For the induction step, we obtain

$$\begin{aligned}
\partial_z^{k+1} m^{-1} &= \sum_{n=1}^{k}\sum_{j=1}^{k+1} \sum_{\substack{a_1+\ldots+a_{n+1}=k+1 \\ 1\le a_i\le k \\ a_j=1}} \frac{k!}{a_1!\ldots a_{n+1}!}(-1)^{n+1}m^{-1}m^{(a_1)}\ldots m^{(a_{n+1})}m^{-1} \\
&\quad + \sum_{n=1}^{k} \sum_{\substack{a_1+\ldots+a_n=k \\ 1\le a_i\le k}} \sum_{j=1}^{n} \frac{k!}{a_1!\ldots a_n!}(-1)^n m^{-1}m^{(a_1)}\ldots m^{(a_j+1)}\ldots m^{(a_n)}m^{-1} \\
&= \sum_{j=1}^{k+1} k!(-1)^{k+1}m^{-1}m^{(1)}m^{-1}\ldots m^{-1}m^{(1)}m^{-1} \\
&\quad + \sum_{n=2}^{k}\sum_{j=1}^{n}\Bigg(\sum_{\substack{a_1+\ldots+a_n=k+1 \\ 1\le a_i\le k+1 \\ a_j=1}} + \sum_{\substack{a_1+\ldots+a_n=k+1 \\ 1\le a_i\le k+1 \\ a_j\ge 2}} \Bigg) \frac{k!a_j}{a_1!\ldots a_n!}(-1)^n m^{-1}m^{(a_1)}\ldots m^{(a_n)}m^{-1} \\
&\quad + \frac{(k+1)!}{(k+1)!}(-1)^1 m^{-1}m^{(k+1)}m^{-1} \\
&= \sum_{n=1}^{k+1} \sum_{\substack{a_1+\ldots+a_n=k+1 \\ 1\le a_i\le k+1}} \frac{k!}{a_1!\ldots a_n!}(-1)^n m^{-1}m^{(a_1)}\ldots m^{(a_n)}m^{-1}.
\end{aligned}$$

Here, we used the product rule in the first step, where the first summand originates from differentiating the m^{-1} factors and the second summand from differentiating $m^{(a_j)}$. In the last step, we employed $a_1 + \ldots + a_n = k + 1$. This completes the proof of (8.5.49).

Second, we also have the following auxiliary bound. For all $k \in \mathbb{N}$, $n \in \mathbb{N}$ with $n \le k$ and $\alpha > 1$, we have

$$\sum_{\substack{a_1+\ldots+a_n=k \\ 1\le a_i\le k}} \frac{1}{a_1^\alpha \cdots a_n^\alpha} \le \frac{(2^{\alpha+1}\zeta(\alpha))^{n-1}}{k^\alpha}, \tag{8.5.50}$$

where $\zeta(\alpha) = \sum_{n=1}^{\infty} n^{-\alpha}$ is Riemann's zeta function.

We now prove (8.5.50) by induction on n and remark that the case $n = 1$ is trivial as the left- and right-hand side of (8.5.50) agree in this case. For the induction step, we assume $n + 1 \le k$ and obtain

$$\begin{aligned}\sum_{a_1+\ldots+a_{n+1}=k} \frac{1}{a_1^\alpha \ldots a_{n+1}^\alpha} &= \sum_{a=1}^{k} \frac{1}{a^\alpha} \sum_{a_1+\ldots+a_n=k-a} \frac{1}{a_1^\alpha \ldots a_n^\alpha} \\ &\le \sum_{a=1}^{k} \frac{(2^{\alpha+1}\zeta(\alpha))^{n-1}}{a^\alpha (k-a)^\alpha} \\ &\le 2\sum_{a=1}^{k/2} \frac{(2^{\alpha+1}\zeta(\alpha))^{n-1}}{a^\alpha (k/2)^\alpha} \\ &\le \frac{(2^{\alpha+1}\zeta(\alpha))^{n}}{k^\alpha}\end{aligned}$$

for $\alpha > 1$. Here, we used the induction hypothesis in the second step and $a(k-a) \ge ak/2$ for $1 \le a \le k$ in the third step. This completes the proof of (8.5.50).

We now show (8.5.47) and (8.5.48) by induction on k. The initial step of the induction with $k = 1$ has been established in (8.5.43) with $\beta_1 = \langle l\,, m^2\rangle/(\beta\langle l\,, b\rangle)$, $q_1 = B^{-1}Q[m^2]$ and some sufficiently large $C_1 \sim 1$. Next, we establish the induction step by proving (8.5.47) and (8.5.48) under the assumption that they hold true for all derivatives of lower order. From the induction hypothesis, we conclude

$$\|m^{(a)}\| \le \frac{k! C_1 C_2^{a-1}}{k^\alpha \rho^{3a-1}} (\|b\| + \rho) \tag{8.5.51}$$

for all $a \in \mathbb{N}$ satisfying $1 \le a \le k-1$.

For $k \ge 2$, we differentiate (8.2.3) k-times and obtain

$$B[\partial_z^k m] = r_k := \partial_z^k m + m\Big(\partial_z^k m^{-1}\Big)m. \tag{8.5.52}$$

By separating the contributions for $n = 1$ and $n \ge 2$ in (8.5.49), we conclude

$$\begin{aligned}r_k = &\sum_{n=3}^{k} \sum_{\substack{a_1+\ldots+a_n=k\\ 1\le a_i<k-1}} \frac{k!}{a_1! \ldots a_n!} (-1)^n\, m^{(a_1)} m^{-1} \ldots m^{-1} m^{(a_n)} \\ &+ \sum_{a=1}^{k-1} \frac{k!}{a!(k-a)!} m^{(a)} m^{-1} m^{(k-a)}.\end{aligned} \tag{8.5.53}$$

Since n is at least 3 in the first sum, we obtain from (8.5.51) and (8.5.50) that

$$\sum_{n=3}^{k} \sum_{\substack{a_1+\ldots+a_n=k \\ 1\le a_i<k-1}} \frac{k!}{a_1!\ldots a_n!}\|m^{(a_1)}m^{-1}\ldots m^{-1}m^{(a_n)}\| \le \frac{k!}{k^\alpha \rho^{3k-3}}(\|b\|+\rho)\sum_{n=3}^{k} C_1^n M_\alpha^{n-1} C_2^{k-n}, \tag{8.5.54}$$

where $M_\alpha := 2^{\alpha+2}\zeta(\alpha)\|m^{-1}\|(\|b\|+\rho)$. A similar argument yields

$$\sum_{a=1}^{k-1} \frac{k!}{a!(k-a)!}\|m^{(a)}m^{-1}m^{(k-a)}\| \le \frac{k!}{k^\alpha\rho^{3k-2}}(\|b\|+\rho)C_1^2 M_\alpha C_2^{k-2}.$$

Thus, we choose $C_2 \ge 2M_\alpha C_1$ and conclude

$$\|r_k\| \le \frac{k!}{k^\alpha\rho^{3k-2}}\frac{(\|b\|+\rho)M_\alpha C_1^2 C_2^k}{C_2^2(1-M_\alpha C_1/C_2)}.$$

Therefore, we obtain the bound on $\|q_k\|$ in (8.5.48) for $C_2 \sim 1$ sufficiently large since $q_k = Q[\partial_z^k m] = B^{-1}Q[r_k]$ and $\|B^{-1}Q\| \lesssim 1$.

Moreover, $\beta_k = \langle l\,, r_k\rangle/(\beta\langle l\,, b\rangle)$. Hence, by using the decomposition of r_k in (8.5.53) and (8.5.54), we obtain

$$|\beta_k| \le \frac{k!C_1C_2^{k-1}}{k^\alpha\rho^{3k-1}}\frac{\|l\|\rho^2}{|\beta\langle l\,,b\rangle|}\frac{(\|b\|+\rho)C_1^2M_\alpha^2}{C_2^2(1-M_\alpha C_1/C_2)} + \sum_{a=1}^{k-1}\frac{k!}{a!(k-a)!}\frac{|\langle l\,, m^{(a)}m^{-1}m^{(k-a)}\rangle|}{|\beta\langle l\,,b\rangle|}$$

We use (8.5.47) for $m^{(a)}$ and $m^{(k-a)}$ in the argument of the last sum, which yields

$$\begin{aligned}\frac{1}{a!(k-a)!}\frac{|\langle l\,, m^{(a)}m^{-1}m^{(k-a)}\rangle|}{|\beta\langle l\,,b\rangle|} &\le \frac{|\beta_a|}{a!}\frac{|\beta_{k-a}|}{(k-a)!}\frac{|\langle l\,, bm^{-1}b\rangle|}{|\beta\langle l\,,b\rangle|} \\ &\quad + \frac{C_1^2C_2^{k-2}}{a^\alpha(k-a)^\alpha\rho^{3k-1}}\frac{\rho^2\|l\|\|m^{-1}\|}{|\beta\langle l\,,b\rangle|}(2\|b\|+\rho) \\ &\le \frac{C_1^2C_2^{k-2}}{a^\alpha(k-a)^\alpha\rho^{3k-1}}\frac{\rho^2}{|\beta\langle l\,,b\rangle|} \\ &\quad \times \Big(|\langle l\,, bm^{-1}b\rangle|\rho^{-1} + \|l\|\|m^{-1}\|(2\|b\|+\rho)\Big)\end{aligned}$$

Here, we applied (8.5.48) to estimate q_a and q_{k-a} as well as β_a and β_{k-a}. Since $|\beta\langle l\,,b\rangle| \gtrsim \rho^2$ as shown below (8.5.45) and $|\langle l\,, bm^{-1}b\rangle| \lesssim \rho$ due to (8.5.46), we obtain the bound on $|\beta_k|$ in (8.5.48) by using (8.5.50) to perform the summation over a. This completes the induction argument, which yields (8.5.47) and (8.5.48) for all $k \in \mathbb{N}$ by possibly increasing $C_2 \sim 1$. By choosing, say, $\alpha = 2$, we immediately conclude Lemma 8.5.7 for $\tau \in I_\theta$ satisfying $\rho(\tau) \le \rho_*$. If $\rho(\tau) > \rho_*$ then $\|B^{-1}\| \lesssim 1$. Hence, a simple induction argument using (8.5.52) and (8.5.53), which hold true for $\rho(\tau) > \rho_*$ as well, yields some $C \sim 1$ such that

$$\|\partial_z^k m(\tau)\| \lesssim k! C^k$$

for all $k \in \mathbb{N}$ satisfying $k \geq 1$. Since $\rho(\tau) \lesssim 1$ for all $\tau \in I_\theta$, we obtain Lemma 8.5.7 in the missing regime. □

8.6. The cubic equation

The following Proposition 8.6.1 is the main result of this section. It asserts that m is determined by the solution to a cubic equation, (8.6.3) below, close to points $\tau_0 \in \operatorname{supp} \rho$ of small density $\rho(\tau_0)$. In Section 8.7, this cubic equation will allow for a classification of the small local minima of $\tau \mapsto \rho(\tau)$. To have a short notation for the elements of $\operatorname{supp} \rho$ of small density, we introduce the set

$$\mathbb{D}_{\varepsilon,\theta} := \{\tau \in \operatorname{supp} \rho \cap I : \rho(\tau) \in [0, \varepsilon],\ \operatorname{dist}(\tau, \partial I) \geq \theta\}$$

for $\varepsilon > 0$ and $\theta > 0$.

The leading order terms of the cubic and quadratic coefficients in (8.6.3) are given by $\psi(\tau_0)$ and $\sigma(\tau_0)$, respectively. For their definitions, we refer to Lemma 8.5.5 (i) and (8.5.12).

Proposition 8.6.1 (Cubic equation for shape analysis)**.** *Let $I \subset \mathbb{R}$ be an open interval and $\theta \in (0, 1]$. If Assumptions 8.4.5 hold true on I for some $\eta_* \in (0, 1]$ then there are thresholds $\rho_* \sim 1$ and $\delta_* \sim 1$ such that, for all $\tau_0 \in \mathbb{D}_{\rho_*,\theta}$, the following hold true:*

(a) For all $\omega \in [-\delta_*, \delta_*]$, we have

$$m(\tau_0 + \omega) - m(\tau_0) = \Theta(\omega) b + r(\omega), \tag{8.6.1}$$

where $\Theta \colon [-\delta_*, \delta_*] \to \mathbb{C}$ and $r \colon [-\delta_*, \delta_*] \to \mathcal{A}$ are defined by

$$\begin{aligned} \Theta(\omega) &:= \left\langle \frac{l}{\langle b, l \rangle}, m(\tau_0 + \omega) - m(\tau_0) \right\rangle, \\ r(\omega) &:= Q[m(\tau_0 + \omega) - m(\tau_0)]. \end{aligned} \tag{8.6.2}$$

Here, $l = l(\tau_0)$, $b = b(\tau_0)$ and $Q = Q(\tau_0)$ are the eigenvectors and spectral projection of $B(\tau_0)$ introduced in Corollary 8.5.2. We have $b = b^* + \mathcal{O}(\rho)$ and $l = l^* + \mathcal{O}(\rho)$ as well as $b + b^* \sim 1$ and $l + l^* \sim 1$ with $\rho = \rho(\tau_0) = \langle \operatorname{Im} m(\tau_0)\rangle/\pi$.

(b) *The function Θ satisfies the cubic equation*

$$\mu_3\Theta^3(\omega) + \mu_2\Theta^2(\omega) + \mu_1\Theta(\omega) + \omega\Xi(\omega) = 0 \tag{8.6.3}$$

for all $\omega \in [-\delta_*, \delta_*]$. The complex coefficients μ_3, μ_2, μ_1 and Ξ in (8.6.3) fulfill

$$\mu_3 = \psi + \mathcal{O}(\rho), \tag{8.6.4a}$$

$$\mu_2 = \sigma + \mathrm{i}\rho\Big(3\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^2), \tag{8.6.4b}$$

$$\mu_1 = 2\mathrm{i}\rho\sigma - 2\rho^2\Big(\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^3), \tag{8.6.4c}$$

$$\Xi(\omega) = \pi(1 + \nu(\omega)) + \mathcal{O}(\rho), \tag{8.6.4d}$$

where $\sigma = \sigma(\tau_0)$ as well as $\psi = \psi(\tau_0)$. For the error term $\nu(\omega)$, we have

$$|\nu(\omega)| \lesssim |\Theta(\omega)| + |\omega| \lesssim |\omega|^{1/3}. \tag{8.6.5}$$

for all $\omega \in [-\delta_*, \delta_*]$. Uniformly for $\tau_0 \in \mathbb{D}_{\rho_*,\theta}$, we have

$$\psi + \sigma^2 \sim 1. \tag{8.6.6}$$

(c) *Moreover, $\Theta(\omega)$ and $r(\omega)$ are bounded by*

$$|\Theta(\omega)| \lesssim \min\Big\{\frac{|\omega|}{\rho^2}, |\omega|^{1/3}\Big\}, \tag{8.6.7a}$$

$$\|r(\omega)\| \lesssim |\Theta(\omega)|^2 + |\omega|, \tag{8.6.7b}$$

uniformly for all $\omega \in [-\delta_*, \delta_*]$.

(d1) *If $\rho > 0$ then Θ and r are differentiable in ω at $\omega = 0$.*

(d2) *If $\rho = 0$ then we have*

$$\begin{aligned} &\operatorname{Im}\Theta(\omega) \geq 0, \\ &|\operatorname{Im}\nu(\omega)| \lesssim \operatorname{Im}\Theta(\omega), \qquad \|\operatorname{Im} r(\omega)\| \lesssim (|\Theta(\omega)| + |\omega|)\operatorname{Im}\Theta(\omega), \end{aligned} \tag{8.6.8}$$

for all $\omega \in [-\delta_*, \delta_*]$ *and* $\operatorname{Re}\Theta$ *is non-decreasing on the connected components of* $\{\omega \in [-\delta_*, \delta_*] \colon \operatorname{Im}\Theta(\omega) = 0\}$.

(e) *The function* $\sigma \colon \mathbb{D}_{\rho_*,\theta} \to \mathbb{R}$ *is uniformly* $1/3$*-Hölder continuous.*

The previous proposition is the analogue of Lemma 9.1 in [4]. The cubic equation for Θ, (8.6.3), will be obtained from an $\mathcal{A}$-valued quadratic equation for $\Delta := m(\tau_0 + \omega) - m(\tau_0)$ and the results of Section 8.5. In fact, we have

$$(\mathrm{Id} - C_m S)[\Delta] = \omega m^2 + \frac{\omega}{2}\Big(m\Delta + \Delta m\Big) + \frac{1}{2}\Big(mS[\Delta]\Delta + \Delta S[\Delta]m\Big), \tag{8.6.9}$$

where $\tau_0, \tau_0 + \omega \in I_\theta := \{\tau \in I \colon \operatorname{dist}(\tau, \partial I) \geq \theta\}$ and $m := m(\tau_0)$ (see the proof of Proposition 8.6.1 in Section 8.6.3 below for a derivation of (8.6.9)). Projecting (8.6.9) onto the direction b and its complement, where b is the unstable direction of B defined in Corollary 8.5.2, yields the cubic equation, (8.6.3), for the contribution Θ of Δ parallel with b. In the next subsection, this derivation is presented in a more abstract and transparent setting of a general $\mathcal{A}$-valued quadratic equation. After that, the coefficients of the cubic equation are computed in Lemma 8.6.3 in the setup of (8.6.9) before we prove Proposition 8.6.1 in Section 8.6.3.

8.6.1. General cubic equation. Let $B, T \colon \mathcal{A} \to \mathcal{A}$ be linear maps, $A \colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ a bilinear map and $K \colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ a map. For $\Delta, e \in \mathcal{A}$, we consider the quadratic equation

$$B[\Delta] - A[\Delta, \Delta] - T[e] - K[e, \Delta] = 0\,. \tag{8.6.10}$$

We view this as an equation for Δ, where e is a (small) error term. This quadratic equation is a generalization of the stability equation (8.6.9) for the Dyson equation, (8.2.3) (see (8.6.23) and (8.6.28) below for the concrete choices of B,T, A and K in the setting of (8.6.9)).

Suppose that B has a non-degenerate isolated eigenvalue β and a corresponding eigenvector b, i.e., $B[b] = \beta b$ and $D_r(\beta) \cap \operatorname{Spec}(B) = \{\beta\}$ for some $r > 0$. We denote the spectral projection corresponding to β and its complementary projection by P and Q, respectively, i.e.,

$$P := -\frac{1}{2\pi\mathrm{i}}\oint_{\partial D_r(\beta)}(B-\omega\mathrm{Id})^{-1}\mathrm{d}\omega = \frac{\langle l\,,\,\cdot\,\rangle}{\langle l\,,b\rangle}b\,, \qquad Q := \mathrm{Id}-P\,. \tag{8.6.11}$$

Here, $l \in \mathcal{A}$ is an eigenvector of B^* corresponding to its eigenvalue $\overline{\beta}$, i.e., $B^*[l] = \overline{\beta}l$. In the following, we will assume that

$$\begin{aligned} &\|B^{-1}Q[x]\| \lesssim \|x\|, \qquad |\langle l\,,b\rangle|^{-1} + \|b\| + \|l\| \lesssim 1, \qquad \|A[x,y]\| \lesssim \|x\|\|y\|, \\ &\|T[e]\| \lesssim \|e\|, \qquad\qquad \|K[e,y]\| \lesssim \|e\|\|y\| \end{aligned} \tag{8.6.12}$$

for all $x, y \in \mathcal{A}$ and the $e \in \mathcal{A}$ from (8.6.10). The guiding idea is that the main contribution in the decomposition

$$\Delta = \Theta b + Q[\Delta], \qquad \Theta := \frac{\langle l\,,\Delta\rangle}{\langle l\,,b\rangle} \tag{8.6.13}$$

is given by Θ, i.e., the coefficient of Δ in the direction b, under the assumption that Δ is small. If $A = K = 0$ then this would be a simple linear stability analysis of the equation $B[\Delta] =$ *small* around an isolated eigenvalue of B. The presence of the quadratic terms in (8.6.10) requires to follow second and third order terms carefully. In the following lemma, we show that the behaviour of Θ is governed by a scalar-valued cubic equation (see (8.6.14) below) and that $Q[\Delta]$ is indeed dominated by Θ. The implicit constants in (8.6.12) are the model parameters in Section 8.6.1.

Lemma 8.6.2 (General cubic equation)**.** *Let β be a non-degenerate isolated eigenvalue of B. Let $\Delta \in \mathcal{A}$ and $e \in \mathcal{A}$ satisfy* (8.6.10)*, Θ be defined as in* (8.6.13) *and the conditions in* (8.6.12) *hold true. Then there is $\varepsilon \sim 1$ such that if $\|\Delta\| \le \varepsilon$ then Θ satisfies the cubic equation*

$$\mu_3\Theta^3 + \mu_2\Theta^2 + \mu_1\Theta + \mu_0 = \tilde{e}, \tag{8.6.14}$$

with some $\tilde{e} = \mathcal{O}(|\Theta|^4 + |\Theta|\|e\| + \|e\|^2)$ and with coefficients

$$\begin{aligned} \mu_3 &= \langle l\,, A[b, B^{-1}QA[b,b]] + A[B^{-1}QA[b,b],b]\rangle, \\ \mu_2 &= \langle l\,, A[b,b]\rangle, \\ \mu_1 &= -\beta\langle l\,,b\rangle, \\ \mu_0 &= \langle l\,, T[e]\rangle. \end{aligned} \tag{8.6.15}$$

Moreover, we have

$$Q[\Delta] = B^{-1}QT[e] + \mathcal{O}(|\Theta|^2 + \|e\|^2). \tag{8.6.16}$$

If we additionally assume that $\operatorname{Im}\Delta \in \overline{\mathcal{A}}_+$, $l = l^*$ *and* $b = b^*$ *as well as*

$$B[x]^* = B[x^*], \quad A[x,y]^* = A[x^*,y^*], \quad T[e]^* = T[e], \quad K[e,y]^* = K[e,y^*] \tag{8.6.17}$$

for all $x, y \in \mathcal{A}$ *then there are* $\varepsilon \sim 1$ *and* $\delta \sim 1$ *such that* $\|\Delta\| \leq \varepsilon$ *and* $\|e\| \leq \delta$ *also imply*

$$\|\operatorname{Im} Q[\Delta]\| \lesssim (|\Theta| + \|e\|)\operatorname{Im}\Theta, \tag{8.6.18a}$$

$$|\operatorname{Im}\tilde{e}| \lesssim (|\Theta|^3 + \|e\|)\operatorname{Im}\Theta. \tag{8.6.18b}$$

PROOF. Setting $r := Q[\Delta]$, the quadratic equation (8.6.10) reads as

$$\Theta\beta b + Br \;=\; T[e] + A[\Delta,\Delta] + K[e,\Delta]. \tag{8.6.19}$$

By applying Q and afterwards B^{-1} to the previous relation, we conclude that

$$\begin{aligned} r &= B^{-1}QT[e] + \Theta^2 B^{-1}QA[b,b] + e_1, \\ e_1 &:= \Theta B^{-1}Q(A[b,r] + A[r,b]) + B^{-1}QA[r,r] + B^{-1}QK[e,\Delta]. \end{aligned} \tag{8.6.20}$$

We have $\|e_1\| \lesssim \|r\||\Theta| + \|r\|^2 + \|e\|\|\Delta\|$ and $\|r\| \lesssim \|e\| + |\Theta|^2 + \|e_1\|$. From the second bound in (8.6.12), we conclude $\|P\| + \|Q\| \lesssim 1$ and, thus, $\|r\| \lesssim \|\Delta\|$. By choosing $\varepsilon \sim 1$ small enough, assuming $\|\Delta\| \leq \varepsilon$ and using $\|r\| \lesssim \|\Delta\|$, we obtain

$$\|r\| \lesssim |\Theta|^2 + \|e\|, \qquad \|e_1\| \lesssim |\Theta|^3 + \|e\||\Theta| + \|e\|^2. \tag{8.6.21}$$

This proves (8.6.16). Defining $e_2 := e_1 + B^{-1}QT[e]$ yields $\Delta = \Theta b + \Theta^2 B^{-1}QA[b,b] + e_2$. By plugging this into (8.6.19) and computing the scalar product with $\langle l\,, \cdot\,\rangle$, we obtain

$$\begin{aligned} \Theta\beta\langle l\,, b\rangle \;&= \langle l\,, T[e]\rangle + \Theta^3\langle l\,, A[b, B^{-1}QA[b,b]] + A[B^{-1}QA[b,b],b]\rangle \\ &\quad + \Theta^2\langle l\,, A[b,b]\rangle - \tilde{e}, \end{aligned} \tag{8.6.22a}$$

$$\begin{aligned} \tilde{e} :=& -\langle l, K[e,\Delta] + \Theta^4 A[B^{-1}QA[b,b], B^{-1}QA[b,b]] \\ &+ A[\Delta, e_2] + A[e_2,\Delta] - A[e_2,e_2]\rangle. \end{aligned} \tag{8.6.22b}$$

Since $\|e_2\| \lesssim |\Theta|^3 + \|e\|$ and $\|\Delta\| \lesssim |\Theta| + \|e\|$ by (8.6.21) and (8.6.16), we conclude $\tilde{e} = \mathcal{O}(|\Theta|^4 + |\Theta|\|e\| + \|e\|^2)$. Therefore, Θ satisfies (8.6.14) with the coefficients from (8.6.15).

For the rest of the proof, we additionally assume that the relations in (8.6.17) hold true. Taking the imaginary part of (8.6.20) and arguing similarly as after (8.6.20) yield

$$\|\mathrm{Im}\, e_1\| \lesssim (\|r\| + |\Theta| + \|e\|)(\mathrm{Im}\,\Theta + \|\mathrm{Im}\, r\|), \qquad \|\mathrm{Im}\, r\| \lesssim |\Theta|\mathrm{Im}\,\Theta + \|\mathrm{Im}\, e_1\|.$$

Hence, (8.6.18a) and $\|\mathrm{Im}\, e_1\| \lesssim (|\Theta| + \|e\|)\mathrm{Im}\,\Theta$ follow for $\|\Delta\| \leq \varepsilon$ and $\|e\| \leq \delta$ with some sufficiently small $\varepsilon \sim 1$ and $\delta \sim 1$. From this and taking the imaginary part in (8.6.22b), we conclude (8.6.18b) as $\|\mathrm{Im}\,\Delta\| \lesssim \mathrm{Im}\,\Theta$ by (8.6.18a) and $\mathrm{Im}\, e_2 = \mathrm{Im}\, e_1$. This completes the proof of Lemma 8.6.2. □

8.6.2. Cubic equation associated to Dyson stability equation. Owing to (8.6.15), the coefficients μ_3, μ_2 and μ_1 are completely determined by the bilinear map A and the operator B. For analyzing the Dyson equation, (8.2.3), owing to (8.6.9), the natural choices for A and B are

$$B := \mathrm{Id} - C_m S, \qquad A[x, y] := \frac{1}{2}(mS[x]y + yS[x]m) \tag{8.6.23}$$

with $x, y \in \mathcal{A}$. In particular, Q in (8.6.11) has to be understood with respect to $B = \mathrm{Id} - C_m S$. In the next lemma, we compute μ_3, μ_2 and μ_1 with these choices. This computation involves the inverse of $\mathrm{Id} - C_s F$.

In order to directly ensure its invertibility, we will assume $\mathrm{Im}\, z > 0$. This assumption will be removed in the proof of Proposition 8.6.1 in Section 8.6.3 below.

Lemma 8.6.3 (Coefficients of the cubic for Dyson equation)**.** *Let A and B be defined as in* (8.6.23)*. If Assumptions 8.4.5 hold true on an interval $I \subset \mathbb{R}$ for some $\eta_* \in (0, 1]$ then there is a threshold $\rho_* \sim 1$ such that, for $z \in \mathbb{H}_{I,\eta_*}$ satisfying $\rho(z) + \rho(z)^{-1}\mathrm{Im}\, z \leq \rho_*$, the* *coefficients of the cubic* (8.6.14) *have the expansions*

$$\mu_3 = \psi + \mathcal{O}(\rho + \rho^{-1}\mathrm{Im}\, z), \tag{8.6.24a}$$

$$\mu_2 = \sigma + \mathrm{i}\rho\Big(3\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^2 + \rho^{-1}\mathrm{Im}\, z), \tag{8.6.24b}$$

$$\mu_1 = -\pi\rho^{-1}\mathrm{Im}\, z + 2\mathrm{i}\rho\sigma - 2\rho^2\Big(\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^3 + \mathrm{Im}\, z + \rho^{-2}(\mathrm{Im}\, z)^2). \tag{8.6.24c}$$

Moreover, we also have

$$\langle l, mS[b]b\rangle = \sigma + \mathrm{i}\rho\Big(3\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^2 + \rho^{-1}\mathrm{Im}\, z). \tag{8.6.25}$$

PROOF. In this proof, we use the convention that concatenation of maps on $\mathcal{A}$ and evaluation of these maps in elements of $\mathcal{A}$ are prioritized before the multiplication in $\mathcal{A}$, i.e.,

$$AB[b]c := (A[B[b]])c$$

if A and B are maps on $\mathcal{A}$ and $b, c \in \mathcal{A}$. We will obtain all expansions in (8.6.24) from (8.6.15) by using the special choices for A and B from (8.6.23). Before starting with the proof of (8.6.24a), we establish a few identities. Recalling $m = q^*uq$ from (8.3.2) and (8.3.4), we first notice the following alternative expression for A

$$A[x,y] = \frac{1}{2}C_{q^*,q}\Big[uFC_{q^*,q}^{-1}[x]C_{q^*,q}^{-1}[y] + C_{q^*,q}^{-1}[y]FC_{q^*,q}^{-1}[x]u\Big] \tag{8.6.26}$$

with $x, y \in \mathcal{A}$. Owing to (8.4.21), the operators $C_{q^*,q}$ and $C_{q^*,q}^{-1}$ are bounded. We choose $\rho_* \sim 1$ small enough so that Lemma 8.5.1 is applicable. By using $u = s + \mathrm{i}\mathrm{Im}\, u + \mathcal{O}(\rho^2)$ due to (8.5.2) as well as (8.5.4), (8.5.5) and (8.5.13a) in (8.6.26), we obtain

$$A[b_0, b_0] = C_{q^*,q}[sf_u^2 + \mathrm{i}\rho f_u^3] + \mathcal{O}(\rho^2 + \rho^{-1}\mathrm{Im}\, z). \tag{8.6.27}$$

Combining (8.6.27) and (8.5.18) implies

$$B_0^{-1}Q_0A[b_0, b_0] = C_{q^*,q}(\mathrm{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] + \mathcal{O}(\rho + \rho^{-1}\mathrm{Im}\, z).$$

We now prove the expansion (8.6.24a) for μ_3 by starting from (8.6.15) and using $l = l_0 + \mathcal{O}(\rho)$, $b = b_0 + \mathcal{O}(\rho)$ by (8.5.14), $B^{-1}Q = B_0^{-1}Q_0 + \mathcal{O}(\rho)$ due to $B = B_0 + \mathcal{O}(\rho)$ and

Lemma 8.5.1 and the previous identities. This yields

$$
\begin{aligned}
\mu_3 &= \langle l_0\,, A[B_0^{-1}Q_0A[b_0,b_0],b_0] + A[b_0,B_0^{-1}Q_0A[b_0,b_0]]\rangle + \mathcal{O}(\rho)\\
&= \langle f_u\,, uF(\mathrm{Id}-C_sF)^{-1}Q_{s,F}[sf_u^2]f_u + uF[f_u](\mathrm{Id}-C_sF)^{-1}Q_{s,F}[sf_u^2]\rangle + \mathcal{O}(\rho+\rho^{-1}\mathrm{Im}\,z)\\
&= \langle sf_u^2\,, (\mathrm{Id}+F)(\mathrm{Id}-C_sF)^{-1}Q_{s,F}[sf_u^2]\rangle + \mathcal{O}(\rho+\rho^{-1}\mathrm{Im}\,z).
\end{aligned}
$$

Here, we also used $F[f_u] = f_u + \mathcal{O}(\rho^{-1}\mathrm{Im}\,z)$ by (8.5.5) and $u = s+\mathcal{O}(\rho)$ by (8.5.2). This shows (8.6.24a).

In order to compute μ_2, we define

$$
b_1 := 2\mathrm{i}\rho C_{q^*,q}(\mathrm{Id}-C_sF)^{-1}Q_{s,F}[sf_u^2], \qquad l_1 := -2\mathrm{i}\rho C_{q,q^*}^{-1}(\mathrm{Id}-FC_s)^{-1}Q_{s,F}^*F[sf_u^2].
$$

Then we use (8.5.14a) as well as (8.5.14b) and obtain

$$
\begin{aligned}
\langle l\,, A[b,b]\rangle &= \langle l_0\,, A[b_0,b_0]\rangle + \langle l_1\,, A[b_0,b_0]\rangle + \langle l_0\,, A[b_1,b_0]\rangle + \langle l_0\,, A[b_0,b_1]\rangle + \mathcal{O}(\rho^2+\mathrm{Im}\,z)\\
&= \langle sf_u^3\rangle + \mathrm{i}\rho\langle f_u^4\rangle + 2\mathrm{i}\rho\langle sf_u^2\,, (\mathrm{Id}+2F)(\mathrm{Id}-C_sF)^{-1}Q_{s,F}[sf_u^2]\rangle + \mathcal{O}(\rho^2+\rho^{-1}\mathrm{Im}\,z)\\
&= \sigma + \mathrm{i}\rho\Big(3\psi + \frac{\sigma^2}{\langle f_u^2\rangle}\Big) + \mathcal{O}(\rho^2+\rho^{-1}\mathrm{Im}\,z).
\end{aligned}
$$

Here, in the second step, we used (8.5.13a), (8.6.27) and the definition of l_1 to compute the first and second term, (8.5.13a), the definition of b_1 and (8.6.26) to compute the third and fourth term. In the last step, we then employed

$$
\begin{aligned}
&\langle f_u^4\rangle + \langle sf_u^2\,, 2(\mathrm{Id}+2F)(\mathrm{Id}-C_sF)^{-1}Q_{s,F}[sf_u^2]\rangle\\
&\qquad = \langle sf_u^2\,, (\mathrm{Id}+2(\mathrm{Id}+2F)(\mathrm{Id}-C_sF)^{-1})Q_{s,F}[sf_u^2]\rangle + \langle sf_u^2\,, P_{s,F}[sf_u^2]\rangle\\
&\qquad = 3\langle sf_u^2\,, (\mathrm{Id}+F)(\mathrm{Id}-C_sF)^{-1}Q_{s,F}[sf_u^2]\rangle + \frac{\sigma^2}{\langle f_u^2\rangle} + \mathcal{O}(\rho^{-1}\mathrm{Im}\,z).
\end{aligned}
$$

Here, we applied (8.5.17), $C_s = C_s^*$ and $C_s[sf_u^2] = sf_u^2$. Since $\mu_2 = \langle l\,, A[b,b]\rangle$ by (8.6.15), this completes the proof of (8.6.24b). A similar computation as the one for μ_2 yields (8.6.25).

Since $\mu_1 = -\beta\langle l\,, b\rangle$ by (8.6.15), the expansion in (8.5.14c) immediately yields (8.6.24c). This completes the proof of the lemma. $\square$

8.6.3. The cubic equation for the shape analysis. In this subsection, we will prove Proposition 8.6.1 by using Lemma 8.6.2 and Lemma 8.6.3. Therefore, in addition to the choices of A and B in (8.6.23), we choose $\Delta = m(\tau_0 + \omega) - m(\tau_0)$, $\tau_0, \tau_0 + \omega \in I$, $e = \omega \mathbb{1}$ and

$$T[x] = xm^2, \quad K[x, y] = \frac{1}{2}(xmy + ymx) \tag{8.6.28}$$

for $x, y \in \mathcal{A}$ with $m = m(\tau_0)$ in (8.6.10).

Proof of Proposition 8.6.1. We choose $\rho_* \sim 1$ such that Lemma 8.5.1 and Corollary 8.5.2 are applicable. We fix $\tau_0 \in \mathbb{D}_{\rho_*,\theta}$ and set $m = m(\tau_0)$. The statements about l and b in (a) of Proposition 8.6.1 follow from Corollary 8.5.2. In particular, $|\langle l, b\rangle| \sim 1$. Thus, the conditions in (8.6.12) are a direct consequence of Assumptions 8.4.5, (8.4.21), Lemma 8.5.1 and Corollary 8.5.2. Furthermore, if $\rho = 0$ then we have $m = m^*$ and, thus, (8.6.17) follows. For $\omega \in [-\delta_*, \delta_*]$, $\delta_* := \theta/2$, we set $\Delta = m(\tau_0 + \omega) - m$. Since $\Theta(\omega)b = P[\Delta]$, $r(\omega) = Q[\Delta]$ and $P + Q = \mathrm{Id}$, we immediately obtain (8.6.1). This proves (a).

Next, we derive (8.6.9) for $\Delta := m(z_0 + \omega) - m(z_0)$ and $m := m(z_0)$ with $z_0 := \tau_0 + \mathrm{i}\eta$, $\tau_0 \in \mathbb{D}_{\rho_*,\theta}$, $\omega \in [-\delta_*, \delta_*]$ and $\eta \in (0, \eta_*]$. We subtract (8.2.3) evaluated at $z = z_0$ from (8.2.3) evaluated at $z = z_0 + \omega$ and obtain (8.6.9) with Δ and m defined at $z_0 = \tau_0 + \mathrm{i}\eta$. Directly taking the limit $\eta \downarrow 0$ yields (8.6.9) with the original choices of Δ and m at $z_0 = \tau_0$ by the Hölder-continuity of m on $\overline{\mathbb{H}}_{I',\eta_*}$, $I' := \{\tau \in I \colon \operatorname{dist}(\tau, \partial I) \geq \theta/2\}$, due to Proposition 8.4.7.

Lemma 8.6.2 is applicable for $|\omega| \leq \delta_*$ with some sufficiently small $\delta_* \sim 1$ since this guarantees $\|\Delta\| \leq \varepsilon$ owing to the Hölder-continuity of m. Hence, Lemma 8.6.2 yields a cubic equation for Θ as defined in (8.6.2) with $l = l(z_0)$, $b = b(z_0)$ and $z_0 = \tau_0 + \mathrm{i}\eta$. The coefficients of this cubic equation are given in Lemma 8.6.2. Owing to the uniform 1/3-Hölder continuity of $z \mapsto m(z)$ on $\overline{\mathbb{H}}_{I',\eta_*}$, we conclude from the definition of Θ and $r := Q[\Delta]$ in (8.6.2), the boundedness of Q and $B^{-1}Q$ as well as (8.6.16) that $|\Theta(\omega)| \lesssim |\omega|^{1/3}$, i.e., the second bound in (8.6.7a), and (8.6.7b) uniformly for $\eta \in [0, \eta_*]$.

We now compute the coefficients of the cubic in (8.6.3) for $\tau_0 \in \mathbb{D}_{\rho_*,\theta}$. Set $z_0 := \tau_0 + \mathrm{i}\eta$. Note that for $\eta = \operatorname{Im} z_0 > 0$ these coefficients were already given in (8.6.24), so the only task is to check their limit behaviour as $\eta \downarrow 0$. Owing to (8.5.26), the expansions in

(8.6.4a), (8.6.4b) and (8.6.4c) follow from (8.6.24a), (8.6.24b) and (8.6.24c), respectively, using the continuity of σ, ψ and f_u on $\overline{\mathbb{H}}_{\mathrm{small}}$ by Lemma 8.5.5 and Lemma 8.5.4, respectively. We now show (8.6.4d). With the definitions of $\tilde{e}$ and μ_0 from Lemma 8.6.2 (see (8.6.22b) and (8.6.15), respectively), we set $\Xi(\omega) := \omega^{-1}(\mu_0 - \tilde{e})$ for arbitrary $|\omega| \leq \delta_*$. Since $l = C_{q,q^*}^{-1}[f_u] + \mathcal{O}(\rho + \rho^{-1}\eta)$ due to (8.5.13a) and (8.5.14b), as well as $m^2 = (\operatorname{Re} m)^2 + \mathcal{O}(\rho) = C_{q^*,q} C_s[qq^*] + \mathcal{O}(\rho)$ due to $\operatorname{Im} m \sim \rho \mathbb{1}$ and (8.5.2), we have

$$\omega^{-1}\mu_0 = \langle l^* m^2 \rangle = \langle f_u q q^* \rangle + \mathcal{O}(\rho + \rho^{-1}\eta) = \pi + \mathcal{O}(\rho + \rho^{-1}\eta). \tag{8.6.29}$$

Here, we also used $C_s[f_u] = f_u$ in the second step and (8.5.19) in the last step. We set $\nu(\omega) := -(\omega\pi)^{-1}\tilde{e}$. We recall $e = \omega\mathbb{1}$. Since $\tilde{e} = \mathcal{O}(|\Theta(\omega)|^4 + |\Theta(\omega)||\omega| + |\omega|^2)$ and $|\Theta(\omega)| \lesssim |\omega|^{1/3}$, we obtain (8.6.5). This yields (8.6.4d) by using (8.5.26) in (8.6.29). Since (8.5.35) implies (8.6.6), this completes the proof of (b) for $\tau_0 \in \mathbb{D}_{\rho_*,\theta}$ and we assume $\eta = 0$ in the following.

If $\rho = \rho(\tau_0) > 0$ then (8.4.20) yields the missing first bound in (8.6.7a) completing the proof of part (c). Moreover, in this case, the definitions of Θ and r imply their differentiability at $\omega = 0$ due to Proposition 8.4.7. This shows (d1).

We now verify (d2). Since $\rho = 0$, we have $\operatorname{Im} m(\tau_0) = 0$ and thus $\operatorname{Im}\Theta(\omega) \geq 0$ by the positive semidefiniteness of $\operatorname{Im} m(\tau_0 + \omega)$. Since μ_0 is real as l and $T[e]$ are self-adjoint, we obtain the second bound in (8.6.8) directly from (8.6.18b) and $|\Theta(\omega)| \lesssim |\omega|^{1/3}$. The third bound in (8.6.8) follows from (8.6.18a) and $e = \omega\mathbb{1}$. Since $\rho = 0$ and hence $b = C_{q^*,q}[f_u]$ by (8.5.14a) and $l = C_{q,q^*}^{-1}[f_u]$ by (8.5.14b) are positive definite elements of $\mathcal{A}$, $\operatorname{Re}\Theta(\omega) + \langle l\,, m(\tau_0)\rangle / \langle l\,, b\rangle$ is the real part of the Stieltjes transform of a positive measure μ evaluated on the real axis. The real part of a Stieltjes transform is non-decreasing on the connected components of the complement in $\mathbb{R}$ of the support of its defining measure. Therefore, as the support of μ is contained in $\mathbb{R} \setminus \{\omega \in [-\delta_*, \delta_*] : \operatorname{Im}\Theta(\omega) = 0\}$ due to $\operatorname{Im} m(\tau_0) = 0$, we conclude that $\operatorname{Re}\Theta(\omega)$ is non-decreasing on the connected components of $\{\omega \in [-\delta_*, \delta_*] : \operatorname{Im}\Theta(\omega) = 0\}$.

Lemma 8.5.5 (i) directly implies the Hölder-continuity in (e), which completes the proof of Proposition 8.6.1. □

8.7. Cubic analysis

The main result of this section, Theorem 8.7.1 below, implies Theorem 8.2.5 and gives even effective error terms. Theorem 8.7.1 describes the behaviour of $\operatorname{Im} m$ close to local minima of ρ inside of $\operatorname{supp}\rho$. This behaviour is governed by the universal shape functions $\Psi_{\mathrm{edge}}\colon [0,\infty) \to \mathbb{R}$ and $\Psi_{\min}\colon \mathbb{R} \to \mathbb{R}$ defined by

$$\Psi_{\mathrm{edge}}(\lambda) := \frac{\sqrt{(1+\lambda)\lambda}}{(1+2\lambda+2\sqrt{(1+\lambda)\lambda})^{2/3} + (1+2\lambda-2\sqrt{(1+\lambda)\lambda})^{2/3}+1}, \tag{8.7.1a}$$

$$\Psi_{\min}(\lambda) := \frac{\sqrt{1+\lambda^2}}{(\sqrt{1+\lambda^2}+\lambda)^{2/3} + (\sqrt{1+\lambda^2}-\lambda)^{2/3}-1} - 1. \tag{8.7.1b}$$

For the definition of the comparison relation $\lesssim$, $\gtrsim$ and $\sim$ in the following Theorem 8.7.1, we refer to Convention 8.3.4 and remark that the model parameters in Theorem 8.7.1 are given by c_1, c_2 and c_3 in (8.3.10), k_3 in (8.4.16) and θ in the definition of I_θ in (8.7.2) below.

Theorem 8.7.1 (Behaviour of $\operatorname{Im} m$ close to local minima of ρ)**.** *Let (a, S) be a data pair such that* (8.3.10) *is satisfied. Let m be the solution to the associated Dyson equation* (8.2.3) *and assume that* (8.4.16) *holds true on $\mathbb{H}_{I,\eta_*}$ for some interval $I \subset \mathbb{R}$ and some $\eta_* \in (0,1]$. We write $v := \pi^{-1}\operatorname{Im} m$ and, for some $\theta \in (0,1]$, we set*

$$I_\theta := \{\tau \in I\colon \operatorname{dist}(\tau, \partial I) \geq \theta\}. \tag{8.7.2}$$

Then there are thresholds $\rho_* \sim 1$ and $\delta_* \sim 1$ such that if $\tau_0 \in \operatorname{supp}\rho \cap I_\theta$ is a local minimum of ρ and $\rho(\tau_0) \leq \rho_*$ then

$$v(\tau_0+\omega) = v(\tau_0) + h\Psi(\omega) + \mathcal{O}\Big(\rho(\tau_0)|\omega|^{1/3}\mathbf{1}(|\omega| \lesssim \rho(\tau_0)^3) + \Psi(\omega)^2\Big) \tag{8.7.3}$$

for $\omega \in [-\delta_*, \delta_*] \cap D$ with some $h = h(\tau_0) \in \mathcal{A}$ satisfying $h \sim 1$. Moreover, the set D and the function Ψ depend only on the type of τ_0 in the following way:

- *(a)* Left edge: *If $\tau_0 \in (\partial \operatorname{supp}\rho) \setminus \{\inf \operatorname{supp}\rho\}$ is the infimum of a connected component of $\operatorname{supp}\rho$ and the lower edge of the corresponding gap is in I_θ, i.e., $\tau_1 := \sup((-\infty,\tau_0) \cap \operatorname{supp}\rho) \in I_\theta$, then* (8.7.3) *holds true with $v(\tau_0) = 0$,*

$D = [0, \infty)$ and

$$\Psi(\omega) = \Delta^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{\omega}{\Delta}\Big)$$

where $\Delta := \tau_0 - \tau_1$. If $\tau_0 = \inf \operatorname{supp} \rho$, or more generally $\rho(\tau) = 0$ for all $\tau \in [\tau_0 - \varepsilon, \tau_0]$ with some $\varepsilon \sim 1$, then the same conclusion holds true with $\Delta := 1$.

(b) Right edge: *If $\tau_0 \in \partial \operatorname{supp} \rho$ is the supremum of a connected component then a similar statement as in the case of a left edge holds true.*

(c) Cusp: *If $\tau_0 \notin \partial \operatorname{supp} \rho$ and $\rho(\tau_0) = 0$ then* (8.7.3) *holds true with $D = \mathbb{R}$ and $\Psi(\omega) = |\omega|^{1/3}$.*

(d) Internal minimum: *If $\tau_0 \notin \partial \operatorname{supp} \rho$ and $\rho(\tau_0) > 0$ then there is $\widetilde{\rho} \sim \rho(\tau_0)$ such that* (8.7.3) *holds true with $D = \mathbb{R}$ and*

$$\Psi(\omega) = \widetilde{\rho}\Psi_{\min}\Big(\frac{\omega}{\widetilde{\rho}^3}\Big).$$

If the conditions of Theorem 8.7.1 hold true, i.e., the data pair (a, S) satisfies (8.3.10) and m satisfies (8.4.16) on $\mathbb{H}_{I,\eta_*}$, then Assumptions 8.4.5 are fulfilled on $\mathbb{H}_{I,\eta_*}$ (compare Lemma 8.4.8 (ii)). In fact, Theorem 8.7.1 holds true under Assumptions 8.4.5 which will become apparent from the proof.

Theorem 8.7.1 contains the most important results of the shape analysis. When considering $\rho = \langle v \rangle$ instead of v the coefficient in front of $\Psi(\omega)$ in (8.7.3) can be precisely identified as demonstrated in part (i) of Theorem 8.7.2 below. Moreover, Theorem 8.7.2 contains additional information on the size of the connected components of $\operatorname{supp} \rho$ and the distance between local minima; these are collected in part (ii). Note that the same information were also proven in the commutative setup in Theorem 2.6 of [4] and Theorem 8.7.2 shows that they are also available in our general von Neumann algebra setup.

We remark that $\Psi_{\min}(\omega) = \Psi_{\min}(-\omega)$ for $\omega \in \mathbb{R}$ and, for $\omega > 0$, $\Delta > 0$ and $\widetilde{\rho} > 0$, we have

$$\Delta^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{\omega}{\Delta}\Big) \sim \min\Big\{\frac{\omega^{1/2}}{\Delta^{1/6}}, \omega^{1/3}\Big\}, \tag{8.7.4a}$$

$$\widetilde{\rho}\Psi_{\min}\Big(\frac{\omega}{\widetilde{\rho}^3}\Big) \sim \min\Big\{\frac{\omega^2}{\widetilde{\rho}^5}, \omega^{1/3}\Big\}. \tag{8.7.4b}$$

The comparison relations $\sim$, $\lesssim$ and $\gtrsim$ in the following Theorem 8.7.2 are understood with respect to the constants $k_1, \dots, k_8$ from Assumptions 8.4.5 and θ in the definition of I_θ in (8.7.2).

Theorem 8.7.2 (Behaviour of ρ close to its local minima; Structure of the set of minima of ρ)**.** *Let $I \subset \mathbb{R}$ be an open interval and $\theta \in (0, 1]$. If Assumptions 8.4.5 hold true on I for some $\eta_* \in (0, 1]$ (in particular, if the data pair (a, S) satisfies (8.3.10) and m satisfies (8.4.16) on $\mathbb{H}_{I,\eta_*}$) then the following statements hold true*

(i) There are thresholds $\rho_* \sim 1$, $\sigma_* \sim 1$ and $\delta_* \sim 1$ such that if $\tau_0 \in \operatorname{supp}\rho \cap I_\theta$ is a local minimum of ρ satisfying $\rho(\tau_0) \le \rho_*$ then we set $\Gamma := \sqrt{27}\pi/(2\psi)$ with $\psi = \psi(\tau_0)$ defined as in Lemma 8.5.5 and have

(a) (Left edge) If $\tau_0 \in \partial \operatorname{supp}\rho \setminus \{\inf \operatorname{supp}\rho\}$ is the infimum of a connected component of $\operatorname{supp}\rho$, $|\sigma(\tau_0)| \le \sigma_*$ and the lower edge of the gap lies in I_θ, i.e., $\tau_1 := \sup((-\infty, \tau_0) \cap \operatorname{supp}\rho) \in I_\theta$, then

$$\begin{aligned} \rho(\tau_0 + \omega) &= (4\Gamma)^{1/3}\Psi(\omega) + \mathcal{O}\left(|\sigma(\tau_0)|\Psi(\omega) + \Psi(\omega)^2\right), \\ \Psi(\omega) &:= \Delta^{1/3}\Psi_{\mathrm{edge}}\left(\frac{\omega}{\Delta}\right) \end{aligned} \tag{8.7.5a}$$

for all $\omega \in [0, \delta_*]$. Here, $\Gamma \sim 1$ and $\psi \sim 1$.

(b) (Right edge) If $\tau_0 \in \partial \operatorname{supp}\rho$ is the supremum of a connected component then a similar statement as in the case of a left edge holds true.

(c) (Cusp) If $\tau_0 \notin \partial \operatorname{supp}\rho$ and $\rho(\tau_0) = 0$ then

$$\rho(\tau_0 + \omega) = \frac{\Gamma^{1/3}}{4^{1/3}}|\omega|^{1/3} + \mathcal{O}\left(|\omega|^{2/3}\right) \tag{8.7.5b}$$

for all $\omega \in [-\delta_*, \delta_*]$. Here, $\Gamma \sim 1$ and $\psi \sim 1$.

(d) (Nonzero local minimum) There is $\varepsilon \sim 1$ *such that if* $\tau_0 \notin \partial \operatorname{supp} \rho$ *and* $\rho(\tau_0) > 0$ *then*

$$\rho(\tau_0 + \omega) = \rho(\tau_0) + \begin{cases} \Gamma^{1/3}\Psi(\omega)\Big(1 + \mathcal{O}\Big(\rho(\tau_0)^{1/2} + \frac{|\omega|}{\rho(\tau_0)^3}\Big)\Big), & \text{if } |\omega| \le \varepsilon\rho(\tau_0)^3, \\ \Gamma^{1/3}\Psi(\omega)\Big(1 + \mathcal{O}(\Psi(\omega))\Big), & \text{if } \varepsilon\rho(\tau_0)^3 < |\omega| \le \delta_*, \end{cases}$$
$$\Psi(\omega) := \widetilde{\rho}\Psi_{\min}\left(\frac{\omega}{\widetilde{\rho}^3}\right), \qquad \widetilde{\rho} := \frac{\rho(\tau_0)}{\Gamma^{1/3}} \tag{8.7.5c}$$

for all $\omega \in \mathbb{R}$*. Here,* $\Gamma \sim 1$ *and* $\psi \sim 1$*.*

(ii) If $\operatorname{supp} \rho \cap I_\theta \neq \varnothing$ *then* $\operatorname{supp} \rho \cap I_\theta$ *consists of* $K \sim 1$ *intervals, i.e., there are* $\alpha_1, \dots, \alpha_K \in \partial \operatorname{supp} \rho \cup \partial I_\theta$ *and* $\beta_1, \dots, \beta_K \in \partial \operatorname{supp} \rho \cup \partial I_\theta$*,* $\alpha_i < \beta_i < \alpha_{i+1}$*, such that*

$$\operatorname{supp} \rho \cap \overline{I_\theta} = \bigcup_{i=1}^{K} [\alpha_i, \beta_i] \tag{8.7.6}$$

and $\beta_i - \alpha_i \sim 1$ *if* $\beta_i \neq \sup I_\theta$ *and* $\alpha_i \neq \inf I_\theta$*.*

For $\rho_* > 0$*, we define the set* $\mathbb{M}_{\rho_*}$ *of small local minima* τ *of* ρ *which are not edges of* $\operatorname{supp} \rho$*, i.e.,*

$$\mathbb{M}_{\rho_*} := \{\tau \in (\operatorname{supp} \rho \setminus \partial \operatorname{supp} \rho) \cap I_\theta \colon \rho(\tau) \le \rho_*, \\ \rho \text{ has a local minimum at } \tau\}. \tag{8.7.7}$$

There is a threshold $\rho_* \sim 1$ *such that, for all* $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$ *satisfying* $\gamma_1 \neq \gamma_2$ *and for all* $i = 1, \dots, K$*, we have*

$$|\gamma_1 - \gamma_2| \sim 1, \qquad |\alpha_i - \gamma_1| \sim 1, \qquad |\beta_i - \gamma_1| \sim 1. \tag{8.7.8}$$

The factors $4^{1/3}$ and $4^{-1/3}$ in the cases (a) and (c) of part (i) of Theorem 8.7.2 can be eliminated by redefining Γ, Ψ_{edge} and $\Psi_{\min}$ to bring the leading term on the right-hand sides into the uniform $\Gamma^{1/3}\Psi(\omega)$ form. We have not used these redefined versions of Γ, Ψ_{edge} and $\Psi_{\min}$ here in order to be consistent with [4].

We remark that part (i) (a) and (b) of Theorem 8.7.2 cover only the case of $\tau_0 \in \partial \operatorname{supp} \rho$ with sufficiently small $|\sigma(\tau_0)|$. We will establish later that the smallness of $|\sigma(\tau_0)|$ corresponds to the smallness of the adjacent gap $\tau_0 - \tau_1$ (see Lemma 8.7.14 below). If $|\sigma(\tau_0)|$ is not so small then $\rho(\tau_0 + \omega)$ is well approximated by a rescaled version of

$(\omega_\pm)^{1/2}$ (positive and negative part of ω for left and right edge, respectively). The precise statement and scaling are given in Lemma 8.7.16 below.

8.7.1. Shape regular points. In the following definition, we introduce the notion of a *shape regular point* which collects the properties of m necessary for the proof of Theorem 8.7.1. Proposition 8.7.4 below explains how the statements of Theorem 8.7.1 are transferred to this more general setup. In fact, Lemma 8.4.8 (ii) and Proposition 8.6.1 show that, under the assumptions of Theorem 8.7.1, any point $\tau_0 \in \operatorname{supp} \rho \cap I$ of sufficiently small density $\rho(\tau_0)$ is a shape regular point for m in the sense of Definition 8.7.3 below. By explicitly spelling out the properties of m really used in the proof of Theorem 8.7.1 we made our argument modular because a similar analysis around shape regular points will be applied in later works as well.

This modularity, however, requires to reinterpret the concept of comparison relations. In earlier sections we used the comparison relation $\sim$, $\lesssim$ and the $\mathcal{O}$-notation introduced in Convention 8.3.4 to hide irrelevant constants in various estimates that depended only on the model parameters c_1, c_2, c_3 from (8.3.10), k_3 from (8.4.16) and θ from (8.7.2), these are also the model parameters in Theorem 8.7.1. The model parameters in Theorem 8.7.2 are given by $k_1, \ldots, k_8$ in Assumptions 8.4.5 and θ in the definition of I_θ.

The formulation of Definition 8.7.3 also involves comparison relations instead of carrying constants; in the application these constants depend on the original model parameters. When Proposition 8.7.4 is proven, the corresponding constants directly depend on the constants in Definition 8.7.3, hence they also indirectly depend on the original model parameters when we apply it to the proof of Theorem 8.7.1. Since these dependences are somewhat involved and we do not want to overload the paper with different concepts of comparison relations, for simplicity, for the purpose of Theorem 8.7.1, the reader may think of the implicit constants in every $\sim$-relation depending only on the original model parameters c_1, c_2, c_3, k_3 and θ.

Definition 8.7.3 (Admissibility for shape analysis, shape regular points)**.** Let m be the solution of the Dyson equation (8.2.3) associated to a data pair $(a, S) \in \mathcal{A}_{\mathrm{sa}} \times \Sigma$.

(i) Let $\tau_0 \in \mathbb{R}$, $J \subset \mathbb{R}$ be an open interval with $0 \in J$, $\Theta\colon J \to \mathbb{C}$ and $r\colon J \to \mathcal{A}$ be continuous functions and $b \in \mathcal{A}$. We say that m is (J, Θ, b, r)-*admissible for the shape analysis at* τ_0 if the following conditions are satisfied:

(a) The function $m\colon \mathbb{H} \to \mathcal{A}$ has a continuous extension to $\tau_0 + J$, which we also denote by m. The relation (8.6.1) and the bounds (8.6.7a) as well as (8.6.7b) hold true for all $\omega \in J$.

(b) The function Θ satisfies the cubic equation (8.6.3) for all $\omega \in J$ with the coefficients

$$\begin{aligned} \mu_3 &= \psi + \mathcal{O}(\rho), \\ \mu_2 &= \sigma + \mathrm{i}3\psi\rho + \mathcal{O}(\rho^2 + \rho|\sigma|), \\ \mu_1 &= -2\rho^2\psi + \mathrm{i}\kappa_1\rho\sigma + \mathcal{O}(\rho^3 + \rho^2|\sigma|), \\ \Xi(\omega) &= \kappa(1 + \nu(\omega)) + \mathcal{O}(\rho), \end{aligned}$$

where $\rho := \langle \operatorname{Im} m(\tau_0)\rangle/\pi$ and $\psi, \kappa \geq 0$ as well as $\sigma, \kappa_1 \in \mathbb{R}$ are some parameters satisfying (8.6.6) and $\kappa, |\kappa_1| \sim 1$. The function $\nu\colon J \to \mathbb{C}$ satisfies (8.6.5).

(c) The element $b \in \mathcal{A}$ in (8.6.1) fulfils $b = b^* + \mathcal{O}(\rho)$ and $b + b^* \sim 1$.

(d1) If $\rho > 0$ then Θ and r are differentiable in ω at $\omega = 0$.

(d2) If $\rho = 0$ then (8.6.8) holds true for all $\omega \in J$ and $\operatorname{Re}\Theta$ is non-decreasing on the connected components of $\{\omega \in J\colon \operatorname{Im}\Theta(\omega) = 0\}$.

(ii) Let $\tau_0 \in \mathbb{R}$ and $J \subset \mathbb{R}$ be an open interval with $0 \in J$. We say that τ_0 is a *shape regular point for* m *on* J if m is (J, Θ, b, r)-admissible for the shape analysis at τ_0 for some continuous functions $\Theta\colon J \to \mathbb{C}$ and $r\colon J \to \mathcal{A}$ as well as $b \in \mathcal{A}$.

The key technical step in the proof of Theorem 8.7.1 is the following Proposition 8.7.4; it shows that Theorem 8.7.1 holds under more general weaker conditions, in fact shape admissibility is sufficient. For the proof of Theorem 8.7.1 we will first check shape regularity from Proposition 8.6.1 and then we will prove Proposition 8.7.4; both steps are done in Section 8.7.4 below.

Proposition 8.7.4 (Theorem 8.7.1 under weaker assumptions; Structure of the set of minima in $\mathrm{supp}\,\rho \cap I$)**.** *For the solution m to the Dyson equation* (8.2.3)*, we write $v := \pi^{-1}\mathrm{Im}\, m$, $\rho = \langle v \rangle$.*

Then there are thresholds $\rho_* \sim 1$ and $\delta_* \sim 1$ such that if $\rho(\tau_0) \le \rho_*$ and $\tau_0 \in \mathrm{supp}\,\rho$ is a local minimum of ρ as well as a shape regular point for m on J with an open interval $J \subset \mathbb{R}$ satisfying $0 \in J$ then (8.7.3) *holds true for all $\omega \in [-\delta_*, \delta_*] \cap J \cap D$. Here, as in Theorem 8.7.1, $h = h(\tau_0) \in \mathcal{A}$ with $h \sim 1$ and D as well as Ψ depend only on the type of τ_0 in the following way:*

Suppose that $\tau_0 \in \partial\,\mathrm{supp}\,\rho$ is the infimum of a connected component of $\mathrm{supp}\,\rho$. If $\rho(\tau) = 0$ for all $\tau \in [\tau_0 - \varepsilon, \tau_0]$ with some $\varepsilon \sim 1$ (e.g. $\tau_0 = \inf \mathrm{supp}\,\rho$) and $|\inf J| \gtrsim 1$, then the conclusion of case (a) in Theorem 8.7.1 holds true with $\Delta = 1$ and $v(\tau_0) = 0$.

If $\tau_0 \neq \inf \mathrm{supp}\,\rho$ and $\tau_1 := \sup((-\infty, \tau_0) \cap \mathrm{supp}\,\rho)$ is a shape regular point for m, $\Delta \lesssim 1$ with $\Delta := \tau_0 - \tau_1$ and $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim |\tau_0 - \tau_1|^\zeta$ for some constant $\zeta \in (0, 1/3]$ then the conclusion of case (a) in Theorem 8.7.1 holds true with this choice of Δ as well as $v(\tau_0) = 0$.

Similarly to (a), the statement of case (b) in Theorem 8.7.1 can be translated to the current setup. The cases (c) and (d) of Theorem 8.7.1, cusp and internal minimum, respectively, hold true without any changes.

Furthermore, suppose that $\tau_0 \in \mathrm{supp}\,\rho$ is a shape regular point for m and $\rho(\tau_0) = 0$, then τ_0 is a cusp if $\sigma(\tau_0) = 0$ and τ_0 is an edge, in particular $\tau_0 \in \partial\,\mathrm{supp}\,\rho$, if $\sigma(\tau_0) \neq 0$.

Similarly, the following Proposition 8.7.5 is the analogue of Theorem 8.7.2 under the sole requirement of shape admissibility. Owing to the weaker assumptions, the error term in (8.7.9) as well as the result in (8.7.10) of Proposition 8.7.5 are weaker than the corresponding results in Theorem 8.7.2. We will first show Proposition 8.7.5 and then conclude Theorem 8.7.2 by using extra arguments for the stronger conclusions; both proofs will be presented in Section 8.7.5 below.

At a shape regular point $\tau_0 \in \mathbb{R}$, we set $\Gamma := \sqrt{27}\kappa/(2\psi)$ (cf. Theorem 8.7.6 (i) below), where $\kappa = \kappa(\tau_0)$ and $\psi = \psi(\tau_0)$ are defined as in Definition 8.7.3 (i) (b).

Proposition 8.7.5 (Behaviour of ρ close to minima, set of minima of ρ under weaker assumptions)**.** *Let m be the solution to the Dyson equation,* (8.2.3)*, and $\rho = \pi^{-1}\langle \mathrm{Im}\, m \rangle$.*

(i) Then there are thresholds $\rho_* \sim 1$, $\sigma_* \sim 1$ *and* $\delta_* \sim 1$ *such that if* $\tau_0 \in \operatorname{supp}\rho$ *is a shape regular point for* m *on an open interval* $J \subset \mathbb{R}$ *with* $0 \in J$, $\rho(\tau_0) \le \rho_*$ *and* τ_0 *is a local minimum of* ρ *then we have*

(a) (Left edge) If $\tau_0 \in \partial\operatorname{supp}\rho$ *is the infimum of a connected component of* $\operatorname{supp}\rho$, $|\sigma(\tau_0)| \le \sigma_*$ *and* $\tau_1 := \sup((-\infty,\tau_0)\cap \operatorname{supp}\rho)$ *is a shape regular point satisfying* $\Delta \lesssim 1$ *for* $\Delta := \tau_0 - \tau_1$ *and* $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim |\tau_0 - \tau_1|^\zeta$ *for some constant* $\zeta \in (0, 1/3]$ *then* (8.7.5a) *for all* $\omega \in [0, \delta_*] \cap J$.

(b) (Right edge) If $\tau_0 \in \partial\operatorname{supp}\rho$ *is the supremum of a connected component then a similar statement as in the case of a left edge holds true.*

(c) (Cusp) If $\tau_0 \notin \partial\operatorname{supp}\rho$ *and* $\rho(\tau_0) = 0$ *then* (8.7.5b) *holds true for all* $\omega \in [-\delta_*, \delta_*] \cap J$.

(d) (Internal minimum) If $\tau_0 \notin \partial\operatorname{supp}\rho$ *and* $\rho(\tau_0) > 0$ *then*

$$\begin{aligned} \rho(\tau_0+\omega) &= \rho(\tau_0) + \Gamma^{1/3}\Psi(\omega) + \mathcal{O}\left(\frac{|\omega|}{\rho(\tau_0)}\mathbf{1}(|\omega| \lesssim \rho(\tau_0)^3) + \Psi(\omega)^2\right), \\ \Psi(\omega) &:= \widetilde{\rho}\Psi_{\min}\left(\frac{\omega}{\widetilde{\rho}^3}\right), \qquad \widetilde{\rho} := \frac{\rho(\tau_0)}{\Gamma^{1/3}} \end{aligned} \tag{8.7.9}$$

for all $\omega \in [-\delta_*, \delta_*] \cap J$.

(ii) Let $I \subset \mathbb{R}$ *be an open interval with* $\operatorname{supp}\rho \cap I \neq \varnothing$ *and* $|I| \lesssim 1$ *and let* m *have a continuous extension to the closure* $\overline{I}$ *of* I. *Let* $J \subset \mathbb{R}$ *be an open interval with* $0 \in J$ *and* $\operatorname{dist}(0, \partial J) \gtrsim 1$ *such that* $J + (\partial\operatorname{supp}\rho) \cap I \subset I$. *We assume that all points in* $(\partial\operatorname{supp}\rho)\cap I$ *are shape regular points for* m *on* J *and all estimates in Definition 8.7.3 hold true uniformly on* $(\partial\operatorname{supp}\rho)\cap I$. *If* $|\sigma(\tau_0)-\sigma(\tau_1)| \lesssim |\tau_0-\tau_1|^\zeta$ *for some* $\zeta \in (0, 1/3]$ *and uniformly for all* $\tau_0, \tau_1 \in (\partial\operatorname{supp}\rho)\cap I$ *then* $\operatorname{supp}\rho \cap I$ *consists of* $K \sim 1$ *intervals, i.e., there are* $\alpha_1, \dots, \alpha_K \in \partial\operatorname{supp}\rho \cup \partial I$ *and* $\beta_1, \dots, \beta_K \in \partial\operatorname{supp}\rho\cup\partial I$, $\alpha_i < \beta_i < \alpha_{i+1}$, *such that* (8.7.6) *holds true with* $\overline{I_\theta}$ *replaced by* $\overline{I}$ *and* $\beta_i - \alpha_i \sim 1$ *if* $\beta_i \neq \sup I$ *and* $\alpha_i \neq \inf I$.

If $\mathbb{M}_{\rho_*}$ *is defined as in* (8.7.7) *then there is a threshold* $\rho_* \sim 1$ *such that if, in addition to the previous conditions in (ii), all points of* $(\mathbb{M}_{\rho_*}\cup\partial\operatorname{supp}\rho)\cap I$ *are shape regular points for* m *on* J *and all estimates in Definition 8.7.3 hold true uniformly on* $(\mathbb{M}_{\rho_*}\cup\partial\operatorname{supp}\rho)\cap I$ *then, for* $\gamma \in \mathbb{M}_{\rho_*}$, *we have* $|\alpha_i - \gamma| \sim 1$ *and* $|\beta_i - \gamma| \sim 1$ *if* $\alpha_i \neq \inf I$ *and* $\beta_i \neq \sup I$. *Moreover, for any* $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$, *we* *have either*

$$|\gamma_1 - \gamma_2| \sim 1, \qquad or \qquad |\gamma_1 - \gamma_2| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4. \tag{8.7.10}$$

If $\rho(\gamma_1) = 0$ *or* $\rho(\gamma_2) = 0$ *then, for* $\gamma_1 \neq \gamma_2$*, only the first case occurs.*

An important step towards Theorem 8.7.1 and Proposition 8.7.4 will be to prove similar behaviours for Θ as $\operatorname{Im}\Theta$ is the leading term in v. These behaviours are collected in the following theorem, Theorem 8.7.6. It has weaker assumptions than those of Theorem 8.7.1 and those required in Proposition 8.7.4 – in particular, on the coefficient μ_1 in the cubic equation (8.6.3). However, these assumptions will be sufficient for the purpose of Theorem 8.7.6.

Theorem 8.7.6 (Abstract cubic equation)**.** *Let* $\Theta(\omega)$ *be a continuous solution to the cubic equation*

$$\mu_3\Theta(\omega)^3 + \mu_2\Theta(\omega)^2 + \mu_1\Theta(\omega) + \omega\Xi(\omega) = 0 \tag{8.7.11}$$

for $\omega \in J$*, where* $J \subset \mathbb{R}$ *is an open interval with* $0 \in J$*. We assume that the coefficients satisfy*

$$\begin{aligned} \mu_3 &= \psi + \mathcal{O}(\rho), \\ \mu_2 &= \sigma + 3\mathrm{i}\psi\rho + \mathcal{O}(\rho^2 + \rho|\sigma|), \\ \mu_1 &= -2\rho^2\psi + \mathcal{O}(\rho^3 + \rho|\sigma|), \end{aligned} \tag{8.7.12}$$

$$\Xi(\omega) = \kappa(1 + \nu(\omega)) + \mathcal{O}(\rho)$$

with some fixed parameters $\psi \geq 0$*,* $\rho \geq 0$*,* $\sigma \in \mathbb{R}$ *and* $\kappa \sim 1$*. The cubic equation is assumed to be stable in the sense that*

$$\psi + \sigma^2 \sim 1. \tag{8.7.13}$$

Moreover, for all $\omega \in J$*, we require the following bounds on* ν *and* Θ*:*

$$|\nu(\omega)| \lesssim |\omega|^{1/3}, \tag{8.7.14a}$$

$$|\Theta(\omega)| \lesssim |\omega|^{1/3}. \tag{8.7.14b}$$

Then the following statements hold true:

(i) ($\rho > 0$) For any $\Pi_* \sim 1$, there is a threshold $\rho_* \sim 1$ such that if $\rho \in (0, \rho_*]$ and $|\sigma| \le \Pi_* \rho^2$ then we have

$$\operatorname{Im}\Theta(\omega) = \rho\Psi_{\min}\Big(\Gamma\frac{\omega}{\rho^3}\Big) + \mathcal{O}\Big(\min\{\rho^{-1}|\omega|, |\omega|^{2/3}\}\Big), \tag{8.7.15}$$

with $\Gamma := \sqrt{27}\kappa/(2\psi)$. Note that $\Gamma \sim 1$ if $\rho_* \sim 1$ is small enough.

(ii) ($\rho = 0$) If $\rho = 0$ and we additionally assume $\operatorname{Im}\Theta(\omega) \ge 0$ for $\omega \in J$, $\operatorname{Re}\Theta$ is non-decreasing on the connected components of $\{\omega \in J\colon \operatorname{Im}\Theta(\omega) = 0\}$ as well as

$$|\operatorname{Im}\nu(\omega)| \lesssim \operatorname{Im}\Theta(\omega) \tag{8.7.16}$$

for all $\omega \in J$ then we have

(a) If $\sigma = 0$ then $\operatorname{Im}\Theta(\omega)$ has a cubic cusp at $\omega = 0$, i.e.,

$$\operatorname{Im}\Theta(\omega) = \frac{\sqrt{3}}{2}\Big(\frac{\kappa}{\psi}\Big)^{1/3}|\omega|^{1/3} + \mathcal{O}(|\omega|^{2/3}). \tag{8.7.17}$$

(b) If $\sigma \neq 0$ then $\operatorname{Im}\Theta(\omega)$ has a square root edge at $\omega = 0$, i.e., there is $c_* \sim 1$ such that

$$\operatorname{Im}\Theta(\omega) = c\widehat{\Delta}^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{|\omega|}{\widehat{\Delta}}\Big) + \mathcal{O}\Big((|\nu(\omega)| + \varepsilon(\omega))\varepsilon(\omega)\Big), \tag{8.7.18a}$$

if $\operatorname{sign}\omega = \operatorname{sign}\sigma$, and

$$\operatorname{Im}\Theta(\omega) = 0, \tag{8.7.18b}$$

if $\operatorname{sign}\omega = -\operatorname{sign}\sigma$ and $|\omega| \le c_*|\sigma|^3$, where $\widehat{\Delta} \in (0,\infty)$, $c \in (0,\infty)$ and $\varepsilon\colon \mathbb{R} \to [0,\infty)$ are defined by

$$\begin{aligned} \widehat{\Delta} &:= \min\Big\{\frac{4}{27\kappa}\frac{|\sigma|^3}{\psi^2}, 1\Big\}, \qquad c := 3\sqrt{\kappa}\frac{\widehat{\Delta}^{1/6}}{|\sigma|^{1/2}}, \\ \varepsilon(\omega) &:= \min\Big\{\frac{|\omega|^{1/2}}{\widehat{\Delta}^{1/6}}, |\omega|^{1/3}\Big\}. \end{aligned} \tag{8.7.19}$$

We have $\widehat{\Delta} \sim |\sigma|^3$ and $c \sim 1$. Moreover, for $\operatorname{sign}\omega = \operatorname{sign}\sigma$, we have

$$|\Theta(\omega)| \lesssim \varepsilon(\omega). \tag{8.7.20}$$

8.7.2. Cubic equations in normal form. The core of the proof of Theorem 8.7.6 is to bring (8.7.11) into a normal form by a change of variables. We will first explain the analysis of these normal forms, especially the mechanism of choosing the right branch of the solution based upon selection principles that will be derived from the constraints on Θ given in Theorem 8.7.6. Then, in Section 8.7.3, we show how to bring (8.7.11) to these normal forms.

In the following proposition, we study a special solution $\Omega(\lambda)$ to a one-parameter family of cubic equations in normal forms with constant term $\Lambda(\lambda)$ (or $2\Lambda(\lambda)$), where $\Lambda(\lambda)$ is a perturbation of the identity map $\lambda \mapsto \lambda$. Here, a priori, the real parameter λ is always contained in an (possibly unbounded) interval around 0. This range of definition will not be explicitly indicated in the statements but will be explicitly restricted for their conclusions. We compare the solution to this perturbed cubic equation with the solution to the cubic equation with constant term λ. Depending on the precise type of the cubic equation, the choice of the solution is based on some of the following *selection principles*

- **SP1** $\lambda \mapsto \Omega(\lambda)$ is continuous
- **SP2** $\Omega(0) = \Omega_0$ for some given $\Omega_0 \in \mathbb{C}$
- **SP3** $\operatorname{Im}(\Omega(\lambda) - \Omega(0)) \geq 0$,
- **SP4'** $|\operatorname{Im}\Lambda(\lambda)| \leq \gamma|\lambda|\operatorname{Im}\Omega(\lambda)$ for some $\gamma > 0$ and $\operatorname{Re}\Omega(\lambda)$ is non-decreasing on the connected components of $\{\lambda\colon \operatorname{Im}\Omega(\lambda) = 0\}$.

We use the notation **SP4'** to distinguish this selection principle from **SP-4** which was introduced in Lemma 9.9 of [4].

We will make use of the following standard convention for complex powers.

Definition 8.7.7 (Complex powers)**.** We define $\mathbb{C} \setminus (-\infty, 0) \to \mathbb{C},\ \zeta \mapsto \zeta^\gamma$ for $\gamma \in \mathbb{C}$ by $\zeta^\gamma := \exp(\gamma \log \zeta)$, where $\log\colon \mathbb{C} \setminus (-\infty, 0) \to \mathbb{C}$ is a continuous branch of the complex logarithm with $\log 1 = 0$.

With this convention, we record Cardano's formula as follows:

Proposition 8.7.8 (Cardano). *The three roots of* $\Omega^3 - 3\Omega + 2\zeta$, $\zeta \in \mathbb{C}$, *are* $\widehat{\Omega}_+(\zeta)$, $\widehat{\Omega}_-(\zeta)$ *and* $\widehat{\Omega}_0(\zeta)$ *which are defined by*

$$\begin{aligned} \widehat{\Omega}_\pm(\zeta) &:= \frac{1}{2}(\Phi_+(\zeta) + \Phi_-(\zeta)) \pm \frac{\mathrm{i}\sqrt{3}}{2}(\Phi_+(\zeta) - \Phi_-(\zeta)), \\ \widehat{\Omega}_0(\zeta) &:= -(\Phi_+(\zeta) + \Phi_-(\zeta)), \end{aligned} \tag{8.7.21}$$

where

$$\Phi_\pm(\zeta) = \begin{cases} (\zeta \pm \sqrt{\zeta^2 - 1})^{1/3}, & \text{if } \operatorname{Re}\zeta \geq 1, \\ (\zeta \pm \mathrm{i}\sqrt{1 - \zeta^2})^{1/3}, & \text{if } |\operatorname{Re}\zeta| < 1, \\ -(-\zeta \mp \sqrt{\zeta^2 - 1})^{1/3}, & \text{if } \operatorname{Re}\zeta \leq -1. \end{cases}$$

Proposition 8.7.9 (Solution to the cubic in normal form). *Let* $\Omega(\lambda)$ *satisfy* ***SP1*** *and* ***SP2***.

(i) (Non-zero local minimum) Let $\Omega_0 = \sqrt{3}(\mathrm{i} + \chi_1)$ *in* ***SP2*** *and* $\Omega(\lambda)$ *satisfy*

$$\Omega(\lambda)^3 + 3\Omega(\lambda) + 2\Lambda(\lambda) = 0, \qquad \Lambda(\lambda) = (1 + \chi_2 + \mu(\lambda))\lambda + \chi_3, \tag{8.7.22}$$

with $|\mu(\lambda)| \lesssim \alpha|\lambda|^{1/3}$, $\alpha > 0$. *Then there exist* $\delta \sim 1$ *and* $\chi_* \sim 1$ *such that if* $\alpha, |\chi_1|, |\chi_2|, |\chi_3| \leq \chi_*$ *then*

$$\Omega(\lambda) - \Omega_0 = \widehat{\Omega}(\lambda) - \mathrm{i}\sqrt{3} + \mathcal{O}\Big((\alpha + |\chi_2| + |\chi_3|)\min\{|\lambda|, |\lambda|^{2/3}\}\Big) \tag{8.7.23}$$

for all $\lambda \in \mathbb{R}$ *satisfying* $|\lambda| \leq \delta/\alpha^3$, *where* $\widehat{\Omega}(\lambda) := \Phi_{\mathrm{odd}}(\lambda) + \mathrm{i}\sqrt{3}\Phi_{\mathrm{even}}(\lambda)$ *and* Φ_{odd} *and* Φ_{even} *are the odd and even part of the function* $\Phi\colon \mathbb{C} \to \mathbb{C}$, $\Phi(\zeta) := (\sqrt{1 + \zeta^2} + \zeta)^{1/3}$, *respectively.*

Moreover, we have for $|\lambda| \leq \delta/\alpha^3$ *that*

$$|\Omega(\lambda) - \Omega_0| \lesssim \min\{|\lambda|, |\lambda|^{1/3}\}. \tag{8.7.24}$$

In the following, we assume that $\Omega(\lambda)$, *in addition to* ***SP1*** *and* ***SP2***, *also satisfies* ***SP3*** *and* ***SP4'***.

(ii) (Simple edge) Let $\Omega_0 = 0$ *in* ***SP2*** *and* $\Omega(\lambda)$ *be a solution to*

$$\Omega^2(\lambda) + \Lambda(\lambda) = 0, \qquad \Lambda(\lambda) = (1 + \mu(\lambda))\lambda. \tag{8.7.25}$$

If $|\mu(\lambda)| \leq \gamma^{2/3}|\lambda|^{1/3}$ *for the* $\gamma > 0$ *of* ***SP4'*** *then there is* $c_* \sim 1$ *such that*

$$\Omega(\lambda) = \widehat{\Omega}(\lambda) + \mathcal{O}\Big(|\mu(\lambda)||\lambda|^{1/2}\Big),$$
$$\widehat{\Omega}(\lambda) := \begin{cases} \mathrm{i}\lambda^{1/2}, & \text{if } \lambda \in [0, c_*\gamma^{-2}], \\ -(-\lambda)^{1/2}, & \text{if } \lambda \in [-c_*\gamma^{-2}, 0]. \end{cases} \tag{8.7.26}$$

Moreover, we have $\operatorname{Im}\Omega(\lambda) = 0$ *for* $\lambda \in [-c_*\gamma^{-2}, 0]$.

(iii) (Sharp cusp) Let $\Omega_0 = 0$ *in* ***SP2****,* $\gamma \sim 1$ *in* ***SP4'*** *and* $\Omega(\lambda)$ *be a solution to*

$$\Omega^3(\lambda) + \Lambda(\lambda) = 0, \qquad \Lambda(\lambda) = (1 + \mu(\lambda))\lambda. \tag{8.7.27}$$

If $|\mu(\lambda)| \lesssim |\lambda|^{1/3}$ *then there is* $\delta \sim 1$ *such that*

$$\Omega(\lambda) = \widehat{\Omega}(\lambda) + \mathcal{O}\Big(|\mu(\lambda)||\lambda|^{1/3}\Big),$$
$$\widehat{\Omega}(\lambda) := \frac{1}{2}\begin{cases} (-1 + \mathrm{i}\sqrt{3})\lambda^{1/3}, & \text{if } \lambda \in (0, \delta], \\ (1 + \mathrm{i}\sqrt{3})|\lambda|^{1/3}, & \text{if } \lambda \in [-\delta, 0]. \end{cases} \tag{8.7.28}$$

(iv) (Two nearby edges) Let $\Omega_0 = s$ *for some* $s \in \{\pm 1\}$ *in* ***SP2****,* $\gamma \sim 1$ *in* ***SP4'*** *and* $\Omega(\lambda)$ *be a solution to*

$$\Omega(\lambda)^3 - 3\Omega(\lambda) + 2\Lambda(\lambda) = 0, \qquad \Lambda(\lambda) = (1 + \mu(\lambda))\lambda + s. \tag{8.7.29}$$

Then there are $\delta \sim 1$*,* $\varrho \sim 1$ *and* $\gamma_* \sim 1$ *such that if* $|\mu(\lambda)| \lesssim \widehat{\gamma}|\lambda|^{1/3}$ *for some* $\widehat{\gamma} \in [0, \gamma_*]$ *then*

(a) We have

$$\Omega(\lambda) = \widehat{\Omega}_+(1 + |\lambda|) + \mathcal{O}\Big(|\mu(\lambda)| \min\{|\lambda|^{1/2}, |\lambda|^{1/3}\}\Big), \tag{8.7.30}$$

for all $\lambda \in s(0, 2\delta/\widehat{\gamma}^3]$*. (Recall the definition of* $\widehat{\Omega}_+$ *from* (8.7.21)*.) Moreover, for all* $\lambda \in s(0, 2\delta/\widehat{\gamma}^3]$*, we have*

$$|\Omega(\lambda) - \Omega_0| \lesssim \min\Big\{|\lambda|^{1/2}, |\lambda|^{1/3}\Big\}. \tag{8.7.31}$$

(b) For all $\lambda \in -s(0, 2 - \varrho\widehat{\gamma}]$*, we have*

$$\operatorname{Im}\Omega(\lambda) \lesssim \widehat{\gamma}^{1/2}. \tag{8.7.32}$$

(c) We have

$$\operatorname{Im}\Omega(-s(2+\varrho\widehat{\gamma}))>0. \tag{8.7.33}$$

The core of each part in Proposition 8.7.9 is choosing the correct cubic root. For the most complicated part (iv), we state this choice in the following auxiliary lemma. For its formulation, we introduce the intervals

$$I_1:=-s[-\lambda_1,0),\qquad I_2:=-s(0,\lambda_2],\qquad I_3:=-s[\lambda_3,\lambda_1], \tag{8.7.34}$$

where we used the definitions

$$\lambda_1:=2\frac{\delta}{\widehat{\gamma}^3},\qquad \lambda_2:=2-\varrho\widehat{\gamma},\qquad \lambda_3:=2+\varrho\widehat{\gamma}. \tag{8.7.35}$$

These definitions are modelled after (9.105) in [4]. We will choose $\widehat{\gamma}=\widehat{\Delta}^{1/3}$ in the proof of Theorem 8.7.6 below. Then λ_1 corresponds to an expansion range δ in the ω coordinate. Note that with the above choice of $\widehat{\gamma}$, we obtain the same λ_1 as in (9.105) of [4]. However, λ_2 and λ_3 differ slightly from those in [4], where $\lambda_{2,3}$ were set to be $2\mp\varrho|\sigma|$. Nevertheless, we will see below that $\widehat{\gamma}\sim|\sigma|$ but they are not equal in general.

For given $\delta,\varrho\sim 1$, we will always choose $\gamma_*\sim 1$ so small that $\widehat{\gamma}\le\gamma_*$ implies

$$\lambda_1\ge 4,\qquad 1\le\lambda_2<2<\lambda_3\le 3.$$

Therefore, the intervals in (8.7.34) are disjoint and nonempty.

Lemma 8.7.10 (Choice of cubic roots in Proposition 8.7.9 (iv))**.** *Under the assumptions of Proposition 8.7.9 (iv), there are $\delta,\varrho,\gamma_*\sim 1$ such that if $\widehat{\gamma}\le\gamma_*$ then we have*

$$\Omega|_{I_k}=\widehat{\Omega}_+\circ\Lambda|_{I_k}$$

for $k=1,2,3$. Here, $\widehat{\Omega}_+$ is defined as in (8.7.21)*.*

Proof. The proof is the same as the one of Lemma 9.14 in [4] but **SP-4** in [4] is replaced by **SP4'** above. In that proof, **SP-4** is used only in the part titled "Choice of a_2". We redo this part here. Recall that $a_2=0,\pm$ denoted the index such that $\Omega|_{I_2}=\widehat{\Omega}_{a_2}\circ\Lambda|_{I_2}$ and our goal is to show $a_2=+$. Similarly as in [4], we assume without loss of generality $s=-1$. Since $\lim_{\lambda\downarrow -1}\widehat{\Omega}_-(\lambda)=2$ and $\Omega(0)=-1$ by **SP2**, we conclude

$a_2 \neq -$. (In the corresponding step in [4], there was a typo: $\widehat{\Omega}_+(-1+0) = 2$ should have been $\widehat{\Omega}_-(-1+0) = 2$, resulting in the choice $a_2 = +$. This conclusion is only used in the bound (9.137) of [4] which still holds true. The rest of the proof is unaffected.)

We now prove $a_2 \neq 0$. To that end, we take the imaginary part of the cubic equation, (8.7.29), and obtain

$$3((\operatorname{Re}\Omega)^2 - 1)\operatorname{Im}\Omega = -2\lambda \operatorname{Im}\mu(\lambda) + (\operatorname{Im}\Omega)^3. \tag{8.7.36}$$

Suppose that $a_2 = 0$. From the definition of $\widehat{\Omega}_0$, $\Lambda(\lambda) = (1+\mu(\lambda))\lambda - 1$ and $|\mu(\lambda)| \lesssim \widehat{\gamma}|\lambda|^{1/3}$ we obtain

$$\operatorname{Re}\widehat{\Omega}_0(\Lambda(\lambda)) \leq -1 - c|\lambda|^{1/2} + C\widehat{\gamma}^{1/2}\lambda^{2/3}, \qquad |\operatorname{Im}\widehat{\Omega}_0(\Lambda(\lambda))| \lesssim \widehat{\gamma}^{1/2}\lambda^{2/3}, \tag{8.7.37}$$

(compare (9.120) in [4]). Thus, from (8.7.36), we conclude

$$|\lambda|^{1/2}\operatorname{Im}\Omega \lesssim |\lambda|\operatorname{Im}\Omega$$

for small λ as $|\operatorname{Im}\mu(\lambda)| \lesssim \operatorname{Im}\Omega$ by **SP4'** and $|\operatorname{Im}\Lambda| = |\lambda||\operatorname{Im}\mu|$. Hence, $\operatorname{Im}\Omega(\lambda) = 0$ for small enough $|\lambda|$. Thus, $\operatorname{Re}\Omega$ is non-decreasing for such λ by **SP4'**, but from $\Omega(0) = -1$ and the first bound in (8.7.37) we conclude that $\operatorname{Re}\Omega$ has to be decreasing if $\Omega(\lambda) = \widehat{\Omega}_0(\Lambda(\lambda))$. This contradiction shows $a_2 \neq 0$, hence, $a_2 = +$. The rest of the proof in [4] is unchanged. □

Proof of Proposition 8.7.9. For the proof of (i), we mainly follow the proof of Proposition 9.3 in [4] with $\gamma_4 = \chi_1$, $\gamma_5 = \chi_2$ and $\gamma_6 = \chi_3$ in (9.35) and (9.37) of [4].

Following the careful selection of the correct solution of (8.7.22) (cf. (9.36) in [4]) by the selection principles till above (9.50) in [4] yields $\Omega(\lambda) = \widehat{\Omega}(\Lambda(\lambda))$ and hence, in particular, $\widehat{\Omega}(\chi_3) = \Omega_0 = \sqrt{3}(\mathrm{i} + \chi_1)$. ($\widehat{\Omega} = \widehat{\Omega}_+$ in [4].) By defining

$$\Lambda_0(\lambda) := (1 + \chi_2 + \mu(\lambda))\lambda$$

and using $|\mu(\lambda)| \lesssim \alpha|\lambda|^{1/3}$ instead of (9.54) in [4], we obtain

$$\begin{aligned}\widehat{\Omega}(\Lambda_0(\lambda)) - \widehat{\Omega}(0) &= \widehat{\Omega}(\lambda) - \widehat{\Omega}(0) + \mathcal{O}\left((|\chi_2| + |\mu(\lambda)|)\frac{|\lambda|}{1+|\lambda|^{2/3}}\right) \\ &= \widehat{\Omega}(\lambda) - \widehat{\Omega}(0) + \mathcal{O}((\alpha + |\chi_2|)\min\{|\lambda|, |\lambda|^{2/3}\})\end{aligned}$$

instead of (9.56) in [4]. Thus, (9.57) in the proof of Proposition 9.3 in [4] yields

$$\widehat{\Omega}(\chi_3 + \Lambda_0(\lambda)) - \widehat{\Omega}(\chi_3) = \widehat{\Omega}(\lambda) - \widehat{\Omega}(0) + \mathcal{O}((\alpha + |\chi_2| + |\chi_3|)\min\{|\lambda|, |\lambda|^{2/3}\}).$$

Thus, we obtain (8.7.23) since $\widehat{\Omega}(\chi_3) = \Omega_0$ and $\widehat{\Omega}(0) = \mathrm{i}\sqrt{3}$. We remark that (8.7.24) is exactly (9.53) in [4].

The proof of (ii) resembles the proof of Lemma 9.11 in [4] but we replace assumption **SP-4** of [4] by **SP4'**. Since $\Omega(\lambda)$ solves (8.7.25), there is a function $A\colon \mathbb{R} \to \{\pm\}$ such that $\Omega(\lambda) = \widetilde{\Omega}_{A(\lambda)}(\Lambda(\lambda))$ for all $\lambda \in \mathbb{R}$. Here, $\widetilde{\Omega}_\pm\colon \mathbb{C} \to \mathbb{C}$ denote the functions

$$\widetilde{\Omega}_\pm(\zeta) := \pm \begin{cases} \mathrm{i}\zeta^{1/2}, & \text{if } \operatorname{Re}\zeta \geq 0, \\ -(-\zeta)^{1/2}, & \text{if } \operatorname{Re}\zeta < 0. \end{cases}$$

(Note that they were denoted by $\widehat{\Omega}_\pm$ in (9.78) of [4]). By assumption, there is $c_* \sim 1$ such that $|\mu(\lambda)| < 1$ for all $|\lambda| \leq c_*\gamma^{-2}$. Hence, by **SP1**, we find $a_+, a_- \in \{\pm\}$ such that $A(\lambda) = a_\pm$ for $\lambda \in \pm[0, c_*\gamma^{-2}]$.

For $\lambda \geq 0$, we have

$$\operatorname{Im}\widetilde{\Omega}_-(\Lambda(\lambda)) = -\lambda^{1/2} + \mathcal{O}(\mu(\lambda)\lambda^{1/2}).$$

Thus, possibly shrinking $c_* \sim 1$, we obtain $\operatorname{Im}\widetilde{\Omega}_-(\Lambda(\lambda)) < 0$ for $\lambda \in (0, c_*\gamma^{-2}]$. Therefore, the choice $a_+ = -$ would contradict **SP3** and we conclude $a_+ = +$.

We now prove that $a_- = +$. Assume to the contrary that $a_- = -$. For small enough $c_* \sim 1$, we have

$$\operatorname{Re}\widetilde{\Omega}_-(\Lambda(\lambda)) = |\lambda|^{1/2}\operatorname{Re}(1 + \mu(\lambda))^{1/2} \sim |\lambda|^{1/2},$$
$$\operatorname{Im}\widetilde{\Omega}_-(\Lambda(\lambda)) = |\lambda|^{1/2}\operatorname{Im}((1 + \mu(\lambda))^{1/2}) \lesssim |\lambda|^{1/2}$$

for $\lambda \in [-c_*\gamma^{-2}, 0)$ by the definition of $\widetilde{\Omega}_-$ and Λ. Hence, taking the imaginary part of (8.7.25) and using **SP4'** yield

$$|\lambda|^{1/2}\operatorname{Im}\Omega(\lambda) \lesssim \gamma|\lambda|\operatorname{Im}\Omega(\lambda)$$

for $\lambda \in [-c_*\gamma^{-2}, 0)$. By possibly shrinking $c_* \sim 1$, we obtain $\operatorname{Im}\Omega(\lambda) = 0$ for $\lambda \in [-c_*\gamma^{-2}, 0)$. Thus, **SP4'** implies that $\operatorname{Re}\Omega$ is non-decreasing on $[-c_*\gamma^{-2}, 0)$ which contradicts $\operatorname{Re}\widetilde{\Omega}_-(0) = 0$ and $\operatorname{Re}\widetilde{\Omega}_-(\Lambda(\lambda)) \sim |\lambda|^{1/2} > 0$ for $\lambda \in [-c_*\gamma^{-2}, 0)$ with small enough $c_* \sim 1$. Hence, $a_- = +$ which completes the selection of the main term $\widehat{\Omega} = \widetilde{\Omega}_+$ in (8.7.26). The error term in (8.7.26) follows by estimating $\widehat{\Omega}(\Lambda(\lambda))$ directly.

For the proof of (iii), we select the correct root of (8.7.27) as in the proof of Lemma 9.12 in [4] under **SP4'** instead of **SP-4**. Since $\Omega(\lambda)$ solves (8.7.27) there is a function $A\colon \mathbb{R} \to \{0, \pm\}$ such that

$$\Omega(\lambda) = \widetilde{\Omega}_{A(\lambda)}(\Lambda(\lambda))$$

for all $\lambda \in \mathbb{R}$. Here, we introduced the functions $\widetilde{\Omega}_a\colon \mathbb{C} \to \mathbb{C}$, $a = 0, \pm$, defined by

$$\widetilde{\Omega}_0 := -\begin{cases} \zeta^{1/3}, & \text{if } \operatorname{Re}\zeta \geq 0, \\ -(-\zeta)^{1/3}, & \text{if } \operatorname{Re}\zeta < 0, \end{cases} \qquad \widetilde{\Omega}_\pm(\zeta) := \frac{1 \mp \mathrm{i}\sqrt{3}}{2}\widetilde{\Omega}_0(\zeta).$$

(Note that they were denoted by $\widehat{\Omega}_a$, $a \in \{0, \pm\}$, in (9.87) of [4].) By **SP1**, A can only change its value at λ if $\Lambda(\lambda) = 0$. By choosing $\delta \sim 1$ small enough and using $|\mu(\lambda)| \lesssim |\lambda|^{1/3}$, we have $A(\lambda) = a_+$ and $A(-\lambda) = a_-$ for some constants $a_\pm$ and for all $\lambda \in (0, \delta]$.

We will now use **SP3** and **SP4'** to determine the value of a_+ and a_-. As in (9.91) of the proof of Lemma 9.12 in [4], we have

$$\pm(\operatorname{sign}\lambda)\operatorname{Im}\widetilde{\Omega}_\pm(\Lambda(\lambda)) = \frac{\sqrt{3}}{2}|\lambda|^{1/3} + \mathcal{O}(\mu(\lambda)\lambda^{1/3}) \geq |\lambda|^{1/3} - C|\lambda|^{2/3}.$$

By possibly shrinking $\delta \sim 1$, we conclude $\operatorname{Im}\widetilde{\Omega}_-(\Lambda(\lambda)) < 0$ for $\lambda \in (0, \delta]$ and $\operatorname{Im}\widetilde{\Omega}_+(\Lambda(\lambda)) < 0$ for $\lambda \in [-\delta, 0)$. Hence, owing to **SP3**, we conclude $a_+ \neq -$ and $a_- \neq +$.

Next, we will prove $a_+ \neq 0$. For $\lambda \geq 0$, we have

$$\operatorname{Re}\widetilde{\Omega}_0(\Lambda(\lambda)) \leq -\lambda^{1/3} + C\lambda^{2/3}, \qquad \operatorname{Im}\widetilde{\Omega}_0(\Lambda(\lambda)) \lesssim \lambda^{2/3}.$$

Thus, assuming $\Omega(\lambda) = \widetilde{\Omega}_0(\Lambda(\lambda))$ and estimating the imaginary part of (8.7.27) yield

$$\lambda^{2/3}\operatorname{Im}\Omega(\lambda) \lesssim (\operatorname{Im}\Omega(\lambda))^3 + |\operatorname{Im}\Lambda(\lambda)| \lesssim |\lambda|\operatorname{Im}\Omega(\lambda).$$

Hence, we possibly shrink $\delta \sim 1$ and conclude $\operatorname{Im}\Omega(\lambda) = 0$ for $\lambda \in [0, \delta]$. Therefore, $\operatorname{Re}\Omega(\lambda)$ is non-decreasing on $[0, \delta]$ by **SP4'**. Combined with $\Omega_0 = 0$ and $\operatorname{Re}\widetilde{\Omega}_0(\Lambda(\lambda)) \lesssim -\lambda^{1/3}$, we obtain a contradiction. Hence, this implies $a_+ \neq 0$, i.e., $a_+ = +$.

A similar argument excludes $a_- = 0$ and we thus obtain $a_- = -$. Now, (8.7.28) is obtained from the definition of $\widehat{\Omega} = \widetilde{\Omega}_+$, which completes the proof of (iii).

For the proof of (iv), we remark that all estimates follow from Lemma 8.7.10 in the same way as they followed in [4] from Lemma 9.14 in [4]. Indeed, (8.7.30) is the same as (9.129) in [4]. The bound (8.7.31) is shown analogously to (9.129) and (9.130) in [4]. Moreover, (8.7.32) is (9.137) in [4] and (8.7.33) is obtained as (9.109) in [4]. This completes the proof of Proposition 8.7.9. □

8.7.3. Proof of Theorem 8.7.6. Before we prove Theorem 8.7.6, we collect some properties of Ψ_{edge} and $\Psi_{\min}$ which will be useful in the following. We recall that Ψ_{edge} and $\Psi_{\min}$ were defined in (8.7.1).

Lemma 8.7.11 (Properties of $\Psi_{\min}$ and Ψ_{edge})**.**

(i) Let $\widehat{\Omega}$ be defined as in Proposition 8.7.9 (i). Then, for any $\lambda \in \mathbb{R}$, we have

$$\Psi_{\min}(\lambda) = \frac{1}{\sqrt{3}} \mathrm{Im}\, [\widehat{\Omega}(\lambda) - \widehat{\Omega}(0)]. \tag{8.7.38}$$

(ii) Let $\widehat{\Omega}_+$ be defined as in (8.7.21)*. Then, for any $\lambda \geq 0$, we have*

$$\Psi_{\mathrm{edge}}(\lambda) = \frac{1}{2\sqrt{3}} \mathrm{Im}\, \widehat{\Omega}_+(1 + 2\lambda). \tag{8.7.39}$$

(iii) There is a function $\widetilde{\Psi}\colon [0,\infty) \to \mathbb{R}$ with uniformly bounded derivatives and $\widetilde{\Psi}(0) = 0$ such that, for any $\lambda \geq 0$, we have

$$\Psi_{\mathrm{edge}}(\lambda) = \frac{\lambda^{1/2}}{3}(1 + \widetilde{\Psi}(\lambda)), \qquad |\widetilde{\Psi}(\lambda)| \lesssim \min\{\lambda, \lambda^{1/3}\}. \tag{8.7.40}$$

(iv) There is $\varepsilon_* \sim 1$ such that if $|\varepsilon| \leq \varepsilon_*$ then, for any $\lambda \geq 0$, we have

$$\Psi_{\mathrm{edge}}((1+\varepsilon)\lambda) = (1+\varepsilon)^{1/2} \Psi_{\mathrm{edge}}(\lambda) + \mathcal{O}(\varepsilon \min\{\lambda^{3/2}, \lambda^{1/3}\}). \tag{8.7.41}$$

We remark that (8.7.39) was present in (9.127) of [4] but the coefficient $1/(2\sqrt{3})$ was erroneously missing there. The relation in (8.7.41) is identical to (9.145) in [4]. Moreover, we use the proof of [4].

Proof. The parts (i), (ii) and (iii) are direct consequences of the definitions of $\Psi_{\min}$, $\widehat{\Omega}$, Ψ_{edge} and $\widehat{\Omega}_+$.

For the proof of (iv), we choose $\varepsilon_* \le 1/2$ such that $1+\varepsilon \sim 1$ for $|\varepsilon| \le \varepsilon_*$. If $0 \le \lambda \lesssim 1$ then (8.7.41) follows from (8.7.40). For $\lambda \gtrsim 1$, we choose $\varepsilon_* = 1/3$ and then (8.7.41) is a consequence of (8.7.39) above as well as the stability of Cardano's solutions, (9.111) in Lemma 9.17 of [4]. □

In the following proof of Theorem 8.7.6, we will choose appropriate normal coordinates Ω and Λ in each case such that (8.7.11) turns into one of the cubic equations in normal form from Proposition 8.7.9. This procedure has been similarly performed in the proofs of Proposition 9.3, Lemma 9.11, Lemma 9.12 and Section 9.2.2 in [4]. However, owing to the weaker error bounds here, we include the proof for the sake of completeness.

Proof of Theorem 8.7.6. We start with the proof of part (i) (cf. Proposition 9.3 in [4]). Owing to (8.7.14b) and $|\Psi_{\min}(\lambda)| \lesssim |\lambda|^{1/3}$, the statement of (8.7.15) is trivial for $|\omega| \gtrsim 1$ since the error term dominates. Therefore, it suffices to prove (8.7.15) for $|\omega| \le \delta$ with some $\delta \sim 1$.

By possibly shrinking $\rho_* \sim 1$, we can assume that $|\sigma| \le \Pi_* \rho_*^2$ is small enough such that $\psi \sim 1$ by (8.7.13). In the following, we will choose ω-independent complex numbers $\gamma_\nu, \gamma_0, \gamma_1, \ldots, \gamma_7 \in \mathbb{C}$ such that certain relations hold. For each choice, it is easily checked that $|\gamma_k| \lesssim \rho$ for $k = \nu, 0, 1, \ldots, 7$. We divide (8.7.11) by μ_3 and obtain

$$\Theta^3 + \mathrm{i}3\rho(1+\gamma_2)\Theta^2 - 2\rho^2(1+\gamma_1)\Theta + (1+\gamma_0+(1+\gamma_\nu)\nu(\omega))\frac{\kappa}{\psi}\omega = 0, \tag{8.7.42}$$

where γ_ν, γ_0, γ_1 and γ_2 are chosen such that

$$\begin{aligned} \frac{\mu_2}{\mu_3} &= \mathrm{i}3\rho(1+\gamma_2), \\ \frac{\mu_1}{\mu_3} &= -2\rho^2(1+\gamma_1), \\ \frac{\kappa(1+\nu(\omega)) + \mathcal{O}(\rho)}{\mu_3} &= \frac{\kappa}{\psi}(1+\gamma_0+(1+\gamma_\nu)\nu(\omega)). \end{aligned}$$

With these choices, we obtain $\gamma_\nu, \gamma_0, \gamma_1, \gamma_2 = \mathcal{O}(\rho)$, since $\mu_3 = \psi + \mathcal{O}(\rho)$, $\mu_2 = \mathrm{i}3\rho\psi + \mathcal{O}(\rho^2)$ and $\mu_1 = -2\rho^2 + \mathcal{O}(\rho^3)$ owing to (8.7.12), $|\sigma| \le \Pi_*\rho^2$, $\psi \sim 1$ and $|\mu_3| \sim 1$ for sufficiently small $\rho_* \sim 1$. We introduce the normal coordinates

$$\lambda := \Gamma\frac{\omega}{\rho^3}, \qquad \Omega(\lambda) := \sqrt{3}\Big[(1+\gamma_3)\frac{1}{\rho}\Theta\Big(\frac{\rho^3}{\Gamma}\lambda\Big) + \mathrm{i} + \gamma_4\Big], \tag{8.7.43}$$

where $\Gamma := \sqrt{27}\kappa/(2\psi)$. Note that $\Gamma \sim 1$ since $\psi \sim 1$. We choose γ_3 and γ_4 such that the coefficient of the quadratic term of the cubic equation, (8.7.42), in normal coordinates vanishes while the coefficient of the linear term equals to 3. This amounts to the relations

$$-\gamma_4 + \mathrm{i}\gamma_2 + \mathrm{i}\gamma_3 + \mathrm{i}\gamma_2\gamma_3 = 0,$$
$$3(\mathrm{i}+\gamma_4)^2 - \mathrm{i}6(\mathrm{i}+\gamma_4)(1+\gamma_2)(1+\gamma_3) - 2(1+\gamma_1)(1+\gamma_3)^2 = 1.$$

Expressing γ_4 by γ_3 (and γ_2 which has already been chosen) via the first equation and plugging the result into the second equation yield a quadratic equation for γ_3 in terms of γ_1 and γ_2. In this quadratic equation the order one term cancels and hence $\gamma_3 = \mathcal{O}(\rho)$. This also implies $\gamma_4 = \mathcal{O}(\rho)$. Thus, a straightforward computation starting from (8.7.42) shows that $\Omega(\lambda)$ and $\Lambda(\lambda)$ satisfy (8.7.22) with

$$\Lambda(\lambda) := (1+\gamma_5+\mu(\lambda))\lambda + \gamma_6, \qquad \mu(\lambda) := (1+\gamma_7)\nu\Big(\frac{\rho^3}{\Gamma}\lambda\Big),$$

i.e., $\chi_2 = \gamma_5$, $\chi_3 = \gamma_6$ and $\alpha = \rho$ by (8.7.14a).

Here, we chose

$$\gamma_5 = (1+\gamma_3)^3(1+\gamma_0) - 1,$$
$$\gamma_6 = \sqrt{27}(-(\mathrm{i}+\gamma_4)^3 + \mathrm{i}3(1+\gamma_2)(1+\gamma_3)(\mathrm{i}+\gamma_4)^2 + 2(1+\gamma_1)(1+\gamma_3)^2(\mathrm{i}+\gamma_4)),$$
$$\gamma_7 = (1+\gamma_3)^3(1+\gamma_\nu) - 1.$$

Since $\gamma_\nu, \gamma_0, \dots, \gamma_4 = \mathcal{O}(\rho)$, we conclude $\gamma_5, \gamma_6, \gamma_7 = \mathcal{O}(\rho)$. Hence, from (8.7.23) and (8.7.43), we obtain $\delta \sim 1$ and $\chi_* \sim 1$ such that

$$\begin{aligned}\operatorname{Im}\Theta(\omega) &= \operatorname{Im}\frac{\rho}{1+\gamma_3}\frac{1}{\sqrt{3}}[\Omega(\lambda) - \Omega_0] \\ &= \rho\Psi_{\min}\Big(\Gamma\frac{\omega}{\rho^3}\Big) + \mathcal{O}\left(\rho^2\min\{|\lambda|, |\lambda|^{1/3}\} + \rho^2\min\{|\lambda|, |\lambda|^{2/3}\}\right)\end{aligned}$$

for $|\lambda| \le \delta/\rho^3$ if $\rho \le \min\{\chi_*, \rho_*\}$. Here, we also used (8.7.24) to expand $\rho/(1+\gamma_3)$ and (8.7.38). By employing (8.7.43) again and replacing ρ_* by $\min\{\chi_*, \rho_*\}$, we conclude (8.7.15).

We now turn to the proof of part (ii) of Theorem 8.7.6. Since $\rho = 0$, the cubic equation (8.7.11) simplifies to the following equation

$$\psi\Theta(\omega)^3 + \sigma\Theta(\omega)^2 + \kappa(1+\nu(\omega))\omega = 0. \tag{8.7.44}$$

We now prove Theorem 8.7.6 (ii) (a), i.e., the case $\sigma = 0$ (cf. Lemma 9.12 in [4]). For any $\delta \sim 1$, the assertion is trivial for $|\omega| \geq \delta$ since the error term dominates $|\omega|^{1/3}$ and $\operatorname{Im}\Theta(\omega)$ in this case (compare (8.7.14b)). Therefore, it suffices to prove the lemma for $|\omega| \leq \delta$ with some $\delta \sim 1$. We choose the normal coordinates

$$\lambda := \omega, \qquad \Omega(\lambda) := \Big(\frac{\psi}{\kappa}\Big)^{1/3}\Theta(\lambda),$$

and notice that the cubic equation (8.7.44) becomes (8.7.27) with $\mu(\lambda) = \nu(\lambda)$. The bound (8.7.14a) implies $|\mu(\lambda)| \lesssim |\lambda|^{1/3}$. Thus, (8.7.17) is a consequence of Proposition 8.7.9 (iii). This completes the proof of (ii) (a).

For the proof of Theorem 8.7.6 (ii) (b), we first show the following auxiliary lemma (cf. Lemma 9.11 in [4]).

Lemma 8.7.12 (Simple edge)**.** *Let the assumptions of Theorem 8.7.6 (ii) hold true. If $\sigma \neq 0$ then there is $c_* \sim 1$ such that, for $|\omega| \leq c_*|\sigma|^3$, we have*

$$\operatorname{Im}\Theta(\omega) = \begin{cases} \sqrt{\kappa}\Big|\dfrac{\omega}{\sigma}\Big|^{1/2} + \mathcal{O}\Big(\big(|\nu(\omega)| + |\sigma|^{-1}|\Theta(\omega)|\big)\Big|\dfrac{\omega}{\sigma}\Big|^{1/2}\Big), & \textit{if } \operatorname{sign}\omega = \operatorname{sign}\sigma, \\ 0, & \textit{if } \operatorname{sign}\omega = -\operatorname{sign}\sigma. \end{cases} \tag{8.7.45}$$

Moreover, we have $|\Theta(\omega)| \lesssim |\omega/\sigma|^{1/2}$ for $|\omega| \leq c_*|\sigma|^3$.

Proof. Dividing (8.7.44) by $\kappa\sigma$ yields

$$\Big(1 + \frac{\psi}{\sigma}\Theta(\omega)\Big)\frac{\Theta(\omega)^2}{\kappa} + (1+\nu(\omega))\frac{\omega}{\sigma} = 0. \tag{8.7.46}$$

We introduce λ, $\Omega(\lambda)$ and $\mu(\lambda)$ defined by

$$\lambda := \frac{\omega}{\sigma}, \qquad \Omega(\lambda) := \frac{1}{\sqrt{\kappa}}\Theta(\sigma\lambda), \qquad \mu(\lambda) := \frac{1+\nu(\sigma\lambda)}{1+\psi\sigma^{-1}\Theta(\sigma\lambda)} - 1.$$

In the normal coordinates λ and $\Omega(\lambda)$, (8.7.46) viewed as a quadratic equation, fulfills (8.7.25) with the above choice of $\mu(\lambda)$. Since $|\psi\sigma^{-1}\Theta(\sigma\lambda)| \lesssim |\sigma|^{-2/3}|\lambda|^{1/3}$ by (8.7.14b), there is $c_* \sim 1$ such that

$$|\mu(\lambda)| \lesssim |\nu(\sigma\lambda)| + |\sigma|^{-1}|\Theta(\sigma\lambda)| \lesssim |\sigma|^{-2/3}|\lambda|^{1/3}, \qquad |\mathrm{Im}\, \mu(\lambda)| \lesssim |\sigma|^{-1}\mathrm{Im}\, \Theta(\sigma\lambda) \quad (8.7.47)$$

for $|\lambda| \leq c_*|\sigma|^2$ by (8.7.14a), (8.7.14b) and (8.7.16). Hence, we apply Proposition 8.7.9 (ii) with $\gamma \sim |\sigma|^{-1}$ in **SP4'** and obtain (8.7.45) with an error term $\mathcal{O}(|\mu(\lambda)||\lambda|^{1/2})$ instead, as well as $|\Theta(\omega)| \lesssim |\sigma|^{-1/2}|\omega|^{1/2}$. Thus, the first bound in (8.7.47) completes the proof of (8.7.45). □

From the second case in (8.7.45), we conclude the second case in (8.7.18). The first case in (8.7.18) and (8.7.20) are trivial if $|\omega| \gtrsim 1$ due to (8.7.14b) and (8.7.4a). Hence, it suffices to prove this case for $|\omega| \leq \delta$ with some $\delta \sim 1$. If $|\sigma| \gtrsim 1$ then the first case in (8.7.18) also follows from (8.7.45) with $\delta := c_*|\sigma|^3$. Indeed, from (8.7.40), we conclude

$$\sqrt{\kappa}\left|\frac{\omega}{\sigma}\right|^{1/2} = c\widehat{\Delta}^{1/3}\Psi_{\mathrm{edge}}\left(\frac{|\omega|}{\widehat{\Delta}}\right) + \mathcal{O}(|\omega|^{3/2}),$$

where c and $\widehat{\Delta}$ are defined as in (8.7.19). Since $|\omega| \lesssim \varepsilon(\omega)$ for $|\omega| \leq \delta$ and $\varepsilon(\omega)$ defined as in (8.7.19) we obtain the first case in (8.7.18) if $|\sigma| \gtrsim 1$. Similarly, $|\Theta(\omega)| \lesssim |\omega/\sigma|^{1/2}$ by Lemma 8.7.12 yields (8.7.20) if $|\omega| \leq \delta$ and $|\sigma| \gtrsim 1$. Hence, it remains to show the first case in (8.7.18) and (8.7.20) if $|\sigma| \leq \sigma_*$ for some $\sigma_* \sim 1$. In fact, we choose $\sigma_* \sim 1$ so small that $\psi \sim 1$ by (8.7.13) and $\widehat{\Delta} < 1$ for $|\sigma| \leq \sigma_*$. In order to apply Proposition 8.7.9 (iv), we introduce

$$\lambda := \frac{2}{\widehat{\Delta}}\omega, \qquad \Omega(\lambda) := 3\frac{\psi}{|\sigma|}\Theta\left(\frac{\widehat{\Delta}}{2}\lambda\right) + \mathrm{sign}\,\sigma, \qquad \mu(\lambda) := \nu\left(\frac{\widehat{\Delta}}{2}\lambda\right) \qquad (8.7.48)$$

(cf. (9.96) and (9.99) in [4]). The cubic (8.7.44) takes the form (8.7.29) in the normal coordinates λ and $\Omega(\lambda)$ with the above choice of $\mu(\lambda)$ and $s = \mathrm{sign}\,\sigma$ in (8.7.29). By (8.7.14a), we have $|\mu(\lambda)| \lesssim \widehat{\Delta}^{1/3}|\lambda|^{1/3}$. We set $\widehat{\gamma} := \widehat{\Delta}^{1/3}$. Therefore, Proposition 8.7.9 (iv) and (8.7.39) yield $\delta \sim 1$ and possibly smaller $\sigma_* := \min\{\sigma_*, \gamma_*\} \sim 1$ such that the first case in (8.7.18) holds true for $|\sigma| \leq \sigma_*$ and $|\omega| \leq \delta$ as $\mu(\lambda) = \nu(\omega)$ and $\widehat{\Delta} \sim |\sigma|^3$. Moreover, (8.7.31) implies (8.7.20) for $|\omega| \leq \delta$. This completes the proof of (ii) (b) and hence of Theorem 8.7.6. □

8.7.4. Proof of Theorem 8.7.1 and Proposition 8.7.4. In this section, we prove Theorem 8.7.1 and Proposition 8.7.4. Some parts of the following proof resemble the proofs of Theorem 2.6, Proposition 9.3 and Proposition 9.8 in [4]. However, owing to the weaker assumptions, we present it here for the sake of completeness.

Proof of Theorem 8.7.1 and Proposition 8.7.4. We will only prove the statements in Proposition 8.7.4. Theorem 8.7.1 is a direct consequence of this proposition as well as Lemma 8.4.8 (ii) and Proposition 8.6.1.

Along the proof of Proposition 8.7.4, we will shrink $\delta_* \sim 1$ such that (8.7.3) holds true for all $\omega \in [-\delta_*, \delta_*] \cap J \cap D$. We will transfer the expansions of Θ in Theorem 8.7.6 to expansions of v by means of (8.6.1). To that end, we take the imaginary part of (8.6.1) and obtain

$$v(\tau_0 + \omega) = v(\tau_0) + \pi^{-1}\mathrm{Re}\, b \mathrm{Im}\,\Theta(\omega) + \pi^{-1}\mathrm{Im}\, b \mathrm{Re}\,\Theta(\omega) + \pi^{-1}\mathrm{Im}\, r(\omega). \tag{8.7.49}$$

We first establish (8.7.3) at a shape regular point $\tau_0 \in (\mathrm{supp}\,\rho) \setminus \partial\,\mathrm{supp}\,\rho$ which is a local minimum of $\tau \mapsto \rho(\tau)$. If $\rho = \rho(\tau_0) = 0$, i.e., the case of a cusp at τ_0, case (c), then $\sigma = 0$. Indeed, if σ were not 0, then, by the second case in (8.7.18), $\mathrm{Im}\,\Theta(\omega)$ would vanish on one side of τ_0. By the third bound in (8.6.8), this would imply the vanishing of ρ as well, contradicting to $\tau_0 \in \mathrm{supp}\,\rho \setminus \partial\,\mathrm{supp}\,\rho$. Hence, for any $\delta_* \sim 1$, (8.7.17) and (8.7.49) immediately yield (8.7.3) for $\omega \in [-\delta_*, \delta_*] \cap J \cap D$ with $h = (2\pi)^{-1} b\sqrt{3}(\kappa/\psi)^{1/3}$ using (8.6.7a), (8.6.7b) and $b = b^*$ due to $\rho = 0$.

We now assume $\rho > 0$ which corresponds to an internal nonzero minimum at τ_0, case (d). Thus, the following lemma implies that the condition $|\sigma| \le \Pi_* \rho^2$, $\sigma = \sigma(\tau_0)$, needed to apply Theorem 8.7.6 (i) is fulfilled. We will prove Lemma 8.7.13 at the end of this section.

Lemma 8.7.13 (Bound on $|\sigma|$ at nonzero local minimum)**.** *There are thresholds $\rho_* \sim 1$ and $\Pi_* \sim 1$ such that*

$$|\sigma(\tau_0)| \le \Pi_* \rho(\tau_0)^2$$

for all shape regular points $\tau_0 \in \mathrm{supp}\,\rho$ which are a local minimum of ρ and satisfy $0 < \rho(\tau_0) \le \rho_*$.

Hence, (8.7.15), (8.7.49) and (8.6.7b) yield (8.7.3) with $\widetilde{\rho} = \rho\Gamma^{-1/3}$ and $h = \pi^{-1}\Gamma^{1/3}\mathrm{Re}\, b$. Here, we also used

$$\rho|\Theta(\omega)| + |\Theta(\omega)|^2 + |\omega| + \min\{\rho^{-1}|\omega|, |\omega|^{2/3}\} \lesssim \frac{|\omega|}{\rho}\mathbf{1}(|\omega| \lesssim \rho^3) + \Psi(\omega)^2, \qquad (8.7.50)$$

which is a consequence of (8.6.7a), (8.7.4b) for $|\omega| \lesssim 1$, as well as $\mathrm{Re}\, b \sim 1$ and $\mathrm{Im}\, b = \mathcal{O}(\rho)$. This completes the proof of (8.7.3) for shape regular points $\tau_0 \in (\mathrm{supp}\,\rho)\setminus\partial\,\mathrm{supp}\,\rho$, cases (c) and (d).

We now turn to the proof of (8.7.3) at an edge τ_0, case (a), i.e., for a shape regular point $\tau_0 \in \partial\,\mathrm{supp}\,\rho$. We first prove a version of (8.7.3) with $\widehat{\Delta}$ in place of Δ, (8.7.51) below. In a second step, we then replace $\widehat{\Delta}$ by Δ to obtain (8.7.3).

Since $\tau_0 \in \partial\,\mathrm{supp}\,\rho$, we have $\rho = \rho(\tau_0) = 0$. Therefore, $v(\tau_0) = 0$ since $\langle\,\cdot\,\rangle$ is a faithful trace and $v(\tau_0)$ is positive semidefinite. As $\tau_0 \in \partial\,\mathrm{supp}\,\rho$, we have $\sigma(\tau_0) \neq 0$. Indeed, assuming $\sigma(\tau_0) = 0$, using Theorem 8.7.6 (ii) (a), taking the imaginary part of (8.6.1) as well as applying the third bound in (8.6.8) and the second bound in (8.6.7a) yield the contradiction $\tau_0 \in (\mathrm{supp}\,\rho) \setminus \partial\,\mathrm{supp}\,\rho$. Recalling the definitions of $\widehat{\Delta}$ and c from (8.7.19), (8.7.49) and the first case in (8.7.18) yield

$$v(\tau_0 + \omega) = \pi^{-1}c\widehat{\Psi}(\omega)b + \mathcal{O}(\widehat{\Psi}(\omega)^2), \qquad \widehat{\Psi}(\omega) := \widehat{\Delta}^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{|\omega|}{\widehat{\Delta}}\Big) \qquad (8.7.51)$$

for any $\omega \in [-\delta_*, \delta_*] \cap J \cap D$ with $\mathrm{sign}\,\omega = \mathrm{sign}\,\sigma$ and some $\delta_* \sim 1$. Here, we also used $b = b^* \sim 1$, the first bound in (8.6.5), (8.7.20) and $\varepsilon(\omega) \sim \widehat{\Psi}(\omega)$ by (8.7.4b) to obtain

$$|\Theta(\omega)|^2 + |\omega| + (|\Theta(\omega)| + |\omega| + \varepsilon(\omega))\varepsilon(\omega) \lesssim \widehat{\Psi}(\omega)^2$$

for any $\omega \in [-\delta_*, \delta_*] \cap J \cap D$ with $\mathrm{sign}\,\omega = \mathrm{sign}\,\sigma$ and some $\delta_* \sim 1$. This means that we have shown (8.7.3) with Ψ replaced by $\widehat{\Psi}$.

We now replace $\widehat{\Delta}$ by Δ in (8.7.51) to obtain (8.7.3). To that end, we first assume that $|\sigma| \gtrsim 1$ and $\Delta \lesssim 1$. The second part of (8.7.18) implies $|\sigma|^3 \lesssim \Delta \lesssim 1$ and thus $|\sigma|^3 \sim \Delta \sim 1$. Since $|\sigma|^3 \sim \widehat{\Delta}$ we conclude $\widehat{\Delta} \sim \Delta$. Therefore, we obtain

$$\widehat{\Delta}^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{|\omega|}{\widehat{\Delta}}\Big) = \Big(\frac{\Delta}{\widehat{\Delta}}\Big)^{1/6}\Delta^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{|\omega|}{\Delta}\Big) + \mathcal{O}(\min\{|\omega|^{3/2}, |\omega|^{1/3}\}).$$

Here, we used $\Psi_{\mathrm{edge}}(|\lambda|) \lesssim |\lambda|^{1/3}$ for $|\lambda| \gtrsim 1$ and (8.7.40) otherwise. Applying this relation to (8.7.51) yields (8.7.3) for $\omega \in [-\delta_*, \delta_*] \cap J \cap D$ with $\operatorname{sign}\omega = \operatorname{sign}\sigma$, $\delta_* \sim 1$ and $h := \pi^{-1}c(\Delta/\widehat{\Delta})^{1/6}b \sim 1$ for $|\sigma| \gtrsim 1$ and $\Delta \lesssim 1$.

The next lemma shows that $|\sigma| \gtrsim 1$ at the edge of a gap of size $\Delta \gtrsim 1$. We postpone its proof until the end of this section.

Lemma 8.7.14 (σ at an edge of a large gap)**.** *Let $\tau_0 \in \partial \operatorname{supp} \rho$ be a shape regular point for m on J. If $|\inf J| \gtrsim 1$ and there is $\varepsilon \sim 1$ such that $\rho(\tau) = 0$ for all $\tau \in [\tau_0 - \varepsilon, \tau_0]$ then $|\sigma| \sim 1$. We also have $|\sigma| \sim 1$ if $\sup J \gtrsim 1$ and $\rho(\tau) = 0$ for all $\tau \in [\tau_0, \tau_0 + \varepsilon]$ and some $\varepsilon \sim 1$.*

Under the assumptions of the previous lemma, we set $\Delta := 1$ and obtain trivially $\widehat{\Delta} \sim 1 \sim \Delta$. Thus, (8.7.51) implies (8.7.3) by the same argument as in the case $\Delta \lesssim 1$.

For $|\sigma| \le \sigma_*$ with some sufficiently small $\sigma_* \sim 1$, we will prove below with the help of the following Lemma 8.7.15 and (8.7.41) that replacing $\widehat{\Delta}$ by Δ in (8.7.51) yields an affordable error. We present the proof of Lemma 8.7.15 at the end of this section.

Lemma 8.7.15 (Size of small gap)**.** *Let $\tau_0, \tau_1 \in \partial \operatorname{supp} \rho$, $\tau_1 < \tau_0$, be two shape regular points for m on J_0 and J_1, respectively, where $J_0, J_1 \subset \mathbb{R}$ are two open intervals with $0 \in J_0 \cap J_1$. We assume $|\inf J_0| \gtrsim 1$ and $\sup J_1 \gtrsim 1$ as well as $(\tau_1, \tau_0) \cap \operatorname{supp} \rho = \varnothing$. We set $\Delta(\tau_0) := \tau_0 - \tau_1$. Then there is $\widetilde{\sigma} \sim 1$ such that if $|\sigma(\tau_0)| \le \widetilde{\sigma}$ and $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim |\tau_0 - \tau_1|^\zeta$ for some $\zeta \in (0, 1/3]$ then*

$$\frac{\Delta(\tau_0)}{\widehat{\Delta}(\tau_0)} = 1 + \mathcal{O}(\sigma(\tau_0)).$$

The same statement holds true when τ_0 is replaced by τ_1 with $\Delta(\tau_1) := \tau_0 - \tau_1$.

From Lemma 8.7.15, we conclude that there is $\gamma \in \mathbb{C}$ such that $|\gamma| \lesssim 1$ and $\Delta = (1 + \gamma|\sigma|)\widehat{\Delta}$. By possibly shrinking $\sigma_* \sim 1$, we can assume that $|\gamma\sigma| \le \varepsilon_*$ for $|\sigma| \le \sigma_*$, where $\varepsilon_* \sim 1$ is chosen as in Lemma 8.7.11 (iv). Thus, (8.7.41) yields

$$\widehat{\Delta}^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{|\omega|}{\widehat{\Delta}}\Big) = \Big(\frac{\Delta}{\widehat{\Delta}}\Big)^{1/6}\Delta^{1/3}\Psi_{\mathrm{edge}}\Big(\frac{|\omega|}{\Delta}\Big) + \mathcal{O}\Big(\min\Big\{\frac{|\omega|^{3/2}}{\Delta^{5/6}}, |\omega|^{1/3}\Big\}\Big).$$

Hence, choosing $h := \pi^{-1}c(\Delta/\widehat{\Delta})^{1/6}b$ as before and noticing $h \sim 1$ yields (8.7.3) in the missing regime. This completes the proof of Proposition 8.7.4. As we have already explained, Theorem 8.7.1 follows immediately. □

The core of the proof of Lemma 8.7.13 is an effective monotonicity estimate on v, see (8.7.52) below, which is the analogue of (9.20) in Lemma 9.2 of [4]. Owing to the weaker assumptions on the coefficients of the cubic equation, we need to present an upgraded proof here. In fact, the bound in (9.20) of [4] contained a typo. It should have read as

$$(\operatorname{sign}\sigma(\tau))\partial_\tau v(\tau) \gtrsim \frac{1}{\langle v(\tau)\rangle(1+|\sigma(\tau)|)}$$

for $\tau \in \mathbb{D}_{\varepsilon_*}$ satisfying $\Pi(\tau) \geq \Pi_*$. However, this does not affect the correctness of the argument in [4].

Proof of Lemma 8.7.13. In the whole proof, we will use the notation of Definition 8.7.3. We will show below that there are $\rho_* \sim 1$ and $\Pi_* \sim 1$ such that

$$(\operatorname{sign}\kappa_1\sigma(\tau))\partial_\tau v(\tau) \gtrsim \rho(\tau)^{-1} \tag{8.7.52}$$

for all $\tau \in \mathbb{R}$ which satisfy $\rho(\tau) \in (0, \rho_*]$ and $|\sigma(\tau)| \geq \Pi_*\rho(\tau)^2$ and are admissible points for the shape analysis.

Now, we first conclude the statement of the lemma from (8.7.52) through a proof by contradiction. If τ_0 satisfies the conditions of Lemma 8.7.13 then $\partial_\tau\rho(\tau_0) = 0$ as τ_0 is a local minimum of ρ. Assuming $|\sigma(\tau_0)| \geq \Pi_*\rho(\tau_0)^2$ and applying $\langle\,\cdot\,\rangle$ to (8.7.52) yield the contradiction $\partial_\tau\rho(\tau_0) > 0$.

For the proof of (8.7.52) we start by proving a relation for $\partial_\tau v(\tau)$. We divide (8.6.1) by ω, use $\Theta(0) = 0$ and $r(0) = 0$ as well as take the limit $\omega \to 0$ to obtain $\partial_\tau m(\tau) = b\partial_\omega\Theta(0) + \partial_\omega r(0)$. Taking the imaginary part of the previous relation yields

$$\pi\partial_\tau v(\tau) = \operatorname{Im}[b\partial_\omega\Theta(0)] + \operatorname{Im}\partial_\omega r(0). \tag{8.7.53}$$

We divide (8.6.7b) by ω, employ the first bound in (8.6.7a) and obtain

$$\left\|\frac{r(\omega)}{\omega}\right\| \lesssim 1 + \left|\frac{\Theta(\omega)}{\omega}\right|^2 \lesssim 1 + \frac{|\omega|}{\rho^4}.$$

By sending $\omega \to 0$ and using $r(0) = 0$, we conclude

$$\|\operatorname{Im}\partial_\omega r(0)\| \lesssim 1. \tag{8.7.54}$$

We divide (8.6.3) by $\mu_1\omega$, take the limit $\omega \to 0$ and use $\lim_{\omega\to 0}\Theta(\omega) = \Theta(0) = 0$ to obtain

$$\begin{aligned}\partial_\omega\Theta(0) = -\frac{\Xi(0)\bar{\mu}_1}{|\mu_1|^2} &= \frac{(\kappa+\mathcal{O}(\rho))(\mathrm{i}\kappa_1\rho\sigma + 2\rho^2\psi + \mathcal{O}(\rho^3+\rho^2|\sigma|))}{4\rho^4|\psi+\mathcal{O}(\rho+|\sigma|)|^2 + \rho^2|\kappa_1\sigma + \mathcal{O}(\rho^2+\rho|\sigma|)|^2} \\ &= \frac{\kappa}{\rho}\frac{\mathrm{i}\kappa_1\sigma + 2\rho\psi + \mathcal{O}(\rho^2+\rho|\sigma|)}{4\rho^2|\psi+\mathcal{O}(\rho+|\sigma|)|^2 + |\kappa_1\sigma+\mathcal{O}(\rho^2+\rho|\sigma|)|^2},\end{aligned} \tag{8.7.55}$$

where we employed $|\mu_1|^2 = 4\rho^4|\psi+\mathcal{O}(\rho+|\sigma|)|^2+\rho^2|\kappa_1\sigma+\mathcal{O}(\rho^2+\rho|\sigma|)|^2$ as $\rho,\psi,\kappa_1,\sigma \in \mathbb{R}$. Thus, we obtain

$$\rho|\mathrm{Re}\,\partial_\omega\Theta(0)| \lesssim \frac{\rho+\rho|\sigma|}{\rho^2|\psi+\mathcal{O}(\rho+|\sigma|)|^2 + |\kappa_1\sigma+\mathcal{O}(\rho^2+\rho|\sigma|)|^2}. \tag{8.7.56}$$

Therefore, using $b = b^* + \mathcal{O}(\rho)$, $b + b^* \sim 1$, $\kappa \sim 1$ and $|\kappa_1| \sim 1$ yields

$$(\mathrm{sign}\,\kappa_1\sigma)\mathrm{Im}\,[b\partial_\omega\Theta(0)] \gtrsim \frac{\rho^{-1}|\sigma| + \mathcal{O}(\rho+|\sigma|) + \mathcal{O}(\rho+\rho|\sigma|)}{|\sigma+\mathcal{O}(\rho^2+\rho|\sigma|)|^2 + \rho^2|\psi+\mathcal{O}(\rho+|\sigma|)|^2} \gtrsim \frac{|\sigma|}{|\sigma|^2+\rho^2}\frac{1}{\rho}.$$

Here, in the first step, the error term $\mathcal{O}(\rho+\rho|\sigma|)$ in the numerator originates from the second term in

$$\begin{aligned}(\mathrm{sign}\,\kappa_1\sigma)\mathrm{Im}\,[b\partial_\omega\Theta(0)] &= (\mathrm{sign}\,\kappa_1\sigma)\Big(\mathrm{Re}\,b\mathrm{Im}\,\partial_\omega\Theta(0) + \mathrm{Im}\,b\mathrm{Re}\,\partial_\omega\Theta(0)\Big) \\ &\gtrsim (\mathrm{sign}\,\kappa_1\sigma)\mathrm{Im}\,\partial_\omega\Theta(0) - \rho|\mathrm{Re}\,\partial_\omega\Theta(0)|\end{aligned} \tag{8.7.57}$$

and applying (8.7.56) to it. We applied (8.7.55) to the first term on the right-hand side of (8.7.57). In the last estimate, we used $\psi, |\sigma|, \rho \lesssim 1$ and $|\sigma| \geq \Pi_*\rho^2$ for some large $\Pi_* \sim 1$ as well as $\rho \leq \rho_*$ for some small $\rho_* \sim 1$. Employing $|\sigma| \geq \Pi_*\rho^2$ once more, the factor $|\sigma|/(|\sigma|^2+\rho^2)$ on the right-hand side scales like $(1+|\sigma|)^{-1} \gtrsim 1$. Hence, we conclude from (8.7.53) and (8.7.54) that

$$(\mathrm{sign}\,\kappa_1\sigma)\partial_\tau v(\tau) \gtrsim \frac{1}{\rho} + \mathcal{O}(1).$$

By choosing $\rho_* \sim 1$ sufficiently small, we obtain (8.7.52). This completes the proof of Lemma 8.7.13. □

Proof of Lemma 8.7.14. We prove both cases, $\rho(\tau) = 0$ for all $\tau \in [\tau_0 - \varepsilon, \tau_0]$ or for all $\tau \in [\tau_0, \tau_0 + \varepsilon]$, in parallel. We can assume that $|\sigma| \leq \widetilde{\sigma}$ for any $\widetilde{\sigma} \sim 1$ as the statement trivially holds true otherwise. We choose $(\delta, \varrho, \gamma_*)$ as in Proposition 8.7.9 (iv),

$\widehat{\Delta}$ as in (8.7.19), normal coordinates $(\lambda, \Omega(\lambda))$ as in (8.7.48) as well as $\widehat{\gamma} = \widehat{\Delta}^{1/3}$ and $s = \operatorname{sign}\sigma$. We set $\lambda_3 := 2 + \varrho\widehat{\Delta}^{1/3}$ (cf. (8.7.35)) and $\omega_3 := \widehat{\Delta}\lambda_3/2$. There is $\widetilde{\sigma} \sim 1$ such that $\widehat{\Delta} \le \gamma_*^3$ for $|\sigma| \le \widetilde{\sigma}$ due to $\widehat{\Delta} \sim |\sigma|^3$ by (8.6.6) and the definition of $\widehat{\Delta}$ in (8.7.19). Hence, $\omega_3 \le C|\sigma|^3$ and, by possibly shrinking $\widetilde{\sigma} \sim 1$, we obtain $-\omega_3 \operatorname{sign}\sigma \in J$ for $|\sigma| \le \widetilde{\sigma}$ due to the assumption on J ($|\inf J| \gtrsim 1$ or $\sup J \gtrsim 1$). From (8.7.33), we obtain $\operatorname{Im}\Omega(-\lambda_3 \operatorname{sign}\sigma) > 0$. Hence, $\operatorname{Im}\Theta(-\omega_3 \operatorname{sign}\sigma) > 0$. From the third bound in (8.6.8), the second bound in (8.6.7a) and $\omega_3 \lesssim |\sigma|^3$, we conclude $v(-\omega_3 \operatorname{sign}\sigma) > 0$ for $|\sigma| \le \widetilde{\sigma}$ and sufficiently small $\widetilde{\sigma} \sim 1$. Thus, $\rho(-\omega_3 \operatorname{sign}\sigma) > 0$ which implies $\omega_3 > \varepsilon$. Therefore, $|\sigma|^3 \gtrsim \omega_3 > \varepsilon \sim 1$ which completes the proof of Lemma 8.7.14. □

We finish this section by proving Lemma 8.7.15. It is similarly proven as Lemma 9.17 in [4]. We present the proof due to the weaker assumptions of Lemma 8.7.15. The main difference is the proof of (8.7.59) below (cf. (9.138) in [4]). In [4], Θ could be explicitly represented in terms of m, i.e,

$$\Theta(\omega) = \langle f, m(\tau_0 + \omega) - m(\tau_0)\rangle$$

(cf. (9.8) and (8.10c) in [4] with $\alpha = 0$). In our setup, b and r do not necessarily define an orthogonal decomposition (cf. (8.6.1)).

Proof of Lemma 8.7.15. Let $(\delta, \varrho, \gamma_*)$ be chosen as in Proposition 8.7.9 (iv). We choose $\widehat{\Delta}$ as in (8.7.19) and normal coordinates as in (8.7.48) as well as $\widehat{\gamma} = \widehat{\Delta}^{1/3}$ and $s = \operatorname{sign}\sigma$. We assume $\widehat{\Delta} \le \gamma_*^3$ in the following and define λ_3 as in (8.7.35). By using $|\inf J_0| \gtrsim 1$ as in the proof of Lemma 8.7.14, we find $\widetilde{\sigma} \sim 1$ such that $-\omega_3 \in J_0$ for $\omega_3 := \lambda_3\widehat{\Delta}/2$ and $|\sigma| \le \widetilde{\sigma}$. Thus, $-\Delta = \tau_1 - \tau_0 \in J_0$. We set

$$\lambda_0 := \inf\{\lambda > 0 \colon \operatorname{Im}\Omega(\lambda) > 0\}$$

and remark that $\lambda_0 = 2\Delta/\widehat{\Delta}$ due to the definition of Δ and the third bound in (8.6.8). From (8.7.33), we conclude $\lambda_0 \le \lambda_3$. Thus, $\Delta \le \widehat{\Delta}(1 + \mathcal{O}(\widehat{\gamma})) = \widehat{\Delta}(1 + \mathcal{O}(|\sigma|))$ as $\varrho \sim 1$ and $\widehat{\gamma} \sim |\sigma|$. Therefore, it suffices to show the opposite bound,

$$\Delta \ge \widehat{\Delta}(1 + \mathcal{O}(|\sigma|)). \tag{8.7.58}$$

If $\lambda_0 \geq \lambda_2 := 2 - \varrho\widehat{\Delta}^{1/3}$ (cf. (8.7.35)) then we have (8.7.58) as $\widehat{\Delta}^{1/3} \sim |\sigma|$ and $\varrho \sim 1$. If $\lambda_0 < \lambda_2$ then we will prove below that

$$\operatorname{Im}\Omega(\lambda_0 + \xi) \gtrsim \xi^{1/2} \tag{8.7.59}$$

for $\xi \in [0, 1]$. From (8.7.32), we then conclude

$$c_0(\lambda_2 - \lambda_0)^{1/2} \leq \operatorname{Im}\Omega(\lambda_2) \leq C_1|\sigma|^{1/2}$$

as $\widehat{\gamma} \sim |\sigma|$. Hence,

$$\lambda_0 \geq \lambda_2 - (C_1/c_0)^2|\sigma| \geq 2 - C|\sigma|,$$

where we used $\lambda_2 = 2 - \varrho\widehat{\gamma}$ and $\varrho \sim 1$ in the last step. This shows (8.7.58) also in the case $\lambda_0 < \lambda_2$. Therefore, the proof of the lemma will be completed once (8.7.59) is proven.

In order to prove (8.7.59), we translate it into the coordinates ω relative to τ_0 and v. From $\lambda_0 < \lambda_2$, we obtain

$$\Delta < (1 - \varrho\widehat{\Delta}^{1/3})\widehat{\Delta} \lesssim |\sigma|^3. \tag{8.7.60}$$

Since

$$\pi v(\tau_0 - \Delta - \widetilde{\omega}) = b\operatorname{Im}\Theta(-\Delta - \widetilde{\omega}) + \operatorname{Im} r(-\Delta - \widetilde{\omega}),$$

the bound (8.7.59) would follow from

$$v(\tau_0 - \Delta - \widetilde{\omega}) \gtrsim \widehat{\Delta}(\tau_0)^{-1/6}|\widetilde{\omega}|^{1/2} \tag{8.7.61}$$

for sufficiently small $\Delta \lesssim |\sigma|^3 \leq \widetilde{\sigma}^3$ and $\widetilde{\omega} \leq \widetilde{\delta}$ due to the third bound in (8.6.8). Since $v(\tau_1) = 0$ and $\tau_1 = \tau_0 - \Delta$ is a shape regular point, we conclude from (8.7.51) that

$$v(\tau_1 - \widetilde{\omega}) \gtrsim \widehat{\Delta}(\tau_1)^{-1/6}|\widetilde{\omega}|^{1/2}$$

for $|\widetilde{\omega}| \leq \delta$. Therefore, it suffices to show that

$$\widehat{\Delta}(\tau_1) \lesssim \widehat{\Delta}(\tau_0) \tag{8.7.62}$$

in order to verify (8.7.61). Owing to $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim \Delta^\zeta$ and (8.7.60), we have

$$|\sigma(\tau_1)| \lesssim |\sigma(\tau_0)| + \Delta^\zeta \lesssim |\sigma(\tau_0)|^{3\zeta}.$$

We allow for a smaller choice of $\widetilde{\sigma} \sim 1$ and assume $\psi(\tau_1) \sim \psi(\tau_0) \sim 1$ by (8.6.6). Assuming without loss of generality $\widehat{\Delta}(\tau_0) < 1$ and $\widehat{\Delta}(\tau_1) < 1$, we obtain (8.7.62) by the definition of $\widehat{\Delta}$ in (8.7.19). We thus get (8.7.62) and hence (8.7.61). This proves (8.7.59) and completes the proof of Lemma 8.7.15. □

8.7.5. Proofs of Theorem 8.7.2 and Proposition 8.7.5.

Proof of Proposition 8.7.5. We start with the proof of part (i). We apply $\langle \cdot \rangle$ to (8.7.3), use $\rho = \langle v \rangle$ and obtain $\langle h \rangle$ from the definitions of h in the four cases given in the proof of Proposition 8.7.4. Indeed, by using the relations

$$\langle b \rangle = \pi + \mathcal{O}(\rho), \qquad c^3 = 4\Gamma, \tag{8.7.63}$$

which are proven below, as well as Lemma 8.7.15 in the cases (a) and (b) and the stronger error estimate (8.7.50) in case (d), we conclude part (i) of Proposition 8.7.5 up to the proof of (8.7.63).

The first relation in (8.7.63) follows from applying $\langle \cdot \rangle$ to (8.5.14a) and using (8.5.13a), Corollary 8.14.2 with $\tau_0 \in \operatorname{supp} \rho$, the cyclicity of $\langle \cdot \rangle$ and (8.5.19). The second relation in (8.7.63) is a consequence of the definition of c in (8.7.19) and the definition of Γ in Theorem 8.7.6 (i). This completes the proof of part (i).

We now turn to the proof of part (ii) of Proposition 8.7.5 and assume that all points of $(\partial \operatorname{supp} \rho) \cap I$ are shape regular for m and all estimates in Definition 8.7.3 hold true uniformly on this set. As in the proof of Proposition 8.7.4, we conclude $\sigma(\tau_0) \neq 0$ for all $\tau_0 \in (\partial \operatorname{supp} \rho) \cap I$. Owing to $\operatorname{dist}(0, \partial J) \gtrsim 1$ and the Hölder-continuity of σ on $(\partial \operatorname{supp} \rho) \cap I$, Proposition 8.7.4 is applicable to every $\tau_0 \in (\partial \operatorname{supp} \rho) \cap I$. Hence, (8.7.4a) and $\operatorname{dist}(0, \partial J) \gtrsim 1$ imply the existence of $\delta_1, c_1 \sim 1$ such that

$$\rho(\tau_0 + \omega) \geq c_1 |\omega|^{1/2} \tag{8.7.64}$$

for all $\omega \in -\operatorname{sign} \sigma(\tau_0)[0, \delta_1]$ and $\tau_0 \in (\partial \operatorname{supp} \rho) \cap I$. In particular, $\tau_0 - \operatorname{sign} \sigma(\tau_0)[0, \delta_1] \subset \operatorname{supp} \rho$ for all $\tau_0 \in (\partial \operatorname{supp} \rho) \cap I$. Since $|I| \lesssim 1$, this implies that $\operatorname{supp} \rho \cap I$ consists of finitely many intervals $[\alpha_i, \beta_i]$ with lengths $\gtrsim 1$, and, thus, their number K satisfies $K \sim 1$ as $\delta_1 \sim 1$ and $\beta_i - \alpha_i \geq \delta_1$ if $\beta_i \neq \sup I$ and $\alpha_i \neq \inf I$.

Additionally, we now assume that the elements of $\mathbb{M}_{\rho_*}$ are shape regular points for m on J and all estimates in Definition 8.7.3 hold true uniformly on $\mathbb{M}_{\rho_*}$. By possibly shrinking $\rho_* \sim 1$, we conclude from (8.7.64) that $|\alpha_i - \gamma| \sim 1$ and $|\beta_i - \gamma| \sim 1$ for any $i = 1, \ldots, K$ and $\gamma \in \mathbb{M}_{\rho_*}$.

Suppose now that $\tau_0 \in \mathbb{M}_{\rho_*}$ with $\rho(\tau_0) = 0$. Then part (i) and $\operatorname{dist}(0, \partial J) \gtrsim 1$ yield the existence of $\delta_2, c_2 \sim 1$ such that

$$\rho(\tau_0 + \omega) \geq c_2 |\omega|^{1/3}$$

for all $|\omega| \leq \delta_2$. By possibly further shrinking $\rho_* \sim 1$, we thus obtain $|\tau_0 - \gamma| \sim 1$ for all $\gamma \in \mathbb{M}_{\rho_*} \setminus \{\tau_0\}$. We thus conclude (8.7.10) in this case.

Finally, let $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$ with $\rho(\gamma_1), \rho(\gamma_2) > 0$. Then applying (i) with $\tau_0 = \gamma_1$ and $\tau_0 = \gamma_2$ yields

$$\Psi_1(\omega) + \Psi_2(\omega) \lesssim |\omega|^{1/3} \Big(\rho(\gamma_1) \mathbf{1}(|\omega| \lesssim \rho(\gamma_1)^3) + \rho(\gamma_2) \mathbf{1}(|\omega| \lesssim \rho(\gamma_2)^3) \Big) + \Psi_1(\omega)^2 + \Psi_2(\omega)^2,$$

where we defined $\omega = \gamma_2 - \gamma_1$ and

$$\Psi_1(\omega) := \widetilde{\rho}_1 \Psi_{\min}\Big(\frac{|\omega|}{\widetilde{\rho}_1^3} \Big), \qquad \Psi_2(\omega) := \widetilde{\rho}_2 \Psi_{\min}\Big(\frac{|\omega|}{\widetilde{\rho}_2^3} \Big)$$

with $\widetilde{\rho}_1 \sim \rho(\gamma_1)$ and $\widetilde{\rho}_2 \sim \rho(\gamma_2)$ (cf. Corollary 9.4 in [4]). Thus, we obtain either $|\omega| \sim 1$ or $|\omega| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$. This completes the proof of (8.7.10) and hence the one of Proposition 8.7.5. □

Finally, we use Proposition 8.7.5 and a Taylor expansion of ρ around a nonzero local minimum τ_0 to obtain the stronger conclusions of Theorem 8.7.2.

Proof of Theorem 8.7.2. We start with the proof of part (i). Let $\tau_0 \in \operatorname{supp} \rho \cap I_\theta$ satisfy the conditions of Theorem 8.7.2 (i). Then, by Proposition 8.6.1, the conditions of Proposition 8.7.5 (i) are fulfilled and all conclusions in Theorem 8.7.2 (i) apart from the case $|\omega| \leq \varepsilon \rho(\tau_0)^3$ in (8.7.5c) follow from Proposition 8.7.5 (i).

For the proof of the missing case, we fix a local minimum $\tau_0 \in \operatorname{supp} \rho \cap I_\theta$ of ρ such that $\rho(\tau_0) \leq \rho_*$. We set $\rho := \rho(\tau_0)$. Owing to the 1/3-Hölder continuity of ρ by Proposition 8.4.7, there is $\varepsilon \sim 1$ such that $\rho(\tau_0 + \omega) \sim \rho$ if $|\omega| \leq \varepsilon \rho^3$. In particular, $\rho(\tau_0 + \omega) > 0$ and using Lemma 8.5.7 with $k = 2, 3$ to compute the second order Taylor expansion of ρ around τ_0 yields

$$f_{\tau_0}(\omega) := \rho(\tau_0 + \omega) - \rho(\tau_0) = \frac{c}{\rho^5}\omega^2 + \mathcal{O}\Big(\frac{|\omega|^3}{\rho^8}\Big) \tag{8.7.65}$$

for all $\omega \in \mathbb{R}$ satisfying $|\omega| \le \varepsilon\rho^3$, where $c = c(\tau_0)$ satisfies $0 \le c \lesssim 1$.

On the other hand, τ_0 is a shape regular point by Proposition 8.6.1 and a nonzero local minimum of ρ. Hence, Proposition 8.7.5 (i) (d) implies

$$f_{\tau_0}(\omega) = \rho\Psi_{\min}\Big(\Gamma\frac{\omega}{\rho^3}\Big) + \mathcal{O}\Big(\frac{|\omega|}{\rho}\Big) = \frac{\Gamma^2}{18\rho^5}\omega^2 + \mathcal{O}\Big(\frac{|\omega|^3}{\rho^8} + \frac{|\omega|}{\rho}\Big) \tag{8.7.66}$$

for $|\omega| \le \varepsilon\rho^3$, where $\Gamma = \Gamma(\tau_0)$. Here, we also used the second order Taylor expansion of $\Psi_{\min}$ defined in (8.7.1b) in the second step. Note that $\Gamma \sim 1$ since $\psi + \sigma^2 \sim 1$ by (8.5.35) and $|\sigma| \lesssim \rho^2$ by Lemma 8.7.13.

We compare (8.7.65) and (8.7.66) and conclude

$$\frac{c}{\rho^5}\omega^2 = \frac{\Gamma^2}{18\rho^5}\omega^2 + \mathcal{O}\Big(\frac{|\omega|^3}{\rho^8} + \frac{|\omega|}{\rho}\Big)$$

for $|\omega| \le \varepsilon\rho^3$. Choosing $\omega = \rho^{7/2}$ and solving for c yield

$$c = \frac{\Gamma^2}{18} + \mathcal{O}(\rho^{1/2}). \tag{8.7.67}$$

By starting from the expansion of f_{τ_0} in (8.7.65), using the Taylor expansion of $\Psi_{\min}$ and (8.7.4b), we obtain (8.7.5c).

We now turn to the proof of (ii) of Theorem 8.7.2. By Proposition 8.6.1, the conditions of Proposition 8.7.5 (ii) are satisfied on $I' := I \cap [-3\kappa, 3\kappa]$, where $\kappa := \|a\| + 2\|S\|^{1/2}$. Since $\|a\| \lesssim 1$ and $\|S\| \le \|S\|_{2\to\|\cdot\|} \lesssim 1$ by Assumptions 8.4.5, we have $|I'| \lesssim 1$. Moreover, $\operatorname{supp}\rho \subset I'$ by (8.2.5a). Hence, by Proposition 8.7.5, it suffices to estimate the distance $|\gamma_1 - \gamma_2|$, where $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$ satisfy $\gamma_1 \ne \gamma_2$.

Let $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$. By (8.7.10) in Proposition 8.7.5 (ii), we know a dichotomy: either $|\gamma_1 - \gamma_2| \gtrsim 1$ or $|\gamma_1 - \gamma_2| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$. For $\gamma_1 \ne \gamma_2$, we now exclude the second case by using the expansions obtained in the proof of (i). If $\rho_* \sim 1$ is chosen sufficiently small then $c(\gamma_1) \sim 1$ and $c(\gamma_2) \sim 1$ by (8.7.67). Hence, by assuming $|\gamma_1 - \gamma_2| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$, we obtain $\rho(\gamma_2) > \rho(\gamma_1)$ from the expansion of $f_{\tau_0}(\omega)$ in (8.7.65) with $\tau_0 = \gamma_1$ and $\omega = \gamma_2 - \gamma_1$. Similarly, as $c(\gamma_2) \sim 1$, the expansion of $f_{\tau_0}(\omega)$ in (8.7.65) with

$\tau_0 = \gamma_2$ and $\omega = \gamma_1 - \gamma_2$ implies $\rho(\gamma_1) > \rho(\gamma_2)$. This is a contradiction. Therefore, the distance of two small local minima of ρ is much bigger than $\min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$ and the dichotomy above completes the proof of (ii). $\square$

8.7.6. Behaviour at a regular edge. We now list a few consequences of the previous results that will be used in the companion paper on the edge universality [17]. As in [17], in this subsection, we also assume that S is flat and a is bounded, i.e., that (8.3.10) is satisfied. In particular, owing to Proposition 8.2.3, there is a Hölder continuous probability density $\rho\colon \mathbb{R} \to [0, \infty)$ such that

$$\langle m(z) \rangle = \int_{\mathbb{R}} \frac{\rho(\tau)}{\tau - z} \, \mathrm{d}\tau,$$

where m is the solution to the Dyson equation, (8.2.3).

In this subsection, we study ρ and its harmonic extension to the complex upper half-plane in the vicinity of $\partial \operatorname{supp} \rho \subset \mathbb{R}$. We say that $\tau_0 \in \partial \operatorname{supp} \rho$ is a *regular edge* of ρ if there is $\varepsilon \sim 1$ such that $\rho(\tau) = 0$ for all $\tau \in [\tau_0 - \varepsilon, \tau_0]$ or $\tau \in [\tau_0, \tau_0 + \varepsilon]$. The following lemma characterizes regular edges and describes the behaviour of ρ close to them.

Lemma 8.7.16 (Behaviour of ρ close to a regular edge)**.** *Let a and S satisfy* (8.3.10) *and m be the solution of the corresponding Dyson equation,* (8.2.3)*. Suppose for some $\tau_0 \in \partial \operatorname{supp} \rho$, there are $m_* > 0$ and $\delta > 0$ such that*

$$\| m(\tau + \mathrm{i}\eta) \| \leq m_*$$

for all $\tau \in [\tau_0 - \delta, \tau_0 + \delta]$ and $\eta \in (0, \delta]$. Then the following implications hold true:

(i) If τ_0 is a regular edge then $|\sigma(\tau_0)| \sim 1$.

(ii) If $|\sigma| \sim 1$, $\sigma := \sigma(\tau_0)$, then τ_0 is a regular edge. Moreover, there is $\delta_* \sim 1$ such that

$$\rho(\tau_0 + \omega) = \begin{cases} \dfrac{\pi^{1/2}}{|\sigma|^{1/2}} |\omega|^{1/2} + \mathcal{O}(|\omega|), & \text{if } \operatorname{sign} \omega = \operatorname{sign} \sigma, \\ 0, & \text{if } \operatorname{sign} \omega = -\operatorname{sign} \sigma, \end{cases}$$

for all $\omega \in [-\delta_*, \delta_*]$.

In this lemma, the comparison relation $\sim$ is understood with respect to c_1, c_2, c_3 from (8.3.10) *as well as δ and m_*.*

Proof. For the entire proof, we remark that, by Lemma 8.4.8 (ii), the conditions of Proposition 8.6.1 are satisfied. Moreover, $\rho(\tau_0) = 0$ due to the continuity of ρ and $\tau_0 \in \partial \operatorname{supp} \rho$.

Thus, part (i) follows directly from Lemma 8.7.14 as τ_0 is a shape regular point by Proposition 8.6.1.

We now turn to the proof of (ii). We choose $\delta_* \sim 1$ as in Proposition 8.6.1. In particular, $\delta_* \leq \delta$. We take the imaginary part of (8.6.1) and apply $\langle \cdot \rangle$ to the result. This yields

$$\rho(\tau_0 + \omega) = \operatorname{Im}\Big(\Theta(\omega)\pi^{-1}\langle b\rangle\Big) + \pi^{-1}\langle \operatorname{Im} r(\omega)\rangle = \operatorname{Im}\Theta(\omega) + \mathcal{O}\Big((|\Theta(\omega)| + |\omega|)\operatorname{Im}\Theta(\omega)\Big)$$

for $|\omega| \leq \delta_*$. Here, we used $\langle b\rangle = \pi$ by (8.7.63) in the proof of Proposition 8.7.5 as well as the third bound in (8.6.8) in the second step.

By Proposition 8.6.1 the assumptions of Theorem 8.7.6 (ii) are satisfied with $\kappa = \pi$. Hence, we conclude (ii) of Lemma 8.7.16 from Lemma 8.7.12 by possibly shrinking $\delta_* \sim 1$ due to $|\sigma| \sim 1$, $|\Theta(\omega)| \lesssim |\omega/\sigma|^{1/2} \lesssim |\omega|^{1/2}$ and $|\nu(\omega)| \lesssim |\Theta(\omega)| + |\omega| \lesssim |\omega|^{1/2}$ by the first bound in (8.6.5). This completes the proof of Lemma 8.7.16. $\square$

The remainder of this section is devoted to understanding the harmonic extension of ρ to the complex upper half-plane. We denote this extension by $\rho(z)$ for $z \in \mathbb{H}$, i.e., $\rho(z) = \langle \operatorname{Im} m(z)\rangle/\pi$ for $z \in \mathbb{H}$.

The results of this subsection will hold true away from points, where m blows up, and away from almost cusp points. We now introduce these sets precisely. For a given $m_* > 0$, we define the set $P_m := P_m^{m_*} \subset \mathbb{H}$, where $\|m(z)\|$ is larger than m_*, i.e.,

$$P_m^{m_*} := \{\tau \in \mathbb{R} : \sup_{\eta > 0} \|m(\tau + \mathrm{i}\eta)\| > m_*\}. \tag{8.7.68}$$

For $\tau \in \mathbb{R} \setminus \operatorname{supp} \rho$, let $\Delta(\tau)$ denote the size of the largest interval that contains τ and is contained in $\mathbb{R} \setminus \operatorname{supp} \rho$. For $\rho_* > 0$ and $\Delta_* > 0$, we define the set $P_{\mathrm{cusp}} = P_{\mathrm{cusp}}^{\rho_*, \Delta_*} \subset \mathbb{R}$ of *almost cusp points* through

$$P_{\mathrm{cusp}}^{\rho_*,\Delta_*} := \{\tau \in \operatorname{supp}\rho \setminus \partial \operatorname{supp}\rho : \tau \text{ is a local minimum of } \rho,\ \rho(\tau) \le \rho_*\} \\ \cup \{\tau \in \mathbb{R} \setminus \operatorname{supp}\rho : \Delta(\tau) \le \Delta_*\}. \tag{8.7.69}$$

The set of points that are away from P_m and P_{cusp} is denoted by $\mathbb{D}$. More precisely, for some $\delta > 0$, we define

$$\mathbb{D} := \{z \in \mathbb{H} : \operatorname{dist}(z, P_m) \ge \delta,\ \operatorname{dist}(z, P_{\mathrm{cusp}}) \ge \delta\}. \tag{8.7.70}$$

In this subsection, the model parameters are c_1, c_2 and c_3 from (8.3.10) as well as m_*, ρ_*, Δ_* and δ from the definitions of P_m, P_{cusp} and $\mathbb{D}$, respectively.

In the next lemma, we establish the behaviour of $\rho(z)$ and $B(z)$ if z is close to a regular edge. Here, closeness means that $\kappa(z) + \operatorname{Im} z \sim \operatorname{dist}(z, \partial \operatorname{supp}\rho)$ is sufficiently small, where $z \in \mathbb{D}$ and $\kappa(z) := \operatorname{dist}(\operatorname{Re} z, \partial \operatorname{supp}\rho)$. By definition of $\mathbb{D}$, $\overline{\mathbb{D}} \cap \partial \operatorname{supp}\rho$ consists only of regular edges.

Lemma 8.7.17. *There is $\varepsilon_* \sim 1$ such that if $z \in \mathbb{D}$ satisfies* $\operatorname{dist}(z, \partial \operatorname{supp}\rho) \le \varepsilon_*$ *then*

(i) For the harmonic extension of the self-consistent density of states ρ, we have

$$\rho(z) \sim \begin{cases} \sqrt{\kappa(z) + \operatorname{Im} z}, & \text{if } \operatorname{Re} z \in \operatorname{supp}\rho, \\ \operatorname{Im} z / \sqrt{\kappa(z) + \operatorname{Im} z}, & \text{if } \operatorname{Re} z \notin \operatorname{supp}\rho. \end{cases} \tag{8.7.71a}$$

$$\rho(z) + \rho(z)^{-1}\operatorname{Im} z \sim \sqrt{\kappa(z) + \operatorname{Im} z}, \tag{8.7.71b}$$

(ii) Let l and b be defined as in Corollary 8.5.2. Setting $\mu_2 := \langle l\,, mS[b]b + bS[b]m\rangle/2$, we have

$$|\langle l\,, mS[b]b\rangle| \sim 1, \qquad |\mu_2(z)| \sim 1. \tag{8.7.72}$$

(iii) Let $B := \mathrm{Id} - C_m S$ and β be its eigenvalue of smallest modulus (cf. Corollary 8.5.2). We have

$$\|B^{-1}(z)\| + \|B^{-1}(z)\|_2 \lesssim (\kappa(z) + \operatorname{Im} z)^{-1/2}, \qquad |\beta(z)| \sim \sqrt{\kappa(z) + \operatorname{Im} z}. \tag{8.7.73}$$

Proof. By assumption, z is ε_*-close to a regular edge. Thus, owing to $\|m\| \lesssim 1$ by definition of $\mathbb{D}$, Theorem 8.7.1 (a), (b) and (8.7.4a) immediately imply (8.7.71a). Moreover, (8.7.71b) is a direct consequence of (8.7.71a).

For the proof of (ii), we shrink $\varepsilon_* \sim 1$ as well as use (8.7.71a), (8.7.71b) and $\operatorname{dist}(z, \partial \operatorname{supp} \rho) \sim \kappa + \operatorname{Im} z$ to guarantee that Lemma 8.5.1 and Corollary 8.5.2 are applicable. Furthermore, we use Lemma 8.7.14 and the definition of $\mathbb{D}$ to obtain $|\sigma(\tau_0)| \sim 1$, where $\tau_0 \in \partial \operatorname{supp} \rho$ is the point in $\partial \operatorname{supp} \rho$ closest to z. The Hölder-continuity of σ from Lemma 8.5.5 (i) implies $|\sigma(z)| \sim 1$ if ε_* is sufficiently small, i.e., z is sufficiently close to τ_0. Therefore, evaluating (8.6.24b) and (8.6.25) at z as well as using $|\sigma(z)| \sim 1$ yield $|\mu_2(z)| \sim 1$ and $|\langle l\,, mS[b]b\rangle| \sim 1$.

For the proof of (iii), we recall $|\sigma(z)| \sim 1$ from the proof of (ii). Therefore, (8.5.24) and (8.7.71b) yield the first bound in (8.7.73). Similarly, we obtain the second bound in (8.7.73) by using $|\sigma(z)| \sim 1$ and (8.7.71b) in (8.5.14c). This completes the proof of Lemma 8.7.17. □

8.8. Band mass formula – Proof of Proposition 8.2.6

Before proving Proposition 8.2.6, we state an auxiliary lemma which will be proven at the end of this section.

Lemma 8.8.1. *Let (a, S) be a data pair, m the solution of the associated Dyson equation,* (8.2.3)*, and ρ the corresponding self-consistent density of states. We assume $\|a\| \leq k_0$ and $S[x] \leq k_1 \langle x \rangle \mathbb{1}$ for all $x \in \overline{\mathcal{A}}_+$ and for some $k_0, k_1 > 0$. Then we have*

(i) If $\tau \in \mathbb{R} \setminus \operatorname{supp} \rho$ then there is $m(\tau) = m(\tau)^* \in \mathcal{A}$ such that

$$\lim_{\eta \downarrow 0} \|m(\tau + \mathrm{i}\eta) - m(\tau)\| = 0.$$

Moreover, $m(\tau)$ is invertible and satisfies the Dyson equation, (8.2.3)*, at $z = \tau$. There is $C > 0$, depending only on k_0, k_1 and $\operatorname{dist}(\tau, \operatorname{supp} \rho)$, such that $\|m(\tau)\| \leq C$ and $\|(\mathrm{Id} - (1-t)C_{m(\tau)}S)^{-1}\| \leq C$ all $t \in [0, 1]$.*

(ii) Fix $\tau \in \mathbb{R} \setminus \operatorname{supp} \rho$. Let m_t be the solution of (8.2.3) *associated to the data pair*

$$(a_t, S_t) := (a - tS[m(\tau)], (1-t)S)$$

for $t \in [0,1]$ *and* ρ_t *the corresponding self-consistent density of states. Then, for any* $t \in [0,1]$*, we have*

$$\lim_{\eta \downarrow 0} \| m_t(\tau + \mathrm{i}\eta) - m(\tau) \| = 0. \tag{8.8.1}$$

Moreover, there is $c > 0$ *such that* $\operatorname{dist}(\tau, \operatorname{supp} \rho_t) \geq c$ *for all* $t \in [0,1]$*.*

Proof of Proposition 8.2.6. We start with the proof of (i) and notice that the existence of $m(\tau)$ has been proven in Lemma 8.8.1 (i). In order to verify (8.2.10), we consider the continuous flow of data pairs (a_t, S_t) from Lemma 8.8.1 (ii) and the corresponding solutions m_t of the Dyson equation, (8.2.3), and prove

$$\rho_t((-\infty, \tau)) = \langle \mathbf{1}_{(-\infty,0)}(m_t(\tau)) \rangle \tag{8.8.2}$$

for all $t \in [0,1]$. Note that $\operatorname{dist}(\tau, \operatorname{supp} \rho_t) \geq c$ for all $t \in [0,1]$ by Lemma 8.8.1 (ii).

In particular, by Lemma 8.8.1 (ii), $m_t(\tau) = m(\tau)$ is constant along the flow, and with it the right-hand side of (8.8.2). The identity (8.8.2) obviously holds for $t = 1$, because $m_1(z) = (a - Sm(\tau) - z)^{-1}$ is the resolvent of a self-adjoint element and $m(\tau)$ satisfies (8.2.3) at $z = \tau$ by Lemma 8.8.1 (i). Thus it remains to verify that the left-hand side of (8.8.2) stays constant along the flow as well. This will show (8.8.2) for $t = 0$ which is (8.2.10).

First we conclude from the Stieltjes transform representation (8.2.4) of m_t that

$$\rho_t((-\infty, \tau)) \; = \; -\frac{1}{2\pi \mathrm{i}} \oint \langle m_t(z) \rangle \, \mathrm{d}z \,, \tag{8.8.3}$$

where the contour encircles $[\min \operatorname{supp} \rho_t, \tau)$ counterclockwise, passing through the real line only at τ and to the left of $\min \operatorname{supp} \rho_t$, and we extended $m_t(z)$ analytically to a neighbourhood of the contour (set $m_t(\bar{z}) := m_t(z)^*$ for $z \in \mathbb{H}$ and use Lemma 8.14.1 (iv) close to the real axis to conclude analyticity in a neighbourhood of the contour).

We now show that the left-hand side of (8.8.3) does not change along the flow. Indeed, differentiating the right-hand side of (8.8.3) with respect to t and writing $m_t = m_t(z)$

yield

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\oint \langle m_t(z)\rangle \mathrm{d}z &= \oint \langle \partial_t m_t(z)\rangle \mathrm{d}z \\ &= \oint \langle (C_{m_t^*}^{-1} - S_t)^{-1}[\mathbb{1}], S[m(\tau)] - S[m_t]\rangle \mathrm{d}z \\ &= \oint \langle (\partial_z m_t)(S[m(\tau)] - S[m_t])\rangle \mathrm{d}z \\ &= \oint \partial_z \Big(\langle m_t S[m(\tau)]\rangle - \frac{1}{2}\langle m_t S[m_t]\rangle\Big) \mathrm{d}z \\ &= 0.\end{aligned}$$

Here, in the second step, we used $\partial_t m_t(z) = (C_{m_t}^{-1} - S_t)^{-1}[-S[m_t] - S[m(\tau)]]$ obtained by differentiating the Dyson equation, (8.2.3), for the data pair (a_t, S_t) defined in Lemma 8.8.1 (ii) and the definition of the scalar product, (8.2.1). In the third step, we employed $(C_{m_t^*}^{-1} - S_t)^{-1}[\mathbb{1}] = (\partial_z m_t(z))^*$ which follows from differentiating the Dyson equation, (8.2.3), for the data pair (a_t, S_t) with respect to z. Finally, we used that m_t is holomorphic in a neighbourhood of the contour. This completes the proof of (i) of Proposition 8.2.6.

For the proof of (ii), we fix a connected component J of $\operatorname{supp}\rho$. Let $\tau_1, \tau_2 \in \mathbb{R}\setminus\operatorname{supp}\rho$ satisfy $\tau_1 < \tau_2$ and $[\tau_1, \tau_2] \cap \operatorname{supp}\rho = J$. By (8.2.10), we have

$$n\rho(J) = n\Big(\rho((-\infty, \tau_2)) - \rho((-\infty, \tau_1))\Big) = \operatorname{Tr}(P_2) - \operatorname{Tr}(P_1) = \operatorname{rank}P_2 - \operatorname{rank}P_1,$$

where $P_i := \pi(\mathbf{1}_{(-\infty,0)}(m(\tau_i)))$ are orthogonal projections in $\mathbb{C}^{n\times n}$ for $i = 1, 2$. Hence, $n\rho(J) \in \mathbb{Z}$. Since $0 < n\rho(J) \le n$ by definition of $\operatorname{supp}\rho$, we conclude $n\rho(J) \in \{1, \dots, n\}$, which immediately implies that $\operatorname{supp}\rho$ has at most n connected components. This completes the proof of Proposition 8.2.6. □

Proof of Lemma 8.8.1. In part (i), the existence of the limit $m(\tau) \in \mathcal{A}$ follows immediately from the implication (v) $\Rightarrow$ (iii) of Lemma 8.14.1. The invertibility of $m(\tau)$ can be seen by multiplying (8.2.3) at $z = \tau + \mathrm{i}\eta$ by $m(\tau + \mathrm{i}\eta)$ and taking the limit $\eta \downarrow 0$. This also implies that $m(\tau)$ satisfies (8.2.3) at $z = \tau$. In order to bound $\|(\mathrm{Id} - (1 - t)C_{m(\tau)}S)^{-1}\|$, we recall the definitions of q, u and F from (8.3.1) and (8.3.4), respectively, and compute

$$\mathrm{Id} - (1-t)C_m S = C_{q^*,q}(\mathrm{Id} - (1-t)C_u F)C_{q^*,q}^{-1}$$

for $m = m(z)$ with $z \in \mathbb{H}$. Hence, by (8.14.1), Lemma 8.4.8 (i) and Lemma 8.12.2, we obtain $\|(\mathrm{Id} - (1-t)C_m S)^{-1}\| \lesssim (1-(1-t)\|F\|_2)^{-1} \leq (1-\|F\|_2)^{-1} \leq C$ for all $z \in \tau + \mathrm{i}N$, where the set $N \subset (0,1]$ with an accumulation point at 0 is given in Lemma 8.14.1 (ii). Taking the limit $\eta \downarrow 0$ under the constraint $\eta \in N$ and possibly increasing C yield the desired uniform bound. This completes the proof of (i).

We start the proof of (ii) with an auxiliary result. Similarly as in the proof of (i), we see that $\mathrm{Id} - (1-t)C_{m^*,m}S$ is invertible for $m = m(z)$, $z \in \tau + \mathrm{i}N$ with N as before. Since $\|F(z)\|_2 \leq 1 - C^{-1}$ for $z \in \tau + \mathrm{i}N$ by Lemma 8.14.1 (ii), Lemma 8.12.3 implies that $(\mathrm{Id} - (1-t)C_{u^*,u}F)^{-1}$, $F = F(z)$, and, thus, $(\mathrm{Id} - (1-t)C_{m^*,m}S)^{-1} = C_{q^*,q}(\mathrm{Id} - (1-t)C_{u^*,u}F)^{-1}C_{q^*,q}^{-1}$ are positivity-preserving for $z \in \tau + \mathrm{i}N$. Taking the limit $\eta = \operatorname{Im} z \downarrow 0$ in N shows that $(\mathrm{Id} - (1-t)C_{m(\tau)}S)^{-1}$ is positivity-preserving for any $t \in [0,1]$. Moreover, (8.12.10) with $x = \mathbb{1}$ yields

$$(\mathrm{Id} - (1-t)C_{m^*,m}S)^{-1}[\mathbb{1}] = C_{q^*,q}(\mathrm{Id} - (1-t)C_{u^*,u}F)^{-1}C_{q^*,q}^{-1}[\mathbb{1}] \geq \mathbb{1}. \tag{8.8.4}$$

Since (8.8.4) holds true uniformly for $z \in \tau + \mathrm{i}N$ and $t \in [0,1]$, taking the limit $\eta = \operatorname{Im} z \downarrow 0$ in N, we obtain

$$(\mathrm{Id} - (1-t)C_{m(\tau)}S)^{-1}[\mathbb{1}] \geq \mathbb{1} \tag{8.8.5}$$

for all $t \in [0,1]$.

We fix $t \in [0,1]$. We write $m = m(\tau)$ and define $\Phi_t \colon \mathcal{A} \times \mathbb{R} \to \mathcal{A}$ through

$$\Phi_t(\Delta, \eta) := (\mathrm{Id} - (1-t)C_m S)[\Delta] - \frac{\mathrm{i}\eta}{2}(m\Delta + \Delta m) - \mathrm{i}\eta m^2 - \frac{1}{2}(1-t)(\Delta S[\Delta]m + mS[\Delta]\Delta)$$

In order to show (8.8.1), we apply the implicit function theorem (see e.g. Lemma 8.14.4 below) to $\Phi_t(\Delta, \eta) = 0$. It is applicable as $\Phi_t(0,0) = 0$ and $\partial_1 \Phi_t(0,0) = \mathrm{Id} - (1-t)C_m S$ which is invertible by (i). Hence, we obtain an $\varepsilon > 0$ and a continuously differentiable function $\Delta_t \colon (-\varepsilon, \varepsilon) \to \mathcal{A}$ such that $\Phi_t(\Delta_t(\eta), \eta) = 0$ for all $\eta \in (-\varepsilon, \varepsilon)$ and $\Delta_t(0) = 0$. We now show that $\Delta_t(\eta) + m(\tau) = m_t(\tau + \mathrm{i}\eta)$ for all sufficiently small $\eta > 0$ by appealing to the uniqueness of the solution to the Dyson equation, (8.2.3), with the choice $z = \tau + \mathrm{i}\eta$, $a = a_t$ and $S = S_t = (1-t)S$. In fact, $m = m(\tau)$ and $m_t = m_t(\tau + \mathrm{i}\eta)$ with $\eta > 0$ satisfy the Dyson equations

$$-m^{-1} = \tau - a + S[m], \quad -m_t^{-1} = \tau + \mathrm{i}\eta - a + tS[m] + (1-t)S[m_t] \tag{8.8.6}$$

and m_t is the unique solution of the second equation under the constraint $\operatorname{Im} m_t > 0$ (compare the remarks around (8.2.3)). A straightforward computation using the first relation in (8.8.6) and $\Phi_t(\Delta_t(\eta), \eta) = 0$ reveals that $\Delta_t(\eta) + m(\tau)$ solves the second equation in (8.8.6) for m_t. Moreover, differentiating $\Phi_t(\Delta_t(\eta), \eta) = 0$ with respect to η at $\eta = 0$ yields

$$\begin{aligned}\partial_\eta \operatorname{Im} \Delta_t(\eta = 0) &= (\mathrm{Id} - (1-t)C_m S)^{-1}[m^2] \\ &\geq \|m^{-1}\|^{-2}(\mathrm{Id} - (1-t)C_m S)^{-1}[\mathbb{1}] \geq \|m^{-1}\|^{-2}\mathbb{1}.\end{aligned}$$

Here, we used that $(\mathrm{Id} - (1-t)C_m S)^{-1}$ is compatible with the involution * and $m = m^*$ in the first step. Then we employed the invertibility of m, $m^2 \geq \|m^{-1}\|^{-2}\mathbb{1}$ and the positivity-preserving property of $(\mathrm{Id} - (1-t)C_m S)^{-1}$ in the second step and, finally, (8.8.5) in the last step. Hence, $\operatorname{Im}(\Delta_t(\eta) + m(\tau)) = \operatorname{Im} \Delta_t(\eta) > 0$ for all sufficiently small $\eta > 0$. The uniqueness of the solution to the Dyson equation for m_t, the second relation in (8.8.6), implies $\Delta_t(\eta) + m(\tau) = m_t(\tau + \mathrm{i}\eta)$ for all sufficiently small $\eta > 0$ and all $t \in [0, 1]$. Therefore, the continuity of Δ_t as a function of η, $\Delta_t(\eta) \to \Delta_t(0) = 0$, yields (8.8.1).

We now conclude from the implication (iii) $\Rightarrow$ (v) of Lemma 8.14.1 that $\operatorname{dist}(\tau, \operatorname{supp} \rho_t) \geq \varepsilon$ for some $\varepsilon > 0$. Lemma 8.14.1 is applicable since $\|a_t\| \leq k_0 + k_1 C$ (cf. Lemma 8.12.2 (i) and Lemma 8.8.1 (i)) and $S_t[x] \leq S[x] \leq k_1\langle x\rangle\mathbb{1}$ for all $x \in \overline{\mathcal{A}}_+$. For any $t \in [0, 1]$, statement (iii) in Lemma 8.14.1 holds true with the same $m = m(\tau)$ by (8.8.1) and S replaced by $S_t = (1-t)S$. By (i), $\|m\| \leq C$ and $\|(\mathrm{Id} - (1-t)C_m S)^{-1}\| \leq C$ for all $t \in [0, 1]$. Hence, owing to Lemma 8.14.1 (v), there is $\varepsilon > 0$ such that $\operatorname{dist}(\tau, \operatorname{supp} \rho_t) \geq \varepsilon$ for all $t \in [0, 1]$. The uniformity of ε in t is a consequence of the effective dependence of the constants in Lemma 8.14.1 on each other (see final remark in Lemma 8.14.1) and the uniform upper bound on $\|(\mathrm{Id} - (1-t)C_m S)^{-1}\|$. This completes the proof of Lemma 8.8.1. $\square$

8.9. Dyson equation for Kronecker random matrices

In this section we present an application of the theory presented in this work to Kronecker random matrices, i.e., block correlated random matrices with variance profiles within the blocks, and their limits. In particular, in Lemma 8.9.1 and Lemma 8.9.3 below, we will provide some sufficient checkable conditions that ensure the flatness of S and the boundedness of $\|m(z)\|$, the main assumptions of Proposition 8.2.4, Theorem 8.2.5 and Theorem 8.7.1, for the self-consistent density of states of Kronecker random matrices introduced in Chapter 7.

8.9.1. The Kronecker setup. We fix $K \in \mathbb{N}$ and a probability space $(\mathfrak{X}, \pi)$ that we view as a possibly infinite set of indices. We consider the von Neumann algebra

$$\mathcal{A} = \mathbb{C}^{K\times K} \otimes \mathrm{L}^\infty(\mathfrak{X}), \tag{8.9.1}$$

with the tracial state

$$\langle \kappa \otimes f \rangle = \frac{\operatorname{Tr} \kappa}{K} \int_{\mathfrak{X}} f \mathrm{d}\pi .$$

For $K = 1$ the algebra $\mathcal{A}$ is commutative and this setup was previously considered in [4, 5]. Now let $(\alpha_\mu)_{\mu=1}^{\ell_1}, (\beta_\nu)_{\nu=1}^{\ell_2}$ be families of matrices in $\mathbb{C}^{K\times K}$ with $\alpha_\mu = \alpha_\mu^*$ self-adjoint and let $(s^\mu)_{\mu=1}^{\ell_1}, (t^\nu)_{\nu=1}^{\ell_2}$ be families of non-negative bounded functions in $\mathrm{L}^\infty(\mathfrak{X}^2)$ and suppose that all s^μ are symmetric, $s^\mu(x,y) = s^\mu(y,x)$. Then we define the self-energy operator $S : \mathcal{A} \to \mathcal{A}$ as

$$S(\kappa \otimes f) := \sum_{\mu=1}^{\ell_1} \alpha_\mu \kappa \alpha_\mu \otimes S_\mu f + \sum_{\nu=1}^{\ell_2} (\beta_\nu \kappa \beta_\nu^* \otimes T_\nu f + \beta_\nu^* \kappa \beta_\nu \otimes T_\nu^* f) , \tag{8.9.2}$$

where the bounded operators $S_\mu, T_\nu, T_\nu^* : \mathrm{L}^\infty(\mathfrak{X}) \to \mathrm{L}^\infty(\mathfrak{X})$ act as

$$(S_\mu f)(x) = \int_{\mathfrak{X}} s^\mu(x,y) f(y) \pi(\mathrm{d}y) ,$$

$$(T_\nu f)(x) = \int_{\mathfrak{X}} t^\nu(x,y) f(y) \pi(\mathrm{d}y) , \qquad (T_\nu^* f)(x) = \int_{\mathfrak{X}} t^\nu(y,x) f(y) \pi(\mathrm{d}y) .$$

Furthermore we fix a self-adjoint $a = a^* \in \mathcal{A}$. With these data we will consider the Dyson equation, (8.2.3).

The following lemma provides sufficient conditions that ensure flatness of S and boundedness of $\|m(z)\|$ uniformly in z up to the real line. We begin with some preparations. We use the notation $x \mapsto v_x$ for $x \in \mathfrak{X}$ and an element $v \in \mathbb{C}^{K\times K} \otimes \mathrm{L}^\infty(\mathfrak{X})$, interpreting it as a function on $\mathfrak{X}$ with values in $\mathbb{C}^{K\times K}$. We also introduce the functions $\gamma \in \mathrm{L}^\infty(\mathfrak{X}^2)$ via

$$\gamma(x,y) := \Big(\int_{\mathfrak{X}} (|s^\mu(x,\cdot) - s^\mu(y,\cdot)|^2 + |t^\nu(x,\cdot) - t^\nu(y,\cdot)|^2 + |t^\nu(\cdot,x) - t^\nu(\cdot,y)|^2)\mathrm{d}\pi\Big)^{1/2} \tag{8.9.3}$$

and $\Gamma : (0,\infty)^2 \to \mathrm{L}^\infty(\mathfrak{X}), (\Lambda,\tau) \mapsto \Gamma_{\Lambda,\cdot}(\tau)$ through

$$\Gamma_{\Lambda,x}(\tau) := \bigg(\int_{\mathfrak{X}} \Big(\frac{1}{\tau} + \|a_x - a_y\| + \gamma(x,y)\Lambda\Big)^{-2} \pi(\mathrm{d}y)\bigg)^{1/2}. \tag{8.9.4}$$

Here, we denoted by $\|\cdot\|$ the operator norm on $\mathbb{C}^{K\times K}$ induced by the Euclidean norm on $\mathbb{C}^K$. The two functions γ and Γ will be important to quantify the modulus of continuity of the data (a, S).

Lemma 8.9.1. *Let m be the solution of the Dyson equation,* (8.2.3)*, on the von Neumann algebra $\mathcal{A}$ from* (8.9.1) *associated to the data (a,S) with S defined as in* (8.9.2)*.*

(i) Define $\Gamma(\tau) := C_{\mathrm{Kr}} \operatorname{ess\,inf}_x \Gamma_{1,x}(\tau)$ with $C_{\mathrm{Kr}} := 4 + 4K(\ell_1+\ell_2)\max_{\mu,\nu}(\|\alpha_\mu\|^2 + \|\beta_\nu\|^2)^{1/2}$, where $\Gamma_{\Lambda,x}(\tau)$ was introduced in (8.9.4) *and assume that for some $z \in \mathbb{H}$ the L^2-upper bound $\|m(z)\|_2 \le \Lambda$ for some $\Lambda \ge 1$ is satisfied. Then we have the uniform upper bound*

$$\|m(z)\| \le \frac{\Gamma^{-1}(\Lambda^2)}{\Lambda}, \tag{8.9.5}$$

where we interpret the right-hand side as ∞ if Λ is not in the range of the strictly monotonously increasing function Γ.

(ii) Suppose that the kernels of the operators S^μ and T^ν, used to define S in (8.9.2)*, are bounded from below, i.e., $\operatorname{ess\,inf}_{x,y} s^\mu(x,y) > 0$ and $\operatorname{ess\,inf}_{x,y} t^\nu(x,y) > 0$. Suppose further that*

$$\inf_\kappa \frac{1}{\operatorname{Tr}\kappa}\Big(\sum_{\mu=1}^{\ell_1} \alpha_\mu\kappa\alpha_\mu + \sum_{\nu=1}^{\ell_2}(\beta_\nu\kappa\beta_\nu^* + \beta_\nu^*\kappa\beta_\nu)\Big) > 0, \tag{8.9.6}$$

where the infimum is taken over all positive definite $\kappa \in \mathbb{C}^{K\times K}$. *Then* S *is flat, i.e.,* $S \in \Sigma_{\mathrm{flat}}$ *(cf.* (8.2.2b)*).*

(iii) Let S *be flat, hence,* $\Lambda := 1 + \sup_{z\in\mathbb{H}}\|m(z)\|_2 < \infty$. *Then* (8.9.5) *holds true with this* Λ.

(iv) If $a = 0$ *then, for each* $\varepsilon > 0$, (8.9.5) *holds true on* $|z| \geq \varepsilon$ *with* $\Lambda := 1 + 2\varepsilon^{-1}$.

Proof of Lemma 8.9.1. We adapt the proof of Proposition 6.6 in [4] to our non-commutative setting in order to prove (i). Recall the definition of $\gamma(x, y)$ in (8.9.3). Estimating the norm $\|m\|_2$ from below, we find

$$\begin{aligned}\|m\|_2^2 = \frac{1}{K}\operatorname{Tr}\int \frac{\pi(\mathrm{d}y)}{m_y^{-1}(m_y^*)^{-1}} &\geq \operatorname{Tr}\int_{\mathfrak{X}} \frac{C_{\mathrm{Kr}}^2\,\pi(\mathrm{d}y)}{m_x^{-1}(m_x^*)^{-1} + \|a_x - a_y\|^2 + \gamma(x,y)^2\|m\|_2^2} \\ &\geq C_{\mathrm{Kr}}^2\Big(\Gamma_{\|m\|_2,x}(\|m_x\|)\Big)^2,\end{aligned} \tag{8.9.7}$$

for π-almost all $x \in \mathfrak{X}$, where we used

$$\begin{aligned}\frac{1}{4}m_y^{-1}(m_y^*)^{-1} &\leq m_x^{-1}(m_x^*)^{-1} + (a_y - a_x)(a_y - a_x)^* + ((Sm)_x - (Sm)_y)((Sm)_x - (Sm)_y)^* \\ &\leq m_x^{-1}(m_x^*)^{-1} + \|a_x - a_y\|^2 + K(\ell_1 + \ell_2)\max_{\mu,\nu}(\|\alpha_\mu\|^2 + \|\beta_\nu\|^2)\gamma(x,y)^2\|m\|_2^2\,.\end{aligned} \tag{8.9.8}$$

We conclude $\Lambda \geq \Lambda^{-1}\Gamma(\Lambda\|m_x\|)$ for any upper bound $\Lambda \geq 1$ on $\|m\|_2$. In particular, (8.9.5) follows.

We turn to the proof of (ii). We view a positive element $r \in \mathcal{A}_+$ as a function $r : [0, 1] \to \mathbb{C}^{K\times K}$ with values in positive semidefinite matrices. Then we find

$$(Sr)_x \;\geq\; c\int_{\mathfrak{X}}\Big(\sum_{\mu=1}^{\ell_1}\alpha_\mu r_y \alpha_\mu + \sum_{\nu=1}^{\ell_2}(\beta_\nu r_y \beta_\nu^* + \beta_\nu^* r_y \beta_\nu)\Big)\pi(\mathrm{d}y)\,,$$

as quadratic forms on $\mathbb{C}^{K\times K}$ for almost every $x \in \mathfrak{X}$. The claim follows now immediately from (8.9.6). Part (iii) is a direct consequence of (i) and (ii) as well as (8.3.11). For the proof of part (iv), we use part (i) and (8.2.6) if $a = 0$. □

8.9.2. $N\times N$**-Kronecker random matrices.** As an application of the general Kronecker setup introduced above, we consider the *matrix Dyson equation* associated to Kronecker random matrices. Let $X_\mu, Y_\nu \in \mathbb{C}^{N\times N}$ be independent centered random matrices such that $Y_\nu = (y_{ij}^\nu)$ has independent entries and $X_\mu = (x_{ij}^\mu)$ has independent entries up to the Hermitian symmetry constraint $X_\mu = X_\mu^*$. Suppose that the entries of $\sqrt{N}X_\mu, \sqrt{N}Y_\nu$ have uniformly bounded moments, $\mathbb{E}(|x_{ij}^\mu|^p + |y_{ij}^\mu|^p) \leq N^{-p/2}C_p$ and define their variance profiles through

$$s^\mu(i,j) := N\mathbb{E}|x_{ij}^\mu|^2\,, \qquad t^\nu(i,j) := N\mathbb{E}|y_{ij}^\nu|^2\,.$$

Then we are interested in the asymptotic spectral properties of the Hermitian *Kronecker random matrix*

$$H := A + \sum_{\mu=1}^{\ell_1} \alpha_\mu \otimes X_\mu + \sum_{\nu=1}^{\ell_2} (\beta_\nu \otimes Y_\nu + \beta_\nu^* \otimes Y_\nu^*) \in \mathbb{C}^{K\times K} \otimes \mathbb{C}^{N\times N}\,, \tag{8.9.9}$$

as $N \to \infty$. Here the expectation matrix A is assumed to be bounded, $\|A\| \leq C$, and block diagonal, i.e.

$$A = \sum_{i=1}^{N} a_i \otimes E_{ii}\,, \tag{8.9.10}$$

with $E_{ii} = (\delta_{il}\delta_{ik})_{l,k=1}^N \in \mathbb{C}^{N\times N}$ and $a_i \in \mathbb{C}^{K\times K}$. In Chapter 7 it was shown that the resolvent $G(z) = (H-z)^{-1}$ of the Kronecker matrix H is well approximated by the solution $M(z)$ of a Dyson equation of Kronecker type, i.e., on the von Neumann algebra $\mathcal{A}$ in (8.9.1) with self-energy S from (8.9.2) and $a = A \in \mathcal{A}$, when we choose $\mathfrak{X} = \{1,\ldots,N\}$ and π the uniform probability distribution. In other words, $\mathrm{L}^\infty(\mathfrak{X}) = \mathbb{C}^N$ with entrywise multiplication.

8.9.3. Limits of Kronecker random matrices. Now we consider limits of Kronecker random matrices $H \in \mathbb{C}^{N\times N}$ with piecewise Hölder-continuous variance profiles as $N \to \infty$. In this situation we can make sense of the continuum limit for the solution $M(z)$ of the associated matrix Dyson equation. The natural setup here is $(\mathfrak{X},\pi) = ([0,1], \mathrm{d}x)$. We fix a partition $(I_l)_{l=1}^L$ of $[0,1]$ into intervals of positive length, i.e., $[0,1] = \dot\cup_l I_l$ and consider non-negative profile functions $s^\mu, t^\nu : [0,1]^2 \to \mathbb{R}$ that are Hölder-continuous with Hölder exponent $1/2$ on each rectangle $I_l \times I_k$. We also fix a function $a : [0,1] \to \mathbb{C}^{K\times K}$ that is $1/2$-Hölder continuous on each I_l. In this piecewise Hölder-continuous setup the Dyson equation on $\mathcal{A}$ with data pair (a, S) describes the asymptotic spectral properties of Kronecker random matrices with fixed variance profiles s^μ and t^ν, i.e., the random matrices H introduced in Subsection 8.9.2 if their variances are given by

$$\mathbb{E}|x_{ij}^{\mu}|^2 = \frac{1}{N}s^{\mu}\Big(\frac{i}{N},\frac{j}{N}\Big), \qquad \mathbb{E}|y_{ij}^{\nu}|^2 = \frac{1}{N}t^{\nu}\Big(\frac{i}{N},\frac{j}{N}\Big),$$

and the matrices a_i in (8.9.10) by $a_i = a(\frac{i}{N})$.

Lemma 8.9.2. *Suppose that a, s^{μ} and t^{ν} are piecewise Hölder-continuous with Hölder exponent 1/2 as described above. The empirical spectral distribution of the Kronecker random matrix H, defined in (8.9.9), with eigenvalues $(\lambda_i)_{i=1}^{KN}$ converges weakly in probability to the self-consistent density of states ρ associated to the Dyson equation with data pair (a, S) as defined in (8.9.2), i.e., for any $\varepsilon > 0$ and $\varphi \in C(\mathbb{R})$ we have*

$$\mathbb{P}\Big(\Big|\frac{1}{KN}\sum_{i=1}^{KN}\varphi(\lambda_i) - \int_{\mathbb{R}}\varphi\,\mathrm{d}\rho\Big| > \varepsilon\Big) \to 0, \qquad N \to \infty.$$

Proof of Lemma 8.9.2. It suffices to prove convergence of the Stieltjes transforms, i.e., in probability $\frac{1}{NK}\operatorname{Tr}_{KN} G(z) \to \langle m(z)\rangle$ for every fixed $z \in \mathbb{H}$, where $G(z) = (H-z)^{-1}$ is the resolvent of the Kronecker matrix H and $m(z)$ is the solution to the Dyson equation with data (a, S).

First we use the Theorem 7.2.7 from Chapter 7 to show that

$$\frac{1}{KN}\operatorname{Tr}_{KN} G(z) - \frac{1}{N}\sum_{i=1}^{N}\operatorname{Tr}_K m_i(z) \to 0$$

in probability, where $M_N = (m_1, \dots, m_N) \in (\mathbb{C}^{K\times K})^N$ denotes the solution to a Dyson equation formulated on the von Neumann algebra $\mathbb{C}^{K\times K}\otimes\mathbb{C}^N$ with entrywise multiplication on vectors in $\mathbb{C}^N$ as explained in Subsection 8.9.2. We recall that in this setup the discrete kernels for S_μ and T_ν from the definition of S in (8.9.2) are given by $N\mathbb{E}|x_{ij}^{\mu}|^2$ and $N\mathbb{E}|y_{ij}^{\nu}|^2$, respectively, and $a = \sum_{i=1}^N a(\frac{i}{N})\otimes e_i$. To distinguish this discrete data pair from the continuum limit over $\mathbb{C}^{K\times K}\otimes \mathrm{L}^\infty[0,1]$, we denote it by (a_N, S_N). Note that in Theorem 7.2.7 of Chapter 7 the test functions were compactly supported in contrast to the function $\tau \mapsto 1/(\tau - z)$ that we used here. However, by Theorem 7.2.4 of Chapter 7 and since the self-consistent density of states is compactly supported (cf. (8.2.5a) and $\|S\| \lesssim 1$) no eigenvalues can be found beyond a certain bounded interval, ensuring that non compactly supported test function are allowed as well.

Now it remains to show that $\langle M_N \rangle \to \langle m \rangle$ as $N \to \infty$ for all $z \in \mathbb{H}$. For this purpose we embed $\mathbb{C}^N$ into $\mathrm{L}^\infty[0,1]$ via $Pv := \sum_{i=1}^N v_i \mathbf{1}_{[(i-1)/N, i/N)}$. With this identification M_N and m satisfy Dyson equations on the same space $\mathbb{C}^{K\times K} \otimes \mathrm{L}^\infty[0,1]$. Evaluating these two equations at $z + \mathrm{i}\eta$, for a fixed $z \in \mathbb{H}$ and any $\eta \geq 0$, and subtracting them from each other yield

$$\begin{aligned} B[\Delta] &= m(S_N - S)[m]\Delta + C_m(S_N - S)[\Delta] + mS_N[\Delta]\Delta \\ &\qquad + C_m(S_N - S)[m] - m(a_N - a)\Delta - C_m[a_N - a], \end{aligned}$$

where $m = m(z + \mathrm{i}\eta)$, $M_N = M_N(z + \mathrm{i}\eta)$, $B = \mathrm{Id} - C_m S$ and $\Delta = M_N - m$. Using the imaginary part of z we have $\mathrm{dist}(z+\mathrm{i}\eta, \mathrm{supp}\,\rho) \geq \mathrm{Im}\, z > 0$. By (7.3.22), (7.3.23), (7.3.11a) and (7.3.11c) in Chapter 7 we infer $\|m\| + \|B^{-1}\|_2 \leq C$ for all $\eta \geq 0$ with a constant C depending on $\mathrm{Im}\, z$. Note that although the proofs in Chapter 7 were performed on $\mathbb{C}^{N\times N}$ all estimates were uniform in N and all algebraic relations in these proof translate to the current setting on a finite von Neumann algebra. Using $\|S_N - S\|_2 \leq \|S_N - S\|$ as well as $\|S_N\| \leq C$ and possibly increasing C, we thus obtain

$$\|\Delta\|_2 \leq C(\Psi_N + \|\Delta\|_2^2), \qquad\qquad \Psi_N := \|a_N - a\| + \|S_N - S\|,$$

where $\Delta = \Delta(z+\mathrm{i}\eta)$, for all $\eta \geq 0$. We choose N_0 sufficiently large such that $2\Psi_N C^2 \leq 1/4$ for all $N \geq N_0$ and define $\eta_* := \sup\{\eta \geq 0\colon\ \|\Delta(z+\mathrm{i}\eta)\|_2 \geq 2C\Psi_N\}$. Since $\|M_N\| + \|m\| \to 0$ for $\eta \to \infty$, we conclude $\eta_* < \infty$.

We now prove $\eta_* = 0$. For a proof by contradiction, we suppose $\eta_* > 0$. Then, by continuity, $\|\Delta(\tau + \mathrm{i}\eta_*)\|_2 = 2C\Psi_N$. Since $2\Psi_N C^2 \leq 1/4$, we have $\|\Delta(z + \mathrm{i}\eta_*)\|_2 \leq 4C\Psi_N/3 < 2C\Psi_N = \|\Delta(z + \mathrm{i}\eta_*)\|_2$. From this contradiction, we conclude $\eta_* = 0$. Therefore, for $N \geq N_0$, we have

$$|M_N(z) - m(z)| \leq \|\Delta(z)\|_2 \leq 2C\Psi_N = 2C(\|S_N - S\| + \|a_N - a\|)\,.$$

Since the right-hand side converges to zero as $N \to \infty$, due to the piecewise Hölder-continuity of the profile functions, and since z was arbitrary, we obtain $\langle M_N \rangle \to \langle m \rangle$ as $N \to \infty$ for all $z \in \mathbb{H}$. This completes the proof of Lemma 8.9.2. □

The boundedness of the solution to the Dyson equation in L^2-norm already implies uniform boundedness in the piecewise Hölder-continuous setup.

Lemma 8.9.3. *Suppose that a, s^μ and t^ν are piecewise $1/2$-Hölder continuous and that $\sup_{z\in\mathbb{D}}\|m(z)\|_2 < \infty$ for some domain $\mathbb{D} \subseteq \mathbb{H}$. Then we have the uniform bound $\sup_{z\in\mathbb{D}}\|m(z)\| < \infty$.*

In particular, if the random matrix H is centered, i.e., $a = 0$, then $m(z)$ is uniformly bounded as long as z is bounded away from zero; and if H is flat in the limit, i.e., S is flat, then $\sup_{z\in\mathbb{H}}\|m(z)\| < \infty$.

PROOF. By (i) of Lemma 8.9.1 the proof reduces to checking that $\lim_{\tau\to\infty}\Gamma(\tau) = \infty$ for piecewise $1/2$-Hölder continuous data in the special case $(\mathfrak{X}, \pi) = ([0,1], \mathrm{d}x)$. But this is clear since in that case $\|a_x - a_y\|^2 + \gamma(x,y)^2 \le C|x-y|$ implies that the integral in (8.9.4) is at least logarithmically divergent as $\tau \to \infty$. ☐

Corollary 8.9.4 (Band mass quantization)**.** *Let ρ be the self-consistent density of states for the Dyson equation with data pair (a, S) and $\tau \in \mathbb{R}\setminus \operatorname{supp}\rho$. Then*

$$\rho((-\infty,\tau)) \in \left\{ \frac{1}{K}\sum_{l=1}^{L} k_l|I_l| : k_l = 1, \dots K \right\}.$$

In particular, in the $L = 1$ case when s^μ, t^μ and a are $1/2$-Hölder continuous on all of $[0,1]^2$ and $[0,1]$, respectively, then $\rho(J)$ is an integer multiple of $1/K$ for every connected component J of $\operatorname{supp}\rho$ and there are at most K such components.

PROOF. Fix $\tau \in \mathbb{R}\setminus\operatorname{supp}\rho$. We denote by $x \mapsto m_x(\tau)$ the self-adjoint solution $m(\tau)$ viewed as a function of $x \in [0,1]$ with values in $\mathbb{C}^{K\times K}$. As is clear from the Dyson equation this function inherits the regularity of the data, i.e., it is continuous on each interval I_l. By the band mass formula (8.2.10) we have

$$\rho((-\infty,\tau)) = \frac{1}{K}\sum_{l=1}^{L}\int_{I_l} \operatorname{Tr}\mathbf{1}_{(-\infty,0)}(m_x(\tau))\mathrm{d}x = \frac{1}{K}\sum_{l=1}^{L} k_l|I_l|,$$

where $k_l = \operatorname{Tr}\mathbf{1}_{(-\infty,0)}(m_x(\tau)) \in \{0,\dots,K\}$ is continuous in $x \in I_l$ with discrete values and therefore does not depend on x. ☐

Remark 8.9.5. We extend the conjecture from Remark 2.9 of [5] to the Kronecker setting. We expect that in the piecewise 1/2-Hölder continuous setting of the current section, the number of connected components of the self-consistent spectrum $\operatorname{supp}\rho$ is at most $K(2L-1)$.

8.10. Perturbations of the data pair

In this section, as an application of our results in Sections 8.4 to 8.7, we show that the Dyson equation, (8.2.3), is stable against small general perturbations of the data pair (a, S) consisting of the bare matrix a and the self-energy operator S. To that end, let $T \subset \mathbb{R}$ contain 0, $S_t\colon \mathcal{A} \to \mathcal{A}$, $t \in T$, be a family of positivity-preserving operators and $a_t = a_t^* \in \mathcal{A}$, $t \in T$, be a family of self-adjoint elements. We set $S := S_{t=0}$ and $a := a_{t=0}$ and will always assume that there are $c_1, \dots, c_5 > 0$ such that

$$c_1\langle x\rangle\mathbb{1} \le S[x] \le c_2\langle x\rangle\mathbb{1}, \qquad \|a\| \le c_3, \qquad \|S - S_t\| \le c_4 t, \qquad \|a - a_t\| \le c_5 t \quad (8.10.1)$$

for all $x \in \overline{\mathcal{A}}_+$ and for all $t \in T$. For any $t \in T$, let m_t be the solution to the Dyson equation associated to the data pair (a_t, S_t), i.e.,

$$-\, m_t(z)^{-1} = z\mathbb{1} - a_t + S_t[m_t(z)] \quad (8.10.2)$$

for $z \in \mathbb{H}$ (cf. (8.2.3)). We also set $m := m_{t=0}$.

The main result of this section, Proposition 8.10.1 below, states that $\|m_t(z) - m(z)\|$ is small for sufficiently small t and all z away from points, where $m(z)$ blows up. In the bulk and away from (almost) cusp points, we obtain stronger estimates on $\|m_t(z) - m(z)\|$.

We now introduce these concepts precisely. We recall the definition of the set $P_m := P_m^{m_*} \subset \mathbb{H}$, where $\|m(z)\|$ is larger than m_* for a given $m_* > 0$, from (8.7.68), i.e.,

$$P_m^{m_*} := \{\tau \in \mathbb{R} : \sup_{\eta>0} \|m(\tau + \mathrm{i}\eta)\| > m_*\}.$$

For any fixed $m_* > 0$ and $\delta > 0$, we introduce the set $\mathbb{D}_{\mathrm{bdd}}$ of points of distance at least δ from P_m, i.e.,

$$\mathbb{D}_{\mathrm{bdd}} := \mathbb{D}_{\mathrm{bdd}}^{m_*,\delta} := \{z \in \mathbb{H} : \operatorname{dist}(z, P_m) \ge \delta\}. \quad (8.10.3)$$

Note that $\|m(z)\| \leq \max\{m_*, \delta^{-1}\}$ for all $z \in \mathbb{D}_{\mathrm{bdd}}$ as $\|m(z)\| \leq (\mathrm{dist}(z, \operatorname{supp} \rho))^{-1}$ by (8.3.7).

We now introduce the concept of the *bulk*. Since $S \in \Sigma_{\mathrm{flat}}$, the self-consistent density of states of m (cf. Definition 8.2.2) has a continuous density $\rho\colon \mathbb{R} \to [0, \infty)$ with respect to the Lebesgue measure (cf. Proposition 8.2.3). We also write ρ for the harmonic extension of ρ to $\mathbb{H}$ which satisfies $\rho(z) = \langle \operatorname{Im} m(z) \rangle / \pi$ for $z \in \mathbb{H}$. For $\rho_* > 0$ and $\delta_s > 0$, we denote those points, where ρ is bigger than ρ_* or which are at least δ_s away from $\operatorname{supp} \rho$, by

$$\mathbb{D}_{\mathrm{bulk}} := \mathbb{D}^{\rho_*}_{\mathrm{bulk}} := \{z \in \mathbb{H} : \rho(z) \geq \rho_*\}, \qquad \mathbb{D}_{\mathrm{out}} := \mathbb{D}^{\delta_s}_{\mathrm{out}} := \{z \in \mathbb{H} \colon \mathrm{dist}(z, \operatorname{supp} \rho) \geq \delta_s\},$$

respectively. We remark that, for fixed ρ_* and δ_s, we have the inclusion $\mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}} \subset \mathbb{D}_{\mathrm{bdd}}$ for all sufficiently large m_* and sufficiently small δ by (8.3.12).

For $\tau \in \mathbb{R} \setminus \operatorname{supp} \rho$, let $\Delta(\tau)$ denote the size of the largest interval that contains τ and is contained in $\mathbb{R} \setminus \operatorname{supp} \rho$. We recall the definition of the set of almost cusp points $P_{\mathrm{cusp}} = P^{\rho_*, \Delta_*}_{\mathrm{cusp}} \subset \mathbb{R}$ for $\rho_* > 0$ and $\Delta_* > 0$ from (8.7.69), which reads as

$$P^{\rho_*, \Delta_*}_{\mathrm{cusp}} := \{\tau \in \operatorname{supp} \rho \setminus \partial \operatorname{supp} \rho : \tau \text{ is a local minimum of } \rho,\ \rho(\tau) \leq \rho_*\} \\ \cup \{\tau \in \mathbb{R} \setminus \operatorname{supp} \rho : \Delta(\tau) \leq \Delta_*\}.$$

For some $\delta_c > 0$, we denote those points which are at least δ_c away from almost cusp points by

$$\mathbb{D}_{\mathrm{nocusp}} := \{z \in \mathbb{H} : \mathrm{dist}(z, P_{\mathrm{cusp}}) \geq \delta_c\}.$$

We remark that $\mathbb{D} = \mathbb{D}_{\mathrm{bdd}} \cap \mathbb{D}_{\mathrm{cusp}}$ with the definition of $\mathbb{D}$ in (8.7.70).

In this section, the model parameters are given by $c_1, \ldots, c_5$ from (8.10.1) as well as the fixed parameters m_*, δ, ρ_*, δ_s, Δ_* and δ_c from the definitions of P_m, $\mathbb{D}_{\mathrm{bdd}}$, $\mathbb{D}_{\mathrm{bulk}}$, $\mathbb{D}_{\mathrm{out}}$, P_{cusp}, and $\mathbb{D}_{\mathrm{nocusp}}$, respectively. Thus, the comparison relation $\sim$ (compare Convention 8.3.4) is understood with respect to these parameters throughout this section.

Proposition 8.10.1. *If the self-adjoint element $a = a_{t=0}$, a_t in $\mathcal{A}$ and the positivity-preserving operators $S = S_{t=0}$, S_t on $\mathcal{A}$ satisfy* (8.10.1) *for each $t \in T$ then there is $t_* \sim 1$ such that*

(a) *Uniformly for all* $z \in \mathbb{D}_{\mathrm{bdd}}$ *and for all* $t \in [-t_*, t_*] \cap T$*, we have*

$$\|m_t(z) - m(z)\| \lesssim |t|^{1/3}.$$

In particular, $\|m_t(z)\| \lesssim 1$ *uniformly for all* $z \in \mathbb{D}_{\mathrm{bdd}}$ *and for all* $t \in [-t_*, t_*] \cap T$*.*

(b) *(Bulk and away from support of* ρ*) Uniformly for all* $z \in \mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}}$ *and for all* $t \in [-t_*, t_*] \cap T$*, we have*

$$\|m_t(z) - m(z)\| \lesssim |t|.$$

(c) *(Away from almost cusps) Uniformly for all* $z \in \mathbb{D}_{\mathrm{nocusp}} \cap \mathbb{D}_{\mathrm{bdd}}$ *and for all* $t \in [-t_*, t_*] \cap T$*, we have*

$$\|m_t(z) - m(z)\| \lesssim |t|^{1/2}.$$

In order to simplify the notation, we set $\Delta m_t = \Delta m_t(z) = m_t(z) - m(z)$. The behaviour of Δm_t will be governed by a scalar-valued cubic equation (see (8.10.5) below). This is the origin of the cubic root $|t|^{1/3}$ in the general estimate on $\|m_t(z) - m(z)\|$ in Proposition 8.10.1. In the special cases, $z \in \mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}}$ and $z \in \mathbb{D}_{\mathrm{nocusp}}$, the cubic equation simplifies to a linear or quadratic equation, respectively, which yield the improved estimates $|t|$ and $|t|^{1/2}$, respectively.

We now define two positive auxiliary functions $\widetilde{\xi}_1(z)$ and $\widetilde{\xi}_2(z)$ for $z \in \mathbb{D}_{\mathrm{bdd}}$ which will control the coefficients in the cubic equation mentioned above. For their definitions, we distinguish several subdomains of $\mathbb{D}_{\mathrm{bdd}}$. The slight ambiguity of the definitions due to overlaps between these domains does, however, not affect the validity of the following statements as the different versions of $\widetilde{\xi}_1$ as well as $\widetilde{\xi}_2$ are comparable with each other with respect to the comparison relation $\sim$ and $\widetilde{\xi}_1$ as well as $\widetilde{\xi}_2$ are only used in bounds with respect to this comparison relation. For $\rho_* \sim 1$ and $\delta_* \sim 1$, we define

- **Bulk:** If $z \in \mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}}$ then we set

$$\widetilde{\xi}_1(z) := \widetilde{\xi}_2(z) := 1. \tag{8.10.4a}$$

- **Around a regular edge:** If $z = \tau_0 + \omega + \mathrm{i}\eta \in \mathbb{D}_{\text{nocusp}} \cap \mathbb{D}_{\text{bdd}}$ with some $\tau_0 \in \partial \operatorname{supp} \rho$, $\omega \in [-\delta_*, \delta_*]$ and $\eta \in (0, \delta_*]$ then we set

$$\widetilde{\xi}_1(z) := (|\omega| + \eta)^{1/2}, \qquad \widetilde{\xi}_2(z) := 1. \tag{8.10.4b}$$

- **Close to an internal edge with a small gap:** Let $\alpha, \beta \in (\partial \operatorname{supp} \rho) \setminus P_m$ satisfy $\beta < \alpha$ and $(\beta, \alpha) \cap \operatorname{supp} \rho = \varnothing$. We set $\Delta := \alpha - \beta$. If $z \in \mathbb{D}_{\text{bdd}}$ satisfies $z = \alpha - \omega + \mathrm{i}\eta$ or $z = \beta + \omega + \mathrm{i}\eta$ for some $\omega \in [-\delta_*, \Delta/2]$ and $\eta \in (0, \delta_*]$ then we define

$$\widetilde{\xi}_1(z) := (|\omega| + \eta)^{1/2}(|\omega| + \eta + \Delta)^{1/6}, \qquad \widetilde{\xi}_2(z) := (|\omega| + \eta + \Delta)^{1/3} \tag{8.10.4c}$$

- **Around a small internal minimum:** If $z = \tau_0 + \omega + \mathrm{i}\eta \in \mathbb{D}_{\text{bdd}}$, where $\tau_0 \in \operatorname{supp} \rho \setminus \partial \operatorname{supp} \rho$ is a local minimum of ρ with $\rho(\tau_0) \leq \rho_*$, $\omega \in [-\delta_*, \delta_*]$ and $\eta \in (0, \delta_*]$ then we define

$$\widetilde{\xi}_1(z) := \rho(\tau_0)^2 + (|\omega| + \eta)^{2/3}, \qquad \widetilde{\xi}_2(z) := \rho(\tau_0) + (|\omega| + \eta)^{1/3}. \tag{8.10.4d}$$

We remark that $\tau_0 \in \partial \operatorname{supp} \rho$ is a *regular edge* if $\rho(\tau) = 0$ for all $\tau \in [\tau_0 - \varepsilon, \tau_0]$ or $\tau \in [\tau_0, \tau_0 + \varepsilon]$ for some $\varepsilon \sim 1$. In fact, $\overline{\mathbb{D}_{\text{nocusp}} \cap \mathbb{D}_{\text{bdd}}} \cap \partial \operatorname{supp} \rho$ consists only of regular edges.

In the proof of Proposition 8.10.1, we will use the following two lemmas, whose proofs we postpone until the end of this section.

Lemma 8.10.2. *Let* $\mathbb{D}_{\text{bdd}}$ *be defined as in* (8.10.3). *Let* a, S *and* $(a_t)_{t \in T}$ *and* $(S_t)_{t \in T}$ *satisfy* (8.10.1). *Then there is* $\varepsilon_1 \sim 1$ *such that if* $\|\Delta m_t(z)\| \leq \varepsilon_1$ *for some* $z \in \mathbb{D}_{\text{bdd}}$, $t \in T$, *then there are* $l, b \in \mathcal{A}$ *depending on* z *such that* $\Theta_t := \langle l, \Delta m_t \rangle / \langle l, b \rangle$ *satisfies a cubic inequality*

$$|\Theta_t^3 + \xi_2 \Theta_t^2 + \xi_1 \Theta_t| \lesssim |t| \tag{8.10.5}$$

with complex coefficients ξ_1 *and* ξ_2 *depending on* z *and* t. *The function* Θ_t *depends continuously on* $\operatorname{Im} z$ *and we also have* $|\Theta_t| \lesssim \|\Delta m_t\|$ *as well as* $\|\Delta m_t\| \lesssim |\Theta_t| + |t|$ *for all* $t \in T$.

The coefficients, ξ_1 *and* ξ_2, *behave as follows: There are* $\delta_* \sim 1$, $\rho_* \sim 1$ *and* $c_* \sim 1$ *such that, with the appropriate definitions of* $\widetilde{\xi}_1$ *and* $\widetilde{\xi}_2$ *from* (8.10.4), *we have*

- *If* $z \in \mathbb{D}_{\mathrm{bdd}}$ *satisfies the conditions for* (8.10.4a) *or* (8.10.4c) *with* $\omega \in [c_*\Delta, \Delta/2]$ *then we have*

$$|\xi_1(z)| \sim \widetilde{\xi}_1(z), \qquad |\xi_2(z)| \lesssim \widetilde{\xi}_2(z). \tag{8.10.6a}$$

- *If* $z \in \mathbb{D}_{\mathrm{bdd}}$ *satisfies the conditions for* (8.10.4b) *or* (8.10.4c) *with* $\omega \in [-\delta_*, c_*\Delta]$ *or* (8.10.4d) *then we have*

$$|\xi_1(z)| \sim \widetilde{\xi}_1(z), \qquad |\xi_2(z)| \sim \widetilde{\xi}_2(z). \tag{8.10.6b}$$

All implicit constants in this lemma are uniform for any $t \in T$.

Lemma 8.10.3. *For* $0 < \eta_* < \eta^* < \infty$*, let* $\xi_1, \xi_2 \colon [\eta_*, \eta^*] \to \mathbb{C}$ *be complex-valued functions and* $\widetilde{\xi}_1, \widetilde{\xi}_2, d \colon [\eta_*, \eta^*] \to \mathbb{R}^+$ *be continuous.*

Suppose that some continuous function $\Theta \colon [\eta_*, \eta^*] \to \mathbb{C}$ *satisfies the cubic inequality*

$$|\Theta^3 + \xi_2\Theta^2 + \xi_1\Theta| \lesssim d \tag{8.10.7}$$

on $[\eta_*, \eta^*]$ *as well as*

$$|\Theta| \lesssim \min\Big\{d^{1/3}, \frac{d^{1/2}}{\widetilde{\xi}_2^{1/2}}, \frac{d}{\widetilde{\xi}_1}\Big\} \tag{8.10.8}$$

at η_**. If one of the following two sets of relations holds true:*

1) (i) $\widetilde{\xi}_2^3/d$, $\widetilde{\xi}_1^3/d^2$, $\widetilde{\xi}_1^2/(d\widetilde{\xi}_2)$ *are monotonically increasing functions,*
   (ii) $|\xi_1| \sim \widetilde{\xi}_1$, $|\xi_2| \sim \widetilde{\xi}_2$,
   (iii) $d^2/\widetilde{\xi}_1^3 + d\widetilde{\xi}_2/\widetilde{\xi}_1^2$ *at* η^* *is sufficiently small depending on the implicit constants in 1) (ii) as well as* (8.10.7) *and* (8.10.8).
2) (i) $\widetilde{\xi}_1^3/d^2$ *is a monotonically increasing function,*
   (ii) $|\xi_1| \sim \widetilde{\xi}_1$, $|\xi_2| \lesssim \widetilde{\xi}_1^{1/2}$.

then, on $[\eta_*, \eta^*]$*, we have the bound*

$$|\Theta| \lesssim \min\Big\{d^{1/3}, \frac{d^{1/2}}{\widetilde{\xi}_2^{1/2}}, \frac{d}{\widetilde{\xi}_1}\Big\}. \tag{8.10.9}$$

Proof of Proposition 8.10.1. We start the proof by introducing the control parameter $M(t)$. Let $\widetilde{\xi}_1$ and $\widetilde{\xi}_2$ be defined as in (8.10.4). For $t \in \mathbb{R}$, we set

$$M(t) := \min\{|t|^{1/3}, \widetilde{\xi}_2^{-1/2}|t|^{1/2}, \widetilde{\xi}_1^{-1}|t|\}. \tag{8.10.10}$$

We remark that M also depends on z as $\widetilde{\xi}_1$ and $\widetilde{\xi}_2$ depend on z.

We will prove below that there are $t_* \sim 1$ and $C \sim 1$ such that, for any fixed $t \in [-t_*, t_*] \cap T \setminus \{0\}$ (if this set is nonempty) and $z \in \mathbb{D}_{\rm bdd}$, we have the implication

$$\|\Delta m_t(\operatorname{Re} z + \mathrm{i}\eta)\| \le \varepsilon_1 \quad \text{for all } \eta \ge \operatorname{Im} z \qquad \Rightarrow \qquad \|\Delta m_t(z)\| \le CM(t), \tag{8.10.11}$$

where $\varepsilon_1 \sim 1$ is from Lemma 8.10.2.

Armed with (8.10.11), by possibly shrinking $t_* \sim 1$, we can assume that $2Ct_*^{1/3} \le \varepsilon_1$. We fix $\tau \in \mathbb{R}$ and $t \in [-t_*, t_*] \cap T \setminus \{0\}$ and set

$$\eta_* := \sup\{\eta > 0 : \|\Delta m_t(\tau + \mathrm{i}\eta)\| \ge 2CM(t)\}.$$

Here, we use the convention $\eta_* = -\infty$ if the set is empty. Note that $\|\Delta m_t(\tau+\mathrm{i}\eta)\| \le 2\eta^{-1}$ since m and m_t are Stieltjes transforms. Hence, $\eta_* < \infty$ as $t \ne 0$.

We prove now that $\eta_* \le \inf\{\operatorname{Im} z : z \in \mathbb{D}_{\rm bdd},\ \operatorname{Re} z = \tau\}$. For a proof by contradiction, we suppose that there is $z_* \in \mathbb{D}_{\rm bdd}$ such that $\operatorname{Re} z_* = \tau$ and $\operatorname{Im} z_* = \eta_*$ (note that if $\tau + \mathrm{i}\eta \in \mathbb{D}_{\rm bdd}$ then $\tau + \mathrm{i}\eta' \in \mathbb{D}_{\rm bdd}$ for any $\eta' \ge \eta$). Since Δm_t is continuous in z, we have $\|\Delta m_t(z_*)\| = 2CM(t)$. Thus, $\|\Delta m_t(\tau + \mathrm{i}\eta)\| \le 2Ct_*^{1/3} \le \varepsilon_1$ for all $\eta \ge \eta_*$ by the choice of t_*. From (8.10.11), we conclude $\|\Delta m_t(z_*)\| \le CM(t)$, which contradicts $\|\Delta m_t(z_*)\| = 2CM(t)$. Thus, $\eta_* \le \inf\{\operatorname{Im} z : z \in \mathbb{D}_{\rm bdd},\ \operatorname{Re} z = \tau\}$.

As τ was arbitrary, this yields $\|\Delta m_t(z)\| \le 2CM(t)$ for all $z \in \mathbb{D}_{\rm bdd}$, which proves part (a) of Proposition 8.10.1 up to (8.10.11). Since $\widetilde{\xi}_1(z) \sim 1$ for $z \in \mathbb{D}_{\rm bulk} \cup \mathbb{D}_{\rm out}$ and $\widetilde{\xi}_2(z) \sim 1$ for $z \in \mathbb{D}_{\rm nocusp} \cap \mathbb{D}_{\rm bdd}$, we also obtain part (b) and (c) from the definition of M in (8.10.10).

Hence, it suffices to show (8.10.11) to complete the proof of Proposition 8.10.1. In order to prove (8.10.11), we use Lemma 8.10.3 with $\Theta(\eta) = \Theta_t(\operatorname{Re} z+\mathrm{i}\eta)$, $\eta \ge \eta_* := \operatorname{Im} z$, $d = |t|$, and ξ_1, ξ_2 and $\widetilde{\xi}_1$, $\widetilde{\xi}_2$ are chosen as in (8.10.5) of Lemma 8.10.2 and (8.10.4), respectively. As $\|\Delta m_t(\operatorname{Re} z + \mathrm{i}\eta)\| \le \varepsilon_1$ for all $\eta \ge \operatorname{Im} z$, we conclude that (8.10.7) is satisfied with $d = |t|$ due to (8.10.5).

We first consider $z \in \mathbb{D}_{\rm bulk} \cup \mathbb{D}_{\rm out}$. If $z \in \mathbb{D}_{\rm bulk} \cup \mathbb{D}_{\rm out}$ then $\operatorname{Re} z + \mathrm{i}\eta \in \mathbb{D}_{\rm bulk} \cup \mathbb{D}_{\rm out}$ and $\xi_1(\operatorname{Re} z+\mathrm{i}\eta) = \xi_2(\operatorname{Re} z+\mathrm{i}\eta) = 1$ for all $\eta \ge \eta_*$ and assumption 2) of Lemma 8.10.3 is always fulfilled. Since $\|\Delta m_t(\operatorname{Re} z + \mathrm{i}\eta)\| \le 2\eta^{-1}$ as remarked above and $t \ne 0$, the condition in (8.10.8) is met for some sufficiently large $\eta > 0$. Hence, by Lemma 8.10.3, there is $C \sim 1$ such that $|\Theta_t(z)| \le CM(t)$. Possibly increasing $C \sim 1$ and using $|t| \le t_* \sim 1$ yield $\|\Delta m_t(z)\| \le CM(t)$ due to $\|\Delta m_t\| \lesssim |\Theta_t| + |t|$ from Lemma 8.10.2.

For each $z \in \mathbb{D}_{\mathrm{bdd}} \setminus \mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}}$, due to (8.10.6), we have $\xi_1(z_\delta) \sim 1$ and $\xi_2(z_\delta) \sim 1$ for $z_\delta := \operatorname{Re} z + \mathrm{i}\delta_*$, where $\delta_* \sim 1$ is as in Lemma 8.10.2. Hence, we conclude $|\Theta_t(z_\delta)| \le CM(t)$ as for $z \in \mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}}$. For each $z \in \mathbb{D}_{\mathrm{bdd}} \setminus \mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}}$, the validity of assumption 1) or assumption 2) of Lemma 8.10.3 can be read off from (8.10.6). Lemma 8.10.3, thus, implies $|\Theta_t(z)| \le CM(t)$. As before, we conclude $\|\Delta m_t(z)\| \le CM(t)$ from Lemma 8.10.2. This completes the proof of (8.10.11) and, hence, the one of Proposition 8.10.1. ☐

Proof of Lemma 8.10.2. We remark that a straightforward computation starting from (8.2.3) and (8.10.2) yields

$$B[\Delta m_t] = A[\Delta m_t, \Delta m_t] + K[\Delta^S, \Delta^a, \Delta m_t] + T[\Delta^S, \Delta^a], \tag{8.10.12}$$

where $B := \mathrm{Id} - C_m S$, $A[x, y] := (mS[x]y + yS[x]m)/2$ are defined as in (8.6.23), $\Delta^S := S_t - S$, $\Delta^a := a_t - a$ and

$$\begin{aligned} K[\Delta^S, \Delta^a, \Delta m_t] &= \frac{1}{2}(m\Delta^S[\Delta m_t]\Delta m_t + \Delta m_t \Delta^S[\Delta m_t]m + m\Delta^S[m]\Delta m_t + \Delta m_t \Delta^S[m]m) \\ &\quad - \frac{1}{2}(m\Delta^a \Delta m_t + \Delta m_t \Delta^a m), \\ T[\Delta^S, \Delta^a] &= m\Delta^S[m]m - m\Delta^a m. \end{aligned}$$

In the following, we will split $\mathbb{D}_{\mathrm{bdd}}$ into two regimes and choose l and b according to the regime. In both cases, we use the definitions

$$\Theta := \Theta_t = \frac{\langle l, \Delta m_t \rangle}{\langle l, b \rangle}, \qquad r = r_t := Q[\Delta m_t], \qquad Q := \mathrm{Id} - \frac{\langle l, \cdot \rangle}{\langle l, b \rangle} b. \tag{8.10.13}$$

In particular, $\Delta m_t = \Theta b + r$. We denote by $\rho(z)$ the harmonic extension of ρ, i.e., $\rho(z) = \langle \operatorname{Im} m(z) \rangle / \pi$.

If z is in the bulk or away from $\operatorname{supp} \rho$ then $\Delta m_t(z)$ is in fact governed by a scalar-valued linear equation for Θ_t with l and b chosen appropriately. Similarly, if z is close to a regular edge or close to an almost cusp point then $\Delta m_t(z)$ is governed by a quadratic or cubic equation, respectively. In order to treat these cases uniformly, we will artificially write all of these equations as a cubic equation by adding and subtracting apparently superfluous terms.

Case 1: We first assume that $z \in \mathbb{D}_{\mathrm{bdd}}$ satisfies $\rho(z) \geq \rho_* $ for some $\rho_* \sim 1$ or $\mathrm{dist}(z, \mathrm{supp}\,\rho) \geq \delta$ for some $\delta \sim 1$, i.e., $z \in \mathbb{D}_{\mathrm{bulk}}^{\rho_*} \cup \mathbb{D}_{\mathrm{out}}^{\delta}$. This implies that B is invertible and $\|B^{-1}\| \lesssim 1$ due to (8.4.1), $\|S\|_{2\to\|\cdot\|} \lesssim 1$, $\|m(z)\| \lesssim 1$ and Lemma 8.12.2 (ii). In this case, we choose $l = b = \mathbb{1}$ and apply QB^{-1} to (8.10.12) to obtain

$$r = QB^{-1}(A[\Delta m_t, \Delta m_t] + K[\Delta^S, \Delta^a, \Delta m_t] + T[\Delta^S, \Delta^a]) = \mathcal{O}(|\Theta|^2 + \|r\|\|\Delta m_t\| + |t|),$$

where we used that $\|m\| \lesssim 1$ on $\mathbb{D}_{\mathrm{bdd}}$ as well as $\|\Delta^S\| + \|\Delta^a\| \lesssim |t|$. Shrinking $\varepsilon_1 \sim 1$, using $\|\Delta m_t\| \leq \varepsilon_1$ and absorbing $\|r\|\|\Delta m_t\|$ into the left-hand side yield $\|r\| \lesssim |\Theta|^2 + |t|$. Thus, $\|\Delta m_t\| \lesssim |\Theta| + |t|$. Hence, applying B^{-1} and $\langle\,\cdot\,\rangle$ to (8.10.12) and using $\langle r\rangle = 0$ as well as $\|\Delta m_t\| \lesssim |\Theta| + |t|$, we find $\xi_2 \in \mathbb{C}$ such that $|\xi_2| \lesssim 1 = \widetilde{\xi}_2$ and

$$\Theta = -\xi_2\Theta^2 + \mathcal{O}(|t||\Theta| + |t|) = -\xi_2\Theta^2 + \mathcal{O}(|t|).$$

Adding and subtracting Θ^3 on the left-hand side as well as setting $\xi_1 := 1 - \Theta^2$ show (8.10.5) in Case 1 for sufficiently small $\varepsilon_1 \sim 1$ as $|\Theta| \lesssim \|\Delta m_t\| \leq \varepsilon_1$ implies $|\xi_1| \sim 1 = \widetilde{\xi}_1$. This completes the proof of (8.10.6a) for $z \in \mathbb{D}_{\mathrm{bulk}} \cup \mathbb{D}_{\mathrm{out}}$.

Case 2: We now prove (8.10.5) for $z \in \mathbb{D}_{\mathrm{bdd}}$ satisfying $\rho(z) \leq \rho_*$ and $\mathrm{dist}(z, \mathrm{supp}\,\rho) \leq \delta$ with sufficiently small $\rho_* \sim 1$ and $\delta \sim 1$. For any $\varepsilon_* \sim 1$, we find $\delta \sim 1$ such that $\rho(z)^{-1}\mathrm{Im}\, z \leq \varepsilon_*$ for all $z \in \mathbb{H}$ satisfying $\mathrm{dist}(z, \mathrm{supp}\,\rho) \leq \delta$ due to (8.5.26) and the 1/3-Hölder continuity of $z \mapsto \rho(z)^{-1}\mathrm{Im}\, z$ by Lemma 8.5.4 (ii). Therefore, using $\rho(z) \leq \rho_*$, we see that Lemma 8.5.1 and Corollary 8.5.2 are applicable for sufficiently small $\rho_* \sim 1$ and $\delta \sim 1$. They yield $l, b \in \mathcal{A}$ which we use to define Θ and r as in (8.10.13), i.e., $\Delta m_t = \Theta b + r$ and $\Theta = \langle l, \Delta m_t\rangle / \langle l, b\rangle$.

In order to derive (8.10.5), we now follow the proof of Lemma 8.6.2 applied to (8.10.12) instead of (8.6.10). Here, Δ^a and Δ^S play the role of e. In fact, by Lemma 8.5.1 and Corollary 8.5.2, the first two bounds in (8.6.12) are fulfilled. Owing to $\|m\| \lesssim 1$, the third bound in (8.6.12) is trivially satisfied. Instead of the last two bounds in (8.6.12), we use

$$\|T[\Delta^S, \Delta^a]\| \lesssim \|\Delta^S\| + \|\Delta^a\|, \qquad \|K[\Delta^S, \Delta^a, \Delta m_t]\| \lesssim (\|\Delta^S\| + \|\Delta^a\|)\|\Delta m_t\|,$$

due to $\|m\| \lesssim 1$ and $\|\Delta m_t\| \lesssim 1$. In fact, the last bound in (8.6.12) will not hold true for a general $y \in \mathcal{A}$ but in the proof of Lemma 8.6.2 it is only used with the special choice $y = \Delta m_t$. We choose $\varepsilon_1 \leq \varepsilon$ for ε from Lemma 8.6.2 and obtain the cubic equation (8.6.14) from Lemma 8.6.2 with $\mu_0 = \langle l\,, T[\Delta^S, \Delta^a]\rangle$ and $\|e\|$ replaced by $|t|$ as $\|\Delta^S\| + \|\Delta^a\| \lesssim |t|$. In particular, $|\mu_0| \lesssim |t|$. We decompose the error term $\widetilde{e} = \mathcal{O}(|\Theta|^4 + |t||\Theta| + |t|^2)$ from (8.6.14) into $\widetilde{e} = \widetilde{e}_1\Theta^3 + \widetilde{e}_2$ with $\widetilde{e}_1, \widetilde{e}_2 \in \mathbb{C}$ satisfying $\widetilde{e}_1 = \mathcal{O}(|\Theta|)$ and $\widetilde{e}_2 = \mathcal{O}(|t||\Theta| + |t|^2)$. With the notation of Lemma 8.6.2, the cubic equation (8.6.14) can be written as

$$(\mu_3 - \widetilde{e}_1)\Theta^3 + \mu_2\Theta^2 + \mu_1\Theta = -\mu_0 + \widetilde{e}_2 = \mathcal{O}(|t|).$$

Since A and B introduced above have the same definitions as in (8.6.23) and μ_3, μ_2 and μ_1 in (8.6.15) depend only on A and B, Lemma 8.6.3 yields the expansions of μ_3, μ_2 and μ_1 in (8.6.24) for sufficiently small $\rho_* \sim 1$ and $\delta \sim 1$. By possibly shrinking $\varepsilon_1 \sim 1$, we find $c \sim 1$ such that $|\mu_3 - \widetilde{e}_1| + |\mu_2| \geq 2c$ as $|\widetilde{e}_1| \lesssim |\Theta| \lesssim \|\Delta m_t\| \leq \varepsilon_1$. Here, we also used $|\mu_3| + |\mu_2| \gtrsim \psi + |\sigma|$ by (8.6.24) as well as (8.5.35).

Consequently, we obtain (8.10.5), where we introduced

$$\begin{aligned} \xi_2 &:= \Big(\mu_2 + (\mu_3 - \widetilde{e}_1 - 1)\Theta\Big)\mathbf{1}(|\mu_2| \geq c) + \frac{\mu_2}{\mu_3 - \widetilde{e}_1}\mathbf{1}(|\mu_2| < c), \\ \xi_1 &:= \mu_1\mathbf{1}(|\mu_2| \geq c) + \frac{\mu_1}{\mu_3 - \widetilde{e}_1}\mathbf{1}(|\mu_2| < c). \end{aligned}$$

Hence, we have $|\xi_2| \sim |\mu_2|$ and $|\xi_1| \sim |\mu_1|$ for sufficiently small $\varepsilon_1 \sim 1$ as $|\widetilde{e}_1| \lesssim |\Theta|$ and $|\Theta| \lesssim \|\Delta m_t\| \leq \varepsilon_1$. This completes the proof of (8.10.5) in Case 2.

It remains to show the scaling relations in (8.10.6) for $z \in \mathbb{D}_{\mathrm{bdd}}$ satisfying $\rho(z) \leq \rho_*$ and $\mathrm{dist}(z, \mathrm{supp}\,\rho) \leq \delta$ in order to complete the proof of Lemma 8.10.2. Starting from $|\xi_1| \sim |\mu_1|$ and $|\xi_2| \sim |\mu_2|$ proven in Case 2, we conclude as in the proof of (10.6) in [4] that

$$|\xi_1| \sim \rho(z)^2 + |\sigma(z)|\rho(z) + \rho(z)^{-1}\mathrm{Im}\, z, \qquad |\xi_2| \sim \rho(z) + |\sigma(z)|,$$

where σ is defined as in (8.5.12). Here, ξ_1 and ξ_2 play the role of π_1 and π_2, respectively, in [4]. Their definitions differ slightly but this does not affect the straightforward estimates. Note that the proof in [4] relies on the expansions of μ_1, μ_2 and μ_3 from (8.33) in [4]. These are the exact analogues of (8.6.24), where ρ plays the role of α from [4].

Based on the singularity analysis of the self-consistent density of states ρ in [4], Corollary A.1 in [4] characterizes the behaviour of the harmonic extension $\rho(z)$ for $z \in \mathbb{H}$ in the vicinity of these singularities. For $z \in \mathbb{D}_{\mathrm{bdd}}$ satisfying $\rho(z) \leq \rho_*$ and $\operatorname{dist}(z, \operatorname{supp}\rho) \leq \delta$, following the proof of Corollary A.1 in [4] and using Theorem 8.7.1 above instead of Theorem 2.6 in [4], we obtain the statements of Corollary A.1 in [4] in our setup as well. Similarly, the proof of (10.7) in [4] yields

$$|\sigma(\beta)| \sim |\sigma(\alpha)| \sim (\alpha - \beta)^{1/3}, \qquad |\sigma(\tau_0)| \lesssim \rho(\tau_0)^2,$$

where $\alpha, \beta \in (\partial \operatorname{supp}\rho) \setminus P_m$ satisfy $\beta < \alpha$ and $(\beta, \alpha) \cap \operatorname{supp}\rho = \varnothing$ and $\tau_0 \in \operatorname{supp}\rho \setminus \partial \operatorname{supp}\rho$ is a local minimum of ρ and $\rho(\tau_0) \leq \rho_*$. Here, we use Lemma 8.7.15 above and $|\sigma|^{1/3} \sim \widehat{\Delta}$ instead of Lemma 9.17 in [4] and Lemma 8.7.13 above instead of Lemma 9.2 in [4]. We then follow the proof of Proposition 4.3 in [7] and use the 1/3-Hölder continuity of σ proven in Lemma 8.5.5 (i). This yields the missing scaling relations in (8.10.6) and, hence, completes the proof of Lemma 8.10.2. □

In the previous proof of Lemma 8.10.2, we have established the following fact.

Remark 8.10.4 (Scaling relations of $\rho(z)$)**.** The scaling relations of $\rho(z)$ in Corollary A.1 of [4] hold true for $z \in \mathbb{D}_{\mathrm{bdd}}$ if there are $c_1, c_2, c_3 > 0$ such that the data pair (a, S) satisfies

$$c_1 \langle x \rangle \mathbb{1} \leq S[x] \leq c_2 \langle x \rangle \mathbb{1}, \qquad \|a\| \leq c_3$$

for all $x \in \overline{\mathcal{A}}_+$.

Proof of Lemma 8.10.3. By dividing the cubic inequality through d and considering $\frac{\Theta}{d^{1/3}}$ instead of Θ, we may assume that $d = 1$. We fix $\varepsilon \in (0,1)$ sufficiently small. First we prove the lemma under assumption 1). Owing to the smallness of $\frac{1}{\widetilde{\xi}_1^3} + \frac{\widetilde{\xi}_2}{\widetilde{\xi}_1^2}$ at η^* as well as the monotonicity of $\widetilde{\xi}_1$ and $\frac{\widetilde{\xi}_1^2}{\widetilde{\xi}_2}$ there are $0 < \eta_1, \eta_2 < \eta^*$ with the following properties: (i) $\widetilde{\xi}_2 \geq \varepsilon^4 \widetilde{\xi}_1^2$ on $[\eta_*, \eta_1]$; (ii) $\widetilde{\xi}_2 \leq \varepsilon^4 \widetilde{\xi}_1^2$ on $[\eta_1, \eta^*]$; (iii) $\varepsilon \widetilde{\xi}_1 \leq 1$ on $[\eta_*, \eta_2]$; (iv) $\varepsilon \widetilde{\xi}_1 \geq 1$ on $[\eta_2, \eta^*]$. Here the intervals $[\eta_*, \eta_2]$ and $[\eta_*, \eta_1]$ may be empty. We will now assume the bound $|\Theta| \lesssim \min\{1, \frac{1}{\widetilde{\xi}_2^{1/2}}, \frac{1}{\widetilde{\xi}_1}\}$ at the initial value η^* and bootstrap it down to η_*. Now we distinguish two cases:

Case 1 ($\eta_1 \geq \eta_2$): On $[\eta_1, \eta^*]$ we have $\varepsilon \widetilde{\xi}_1 \geq 1$ and $\widetilde{\xi}_2 \leq \varepsilon^4 \widetilde{\xi}_1^2$. Thus, by the cubic inequality

$$|\Theta| \lesssim \min\Big\{1, \frac{1}{\widetilde{\xi}_2^{1/2}}\Big\} \quad \text{implies} \quad |\Theta| \lesssim \frac{1}{\widetilde{\xi}_1} \lesssim \min\Big\{\varepsilon, \frac{\varepsilon^2}{\widetilde{\xi}_2^{1/2}}\Big\}.$$

In particular, there is a gap in the values of $|\Theta|$ and by continuity all values lie below the gap on $[\eta_1, \eta^*]$.

The interval $[\eta_*, \eta_1]$ is split again, $[\eta_*, \eta_1] = [\eta_*, \eta_3] \cup [\eta_3, \eta_1]$, where η_3 is chosen such that (i) $\widetilde{\xi}_2\varepsilon^2 \geq 1$ on $[\eta_3, \eta_1]$; (ii) $\widetilde{\xi}_2\varepsilon^2 \leq 1$ on $[\eta_*, \eta_3]$. Here one or both of these intervals may be empty. Using $\widetilde{\xi}_2 \geq \varepsilon^4\widetilde{\xi}_1^2$ we see that on $[\eta_3, \eta_1]$ the bound

$$|\Theta| \lesssim \min\Big\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^3\widetilde{\xi}_1}\Big\} \quad \text{implies} \quad |\Theta| \lesssim \frac{1}{\varepsilon^{3/2}\widetilde{\xi}_2^{1/2}} \lesssim \min\Big\{\frac{1}{\varepsilon^{1/2}}, \frac{1}{\varepsilon^{7/2}\widetilde{\xi}_1}\Big\}.$$

Again the gap in the values of $|\Theta|$ allows us to infer from the bound $|\Theta| \lesssim \min\{1, \frac{1}{\widetilde{\xi}_2^{1/2}}, \frac{1}{\widetilde{\xi}_1}\}$ at η_1 that $|\Theta|$ satisfies the same bound on $[\eta_3, \eta_1]$ up to an ε-dependent multiplicative constant.

Finally, on $[\eta_*, \eta_3]$ we have $\widetilde{\xi}_2 \leq \varepsilon^{-2}$ and $\widetilde{\xi}_1^2 \leq \varepsilon^{-4}\widetilde{\xi}_2 \leq \varepsilon^{-6}$. Using the cubic inequality this immediately implies $|\Theta| \lesssim_\varepsilon 1 \lesssim_\varepsilon \min\{1, \frac{1}{\widetilde{\xi}_2^{1/2}}, \frac{1}{\widetilde{\xi}_1}\}$. Here and in the following, the notation $\lesssim_\varepsilon$ indicates that the implicit constant in the bound is allowed to depend on ε.

Case 2 ($\eta_1 \leq \eta_2$): On $[\eta_2, \eta^*]$ we have $\varepsilon\widetilde{\xi}_1 \geq 1$ and $\widetilde{\xi}_2 \leq \varepsilon^4\widetilde{\xi}_1^2$. So this regime is treated exactly as in the beginning of *Case 1*. On $[\eta_*, \eta_2]$ we have $\varepsilon\widetilde{\xi}_1 \leq 1$ and $\widetilde{\xi}_2 \leq \widetilde{\xi}_2(\eta_2) \leq \varepsilon^4\widetilde{\xi}_1(\eta_2)^2 = \varepsilon^2$, which implies $|\Theta| \lesssim_\varepsilon 1 \lesssim_\varepsilon \min\{1, \frac{1}{\widetilde{\xi}_2^{1/2}}, \frac{1}{\widetilde{\xi}_1}\}$.

Now we prove the lemma under assumption 2). In this case we choose $0 < \eta_1 < \eta^*$ such that (i) $\varepsilon\widetilde{\xi}_1 \geq 1$ on $[\eta_1, \eta^*]$; (ii) $\varepsilon\widetilde{\xi}_1 \leq 1$ on $[\eta_*, \eta_1]$. Here the interval $[\eta_*, \eta_1]$ may be empty.

On $[\eta_1, \eta^*]$ the bound

$$|\Theta| \lesssim 1 \ \text{implies} \ \widetilde{\xi}_1|\Theta| \lesssim 1 + \widetilde{\xi}_1^{1/2}|\Theta|^2 \lesssim \varepsilon^{-1/2} + \varepsilon^{1/2}\widetilde{\xi}_1|\Theta| \ \text{implies} \ |\Theta| \lesssim \frac{1}{\sqrt{\varepsilon}\,\widetilde{\xi}_1} \leq \sqrt{\varepsilon}\,.$$

From the gap in the values of $|\Theta|$ and its continuity we infer $|\Theta| \lesssim \min\{\sqrt{\varepsilon}, \frac{1}{\sqrt{\varepsilon}\widetilde{\xi}_1}\}$. On $[\eta_*, \eta_1]$ we use $\widetilde{\xi}_1 \leq \varepsilon^{-1}$ and $|\xi_2| \lesssim \widetilde{\xi}_1^{1/2} \leq \varepsilon^{-1/2}$ to conclude $|\Theta| \lesssim_\varepsilon 1 \lesssim_\varepsilon \min\{1, \frac{1}{\widetilde{\xi}_1}\}$. This finishes the proof of the lemma. ∎

Lemma 8.10.5 (Hölder continuity of σ and ψ with respect to a and S)**.** *Let $T \subset \mathbb{R}$ contain 0. For each $t \in T$, we assume that the linear operator $S_t\colon \mathcal{A} \to \mathcal{A}$ satisfies*

$$c_1\langle x\rangle\mathbb{1} \leq S_t[x] \leq c_2\langle x\rangle\mathbb{1} \tag{8.10.14}$$

for all $x \in \overline{\mathcal{A}}_+$ and some $c_2 > c_1 > 0$. Moreover, let $a_t = a_t^* \in \mathcal{A}$ be self-adjoint such that S_t and a_t satisfy (8.10.1) *with $a := a_{t=0}$ and $S := S_{t=0}$. Let m_t be the solution to* (8.10.2) *and $\rho(z) := \langle \operatorname{Im} m_0(z)\rangle/\pi$ for $z \in \mathbb{H}$.*

If σ_t and ψ_t are defined according to (8.5.12)*, where m is replaced by m_t, then there are $\rho_* \sim 1$ and $t_* \sim 1$ such that*

$$|\sigma_t(z_1) - \sigma_0(z_1)| \lesssim |t|^{1/3}, \qquad |\psi_t(z_2) - \psi_0(z_2)| \lesssim |t|^{1/3}$$

for all $t \in [-t_*, t_*] \cap T$ and all $z_1, z_2 \in \mathbb{D}_{\mathrm{bdd}} \cap \{z \in \mathbb{H}\colon |z| \leq c_6\}$ satisfying $\rho(z_1) \leq \rho_*$ and $\rho(z_2) + \rho(z_2)^{-1}\operatorname{Im} z_2 \leq \rho_*$. Here, $c_6 > 0$ is also considered a model parameter.

Proof. We choose t_* as in Proposition 8.10.1 and conclude from this result that $\|m_t(z)\| \leq k_3$ for all $t \in [-t_*, t_*] \cap T$, all $z \in \mathbb{D}_{\mathrm{bdd}}$ and some $k_3 \sim 1$. Hence, owing to (8.10.1), (8.10.14) and Lemma 8.4.8 (ii), the conditions of Assumptions 8.4.5 are met on $\mathbb{D}_{\mathrm{bdd}} \cap \{z \in \mathbb{H}\colon |z| \leq c_6\}$. Therefore, the lemma follows from Remark 8.5.6 (ii) and (iii) as well as Proposition 8.10.1 (a). ☐

8.11. Stieltjes transforms of positive operator-valued measures

In this section, we will show some results about the Stieltjes transform of a positive operator-valued measure on $\mathcal{A}$.

We first prove Lemma 8.3.1 by generalizing existing proofs in the matrix algebra setup. Since we have not found the general version in the literature, we provide a proof here for the convenience of the reader. In the proof of Lemma 8.3.1, we will use that a von Neumann algebra is always isomorphically isomorphic as a Banach space to the dual space of a Banach space. In our setup, this Banach space and the identification are simple to introduce which we will explain now. Analogously to L^2 defined in Section 8.4, we define L^1 to be the completion of $\mathcal{A}$ when equipped with the norm $\|x\|_1 := \langle (x^*x)^{1/2}\rangle = \langle |x|\rangle$ for $x \in \mathcal{A}$. Moreover, we extend $\langle\,\cdot\,\rangle$ to L^1 and remark that $xy \in L^1$ for $x \in \mathcal{A}$ and $y \in L^1$. It is well-known (e.g. [138, Theorem 2.18]) that the dual space $(L^1)'$ of L^1 can be identified with $\mathcal{A}$ via the isometric isomorphism

$$\mathcal{A} \to (L^1)',\ x \mapsto \psi_x, \qquad \psi_x\colon L^1 \to \mathbb{C},\ y \mapsto \langle xy \rangle. \tag{8.11.1}$$

We stress that the existence of this isomorphism requires the state $\langle \cdot \rangle$ to be normal.

Proof of Lemma 8.3.1. From (8.3.5), we conclude that

$$\lim_{\eta\to\infty} \mathrm{i}\eta\langle x\,, h(\mathrm{i}\eta)x\rangle = -\langle x\,, x\rangle$$

for all $x \in \mathcal{A}$. Hence, $z \mapsto \langle x\,, h(z)x\rangle$ is the Stieltjes transform of a unique finite positive measure v_x on $\mathbb{R}$ with $v_x(\mathbb{R}) = \|x^*x\|_1$.

For any $x \in \mathcal{A}$, we can find $x_1, \dots x_4 \in \overline{\mathcal{A}}_+$ such that $x = x_1 - x_2 + \mathrm{i}x_3 - \mathrm{i}x_4$. We define

$$\varphi_B(x) := v_{\sqrt{x_1}}(B) - v_{\sqrt{x_2}}(B) + \mathrm{i}v_{\sqrt{x_3}}(B) - \mathrm{i}v_{\sqrt{x_4}}(B) \tag{8.11.2}$$

for $B \in \mathcal{B}$. This definition is independent of the representation of x. Indeed, for fixed $x \in \mathcal{A}$, any representation $x = x_1 - x_2 + \mathrm{i}x_3 - \mathrm{i}x_4$ with $x_1, \dots, x_4 \in \overline{\mathcal{A}}_+$ defines a complex measure $\varphi_\cdot(x)$ through $B \mapsto \varphi_B(x)$ on $\mathbb{R}$ via (8.11.2). However, extending h to the lower half-plane by setting $h(z) := h(\bar{z})^*$ for $z \in \mathbb{C}$ with $\operatorname{Im} z < 0$, the Stieltjes transform of $\varphi_\cdot(x)$ is given by

$$\begin{aligned}\int_{\mathbb{R}} \frac{\varphi_{\mathrm{d}\tau}(x)}{\tau - z} &= \langle \sqrt{x_1}\,, h(z)\sqrt{x_1}\rangle - \langle \sqrt{x_2}\,, h(z)\sqrt{x_2}\rangle + \mathrm{i}\langle \sqrt{x_3}\,, h(z)\sqrt{x_3}\rangle - \mathrm{i}\langle \sqrt{x_4}\,, h(z)\sqrt{x_4}\rangle \\ &= \langle h(z)x\rangle\end{aligned}$$

for all $z \in \mathbb{C}\setminus\mathbb{R}$. This formula shows that the Stieltjes transform of $\varphi_\cdot(x)$ is independent of the decomposition $x = x_1 - x_2 + \mathrm{i}x_3 - \mathrm{i}x_4$. Hence, $\varphi_B(x)$ is independent of this representation for all $B \in \mathcal{B}$ since the Stieltjes transform uniquely determines even a complex measure. A similar argument also implies that, for fixed $B \in \mathcal{B}$, φ_B defines a linear functional on $\mathcal{A}$.

Since $v_{\sqrt{y}}(\mathbb{R}) = \langle y \rangle$ for $y \in \overline{\mathcal{A}}_+$, we obtain for any $x = (\operatorname{Re} x)_+ - (\operatorname{Re} x)_- + \mathrm{i}(\operatorname{Im} x)_+ - \mathrm{i}(\operatorname{Im} x)_- \in \mathcal{A}$ the bound

$$\begin{aligned} |\varphi_B(x)| &\leq v_{\sqrt{(\operatorname{Re} x)_+}}(\mathbb{R}) + v_{\sqrt{(\operatorname{Re} x)_-}}(\mathbb{R}) + v_{\sqrt{(\operatorname{Im} x)_+}}(\mathbb{R}) + v_{\sqrt{(\operatorname{Im} x)_-}}(\mathbb{R}) \\ &\leq \langle (\operatorname{Re} x)_+ + (\operatorname{Re} x)_- + (\operatorname{Im} x)_+ + (\operatorname{Im} x)_- \rangle \leq 2\|x\|_1, \end{aligned}$$

where we used that $(\operatorname{Re} x)_+ + (\operatorname{Re} x)_- = |\operatorname{Re} x|$ and $(\operatorname{Im} x)_+ + (\operatorname{Im} x)_- = |\operatorname{Im} x|$. Therefore, φ_B extends to a bounded linear functional on L^1 as $\mathcal{A}$ is a dense linear subspace of L^1. Using the isomorphism in (8.11.1), for each $B \in \mathcal{B}$, there exists a unique $v(B) \in \mathcal{A}$ such that

$$\varphi_B(x) = \langle v(B) x \rangle$$

for all $x \in \mathcal{A}$. For $y \in \mathcal{A}$, we conclude $v_y(B) = v_{\sqrt{yy^*}}(B) = \varphi_B(yy^*) = \langle y\,, v(B) y \rangle \geq 0$, where we used that $v_y = v_{\sqrt{yy^*}}$ since they have the same Stieltjes transform. Since $\langle v(B) y \rangle \geq 0$ for all $y \in \overline{\mathcal{A}}_+$, we have $v(B) \in \overline{\mathcal{A}}_+$ for all $B \in \mathcal{B}$. Moreover, $v_x = \langle x\,, v(\cdot) x \rangle$, in particular, $\langle x\,, v(\mathbb{R}) x \rangle = v_x(\mathbb{R}) = \langle x\,, x \rangle$, for all $x \in \mathcal{A}$. The polarization identity yields that v is an $\overline{\mathcal{A}}_+$-valued measure on $\mathcal{B}$ satisfying (8.3.6) and $v(\mathbb{R}) = \mathbb{1}$. This completes the proof of Lemma 8.3.1. □

Lemma 8.11.1 (Stieltjes transform inherits Hölder regularity)**.** *Let v be an $\overline{\mathcal{A}}_+$-valued measure on $\mathbb{R}$ and $h\colon \mathbb{H} \to \mathcal{A}$ be its Stieltjes transform, i.e., h satisfies* (8.3.6) *for all $z \in \mathbb{H}$. Let $f\colon I \to \overline{\mathcal{A}}_+$ be a γ-Hölder continuous function on an interval $I \subset \mathbb{R}$ with $\gamma \in (0,1)$ and f be a density of v on I with respect to the Lebesgue measure, i.e.,*

$$\|f(\tau_1) - f(\tau_2)\| \leq C_0 |\tau_1 - \tau_2|^\gamma, \qquad v(A) = \int_A f(\tau)\mathrm{d}\tau$$

for all $\tau_1, \tau_2 \in I$, some $C > 0$ and for all Borel sets $A \subset I$. Moreover, we assume that $\|f(\tau)\| \leq C_1$ for all $\tau \in I$. Let $\theta \in (0,1]$.

Then, for $z_1, z_2 \in \mathbb{H}$ satisfying $\operatorname{Re} z_1, \operatorname{Re} z_2 \in I$ *and* $\operatorname{dist}(\operatorname{Re} z_k, \partial I) \geq \theta$, *$k = 1, 2$, we have*

$$\|h(z_1) - h(z_2)\| \leq \Big(\frac{13 C_0}{\gamma(1-\gamma)} + \frac{14 C_1}{\gamma \theta^\gamma} + \frac{4\|v(\mathbb{R})\|}{\theta^{1+\gamma}} \Big) |z_1 - z_2|^\gamma. \tag{8.11.3}$$

Furthermore, for $z_1, z_2 \in \mathbb{H}$ satisfying $\operatorname{dist}(z_k, \operatorname{supp} v) \geq \theta$, *$k = 1, 2$, we have*

$$\|h(z_1) - h(z_2)\| \leq \frac{2\|v(\mathbb{R})\|}{\theta^2} |z_1 - z_2|^\gamma. \tag{8.11.4}$$

We remark that the proof of Lemma 8.11.1 is very similar to the proof of Lemma A.7 in [4]. Nevertheless, we present it here for the convenience of the reader.

Proof. We will prove (8.11.3) in two steps: First, we will estimate the left-hand side of (8.11.3) for $\operatorname{Im} z_1 = \operatorname{Im} z_2$ and then for $\operatorname{Re} z_1 = \operatorname{Re} z_2$. Combining the estimates in these two special cases, we will then conclude (8.11.3). We set $I_\theta := \{\tau \in I\colon\ \operatorname{dist}(\tau, \partial I) \geq \theta\}$, i.e., $I \supset I_\theta$.

In fact, for $\omega_1, \omega_2 \in I_\theta$ and $\eta > 0$, we now prove

$$\|h(\omega_1 + \mathrm{i}\eta) - h(\omega_2 + \mathrm{i}\eta)\| \leq \Big(\frac{10C_0}{\gamma(1-\gamma)} + \frac{10C_1}{\gamma\theta^\gamma} + \frac{2\|v(\mathbb{R})\|}{\theta^{1+\gamma}}\Big)|\omega_1 - \omega_2|^\gamma. \tag{8.11.5}$$

First, we conveniently decompose $h(\omega_2 + \mathrm{i}\eta) - h(\omega_1 + \mathrm{i}\eta)$. For $k = 1, 2$, we have

$$h(\omega_k + \mathrm{i}\eta) = \mathrm{i}\pi f(\omega_k) + \lim_{R\to\infty}\bigg(\int_{I\cap(\omega_1+[-R,R])}\frac{f(\tau)-f(\omega)}{\tau-\omega_k-\mathrm{i}\eta}\mathrm{d}\tau - \int_{(\omega_1+[-R,R])\setminus I}\frac{f(\omega_k)\mathrm{d}\tau}{\tau-\omega_k-\mathrm{i}\eta}\bigg) + \int_{\mathbb{R}\setminus I}\frac{v(\mathrm{d}\tau)}{\tau-\omega_k-\mathrm{i}\eta}.$$

Here, we used that

$$\lim_{R\to\infty}\int_{-R}^R\frac{1}{\tau-\mathrm{i}\eta}\,\mathrm{d}\tau = \mathrm{i}\pi, \qquad \lim_{R\to\infty}\int_{J_2\setminus J_1}\frac{f(\omega_2)}{\tau-z_2}\,\mathrm{d}\tau = \lim_{R\to\infty}\int_{J_1\setminus J_2}\frac{f(\omega_2)}{\tau-z_2}\,\mathrm{d}\tau = 0,$$

where $J_1 := \omega_1 + [-R, R]$ and $J_2 := \omega_2 + [-R, R]$. Thus, we obtain the decomposition

$$h(\omega_2 + \mathrm{i}\eta) - h(\omega_1 + \mathrm{i}\eta) = \mathrm{i}\pi(f(\omega_2) - f(\omega_1)) + \lim_{R\to\infty}(D_1 + \ldots + D_6) + D_7, \tag{8.11.6}$$

where we introduced

$$\begin{aligned}
D_k &:= (-1)^k\int_{I\cap J_1}\frac{f(\tau)-f(\omega_k)}{\tau-z_k}\mathbf{1}(|\tau-\omega_1|\leq|\omega_1-\omega_2|)\mathrm{d}\tau, \qquad k=1,2,\\
D_3 &:= \int_{I\cap J_1}(f(\tau)-f(\omega_2))\Big(\frac{1}{\tau-z_2}-\frac{1}{\tau-z_1}\Big)\mathbf{1}(|\tau-\omega_1|>|\omega_1-\omega_2|)\mathrm{d}\tau,\\
D_4 &:= (f(\omega_1)-f(\omega_2))\int_{J_1}\frac{1}{\tau-z_1}\mathbf{1}(|\tau-\omega_1|>|\omega_1-\omega_2|)\mathrm{d}\tau,\\
D_5 &:= \int_{J_1\setminus I}\frac{f(\omega_1)}{\tau-z_1}\mathbf{1}(|\tau-\omega_1|\leq|\omega_1-\omega_2|)\mathrm{d}\tau - \int_{J_1\setminus I}\frac{f(\omega_2)}{\tau-z_2}\mathbf{1}(|\tau-\omega_1|\leq|\omega_1-\omega_2|)\mathrm{d}\tau,\\
D_6 &:= -\int_{J_1\setminus I}f(\omega_2)\Big(\frac{1}{\tau-z_2}-\frac{1}{\tau-z_1}\Big)\mathbf{1}(|\tau-\omega_1|>|\omega_1-\omega_2|)\mathrm{d}\tau,\\
D_7 &:= \int_{\mathbb{R}\setminus I}\Big(\frac{1}{\tau-z_2}-\frac{1}{\tau-z_1}\Big)v(\mathrm{d}\tau).
\end{aligned}$$

We remark that $D_1, \ldots, D_6$ depend on R. However, since the following estimates on their norms will hold true uniformly for all large R, they will also hold true for the limes superior of these norms.

In order to estimate $\|D_1\|$ and $\|D_2\|$, we pull the norm inside the integral, use the Hölder-continuity of f, neglect all η's, extend the domain of integration from $I \cap J_1$ to $\mathbb{R}$ and compute the remaining integral. This yields

$$\|D_1\| \le \frac{2C_0}{\gamma}|\omega_1 - \omega_2|^\gamma, \qquad \|D_2\| \le \frac{2C_0}{\gamma}|\omega_1 - \omega_2|^\gamma.$$

For the estimate of $\|D_3\|$, we pull the norm inside the integral, disregard all η's in

$$\left|\frac{1}{\tau - \omega_2 - \mathrm{i}\eta} - \frac{1}{\tau - \omega_1 - \mathrm{i}\eta}\right| \le \frac{|\omega_1 - \omega_2|}{|\tau - \omega_1||\tau - \omega_2|},$$

use the Hölder-continuity of f and extend the domain of integration from $I \cap J_1$ to $\mathbb{R}$. We, thus, obtain

$$\|D_3\| \le C_0 \int_{\mathbb{R}} \frac{|\omega_2 - \omega_1| \mathbf{1}(|\tau - \omega_1| > |\omega_1 - \omega_2|)}{|\tau - \omega_1||\tau - \omega_1 - (\omega_2 - \omega_1)|^{1-\gamma}} \mathrm{d}\tau \le \frac{2C_0}{\gamma(1-\gamma)}|\omega_1 - \omega_2|^\gamma.$$

The real part of the integral in the definition of D_4 vanishes as J_1 and the argument of the characteristic function are symmetric around ω_1. Hence, since the imaginary part of the integral is bounded by π, the Hölder-continuity of f yields

$$\|D_4\| \le C_0 \pi |\omega_1 - \omega_2|^\gamma.$$

To bound $\|D_5\|$, we pull the norm inside of the integrals and use $\omega_1, \omega_2 \in I_\theta$ and $\tau \in \mathbb{R} \setminus I$ to see that θ is a lower bound on $|\tau - \omega_1|$ and $|\tau - \omega_2|$. Moreover, the characteristic function in the integrals yields upper bounds on $|\tau - \omega_1|$ and $|\tau - \omega_2|$, respectively. Hence, we obtain

$$\|D_5\| \le \frac{2\|f(\omega_1)\| + 2^{1+\gamma}\|f(\omega_2)\|}{\gamma\theta^\gamma}|\omega_1 - \omega_2|^\gamma.$$

We now bound $\|D_6\|$ and $\|D_7\|$. Computing the difference on the left-hand side, taking its absolute value to the power γ and using the triangle inequality for the modulus of difference to the power $1-\gamma$ as well as disregarding all η's yield

$$\Big|\frac{1}{\tau-\omega_1-\mathrm{i}\eta}-\frac{1}{\tau-\omega_2-\mathrm{i}\eta}\Big|\leq\frac{2^{1-\gamma}|\omega_2-\omega_1|^\gamma}{\min\{|\tau-\omega_1|,|\tau-\omega_2|\}^{1+\gamma}}.$$

Thus, we pull the norms inside the integrals in the definition of D_6 and D_7, respectively, use the previous bound as well as $\tau\in\mathbb{R}\setminus I$ and $\omega_i\in I_\theta$, i.e., $|\tau-\omega_i|\geq\theta$, and obtain

$$\|D_6\|\leq\frac{2^{2-\gamma}\|f(\omega_2)\|}{\gamma\theta^\gamma}|\omega_1-\omega_2|^\gamma,\qquad\qquad \|D_7\|\leq\frac{2^{1-\gamma}\|v(\mathbb{R})\|}{\theta^{1+\gamma}}|\omega_1-\omega_2|^\gamma.$$

Starting from (8.11.6) and using the Hölder continuity of f for the first term on the right-hand side of (8.11.6) as well as the previous estimates on $\|D_1\|,\dots,\|D_7\|$ complete the proof of (8.11.5).

We now establish the second special case. For $\omega\in I_\theta$, and $\eta_1,\eta_2>0$, we now show the bound

$$\|h(\omega+\mathrm{i}\eta_1)-h(\omega+\mathrm{i}\eta_2)\|\leq\Big(\frac{\sqrt{8}C_0}{\gamma(1-\gamma)}+\frac{4C_1}{\gamma\theta^\gamma}+\frac{2\|v(\mathbb{R})\|}{\theta^{1+\gamma}}\Big)|\eta_1-\eta_2|^\gamma. \tag{8.11.7}$$

Similarly to the proof of (8.11.6), we obtain the decomposition

$$h(\omega+\mathrm{i}\eta_2)-h(\omega+\mathrm{i}\eta_1)=E_1+E_2+E_3,$$

where we introduced

$$\begin{aligned}E_1&:=\int_I(f(\tau)-f(\omega))\Big(\frac{1}{\tau-\omega-\mathrm{i}\eta_2}-\frac{1}{\tau-\omega-\mathrm{i}\eta_1}\Big)\mathrm{d}\tau,\\ E_2&:=\int_{\mathbb{R}\setminus I}f(\omega)\Big(\frac{1}{\tau-\omega-\mathrm{i}\eta_2}-\frac{1}{\tau-\omega-\mathrm{i}\eta_1}\Big)\mathrm{d}\tau,\\ E_3&:=\int_{\mathbb{R}\setminus I}\Big(\frac{1}{\tau-\omega-\mathrm{i}\eta_2}-\frac{1}{\tau-\omega-\mathrm{i}\eta_1}\Big)v(\mathrm{d}\tau).\end{aligned}$$

Next, we verify the following bounds

$$\begin{aligned}\|E_1\|&\leq\frac{\sqrt{8}C_0}{\gamma(1-\gamma)}|\eta_2-\eta_1|^\gamma,\\ \|E_2\|&\leq\frac{2^{2-\gamma}}{\gamma\theta^\gamma}\|f(\omega)\||\eta_2-\eta_1|^\gamma,\\ \|E_3\|&\leq\frac{2^{1-\gamma}}{\theta^{1+\gamma}}\|v(\mathbb{R})\||\eta_2-\eta_1|^\gamma.\end{aligned}\tag{8.11.8}$$

We, thus, notice that (8.11.7) is proven once the estimates in (8.11.8) are established. Since

$$E_1 = \mathrm{i} \int_I \frac{(\eta_2 - \eta_1)(f(\tau) - f(\omega))}{(\tau - \omega - \mathrm{i}\eta_1)(\tau - \omega - \mathrm{i}\eta_2)} \mathrm{d}\tau$$

we obtain

$$\begin{aligned} \|E_1\| \leq C_0 \int_{\mathbb{R}} \frac{|\eta_2 - \eta_1|}{|\tau - \omega|^{1-\gamma} \frac{1}{\sqrt{2}}(|\tau - \omega| + |\eta_2 - \eta_1|)} \mathrm{d}\tau &\leq 2\sqrt{2} C_0 \int_0^\infty \frac{|\eta_2 - \eta_1| \mathrm{d}x}{x^{1-\gamma}(x + |\eta_2 - \eta_1|)} \\ &\leq \frac{\sqrt{8} C_0}{\gamma(1-\gamma)} |\eta_2 - \eta_1|^\gamma. \end{aligned}$$

For the remaining estimates in the proof of (8.11.8), we remark that

$$\left| \frac{1}{\tau - \omega - \mathrm{i}\eta_2} - \frac{1}{\tau - \omega - \mathrm{i}\eta_1} \right| \leq \frac{|\eta_2 - \eta_1|^\gamma}{|\tau - \omega|^{2\gamma}} \frac{2^{1-\gamma}}{|\tau - \omega|^{1-\gamma}} \leq \frac{2^{1-\gamma}}{\theta^{1+\gamma}} |\eta_2 - \eta_1|^\gamma. \tag{8.11.9}$$

Applying the second bound in (8.11.9) to the definition of E_2 yields

$$\|E_2\| \leq 2^{1-\gamma} \|f(\omega)\| |\eta_2 - \eta_1|^\gamma \int_{\mathbb{R} \setminus I} \frac{1}{|\tau - \omega|^{1+\gamma}} \mathrm{d}\tau \leq \frac{2^{2-\gamma}}{\gamma \theta^\gamma} \|f(\omega)\| |\eta_2 - \eta_1|^\gamma,$$

which implies the second bound in (8.11.8). Similarly, we apply the third bound in (8.11.9) to the definition of E_3 and conclude

$$\|E_3\| \leq \frac{2^{1-\gamma}}{\theta^{1+\gamma}} \|v(\mathbb{R})\| |\eta_2 - \eta_1|^\gamma.$$

This completes the proof of (8.11.8) and, hence, the one of (8.11.7) as well. By combining (8.11.5) and (8.11.7), we obtain (8.11.3).

The bound in (8.11.4) is a trivial consequence of

$$\left| \frac{1}{\tau - z_1} - \frac{1}{\tau - z_2} \right| \leq \frac{2^{1-\gamma} |z_1 - z_2|^\gamma}{\min\{|\tau - z_1|, |\tau - z_2|\}^{1+\gamma}} \leq \frac{2}{\theta^2} |z_1 - z_2|^\gamma,$$

where we used $\tau \in \operatorname{supp} v$ and $\operatorname{dist}(z_k, \operatorname{supp} v) \geq \theta$ for $k = 1, 2$. This completes the proof of Lemma 8.11.1. ☐

8.12. Positivity-preserving, symmetric operators on $\mathcal{A}$

Lemma 8.12.1. *Let $T\colon \mathcal{A} \to \mathcal{A}$ be a positivity-preserving, symmetric operator.*

(i) If $T[a] \leq C\langle a \rangle \mathbb{1}$ for some $C > 0$ and all $a \in \overline{\mathcal{A}}_+$ then $\|T\|_2 \leq 2C$. Moreover, $\|T\|_2$ is an eigenvalue of T and there is $x \in \overline{\mathcal{A}}_+ \setminus \{0\}$ such that $T[x] = \|T\|_2 x$.

(ii) *We assume* $\|T\|_2 = 1$ *and that there are* $c, C > 0$ *such that*

$$c\langle a\rangle \mathbb{1} \leq T[a] \leq C\langle a\rangle \mathbb{1} \tag{8.12.1}$$

for all $a \in \mathcal{A}_+$. *Then* 1 *is an eigenvalue of* T *with a one-dimensional eigenspace. There is a unique* $x \in \mathcal{A}_+$ *satisfying* $T[x] = x$ *and* $\|x\|_2 = 1$. *Moreover,* x *is positive definite,*

$$cC^{-1/2}\mathbb{1} \leq x \leq C\mathbb{1}. \tag{8.12.2}$$

Furthermore, the spectrum of T *has a gap of size* $\theta := c^6/(2(c^3 + 2C^2)C^2))$, *i.e.,*

$$\mathrm{Spec}(T) \subset [-1+\theta, 1-\theta] \cup \{1\}. \tag{8.12.3}$$

Lemma 8.12.1 is the analogue of Lemma 4.8 in [6]. Here, we explain how to generalize it to the context of von Neumann algebras. In the proof of Lemma 8.12.1, we will use the following lemma. We omit its proof since the first part is obtained as in (4.2) of [6] and the second part as in (5.28) of [4].

Lemma 8.12.2. *Let* $T\colon \mathcal{A} \to \mathcal{A}$ *be a linear map.*

(i) *If* T *is positivity-preserving such that* $T[a] \leq C\langle a\rangle \mathbb{1}$ *for all* $a \in \mathcal{A}_+$ *and some* $C > 0$ *then* $\|T\| \leq \|T\|_{2\to\|\cdot\|} \leq 2C$.

(ii) *If* $T - \omega\mathrm{Id}$ *is invertible on* $\mathcal{A}$ *for some* $\omega \in \mathbb{C}\setminus\{0\}$ *and* $\|(T-\omega\mathrm{Id})^{-1}\|_2 < \infty$, $\|T\|_{2\to\|\cdot\|} < \infty$ *then we have*

$$\|(T-\omega\mathrm{Id})^{-1}\| \leq |\omega|^{-1}\Big(1 + \|T\|_{2\to\|\cdot\|}\|(T-\omega\mathrm{Id})^{-1}\|_2\Big).$$

Proof of Lemma 8.12.1. For the proof of (i), we remark that Lemma 8.12.2 (i) implies $\|T\|_2 \leq \|T\|_{2\to\|\cdot\|} \leq 2C$. Without loss of generality, we assume $\|T\|_2 = 1$. Since T is positivity-preserving, we have $T[b] \in \mathcal{A}_{\mathrm{sa}}$ for all $b \in \mathcal{A}_{\mathrm{sa}}$. It is easy to check that, for each $a \in \mathcal{A}$, one may find $b \in \mathcal{A}_{\mathrm{sa}}$ such that $\|a\|_2 = \|b\|_2$ and $\|T[a]\|_2 \leq \|T[b]\|_2$. Hence, $\|T|_{\mathcal{A}_{\mathrm{sa}}}\|_2 = \|T\|_2 = 1$ and 1 is contained in the spectrum of $T\colon L^2_{\mathrm{sa}} \to L^2_{\mathrm{sa}}$, where $L^2_{\mathrm{sa}} := \overline{\mathcal{A}_{\mathrm{sa}}}^{\|\cdot\|_2}$, due to the variational principle for the spectrum of self-adjoint operators and $|\langle b, T[b]\rangle| \leq \langle |b|, T[|b|]\rangle$ for all $b \in \mathcal{A}_{\mathrm{sa}}$. This last inequality can be checked easily by decomposing $b = b_+ - b_-$ into positive and negative part.

Hence, due to the symmetry of T, there is a sequence $(y_n)_n$ of approximating eigenvectors in $\mathcal{A}_{\mathrm{sa}}$, i.e., $y_n \in \mathcal{A}_{\mathrm{sa}}$, $\|y_n\|_2 = 1$ and $T[y_n] - y_n$ converges to 0 in L^2 for $n \to \infty$. We set $x_n := |y_n|$. By using $\|T|_{L^2_{\mathrm{sa}}}\|_2 = 1$ and $\langle b\,, T[b]\rangle \leq \langle |b|\,, T[|b|]\rangle$ for all $b \in \mathcal{A}_{\mathrm{sa}}$, we obtain $\|T[x_n] - x_n\|_2^2 \leq 2\|y_n\|_2\|T[y_n] - y_n\|_2$ and, thus,

$$\lim_{n\to\infty} \|T[x_n] - x_n\|_2 = 0. \tag{8.12.4}$$

Since the unit ball in the Hilbert space L^2 is relatively sequentially compact in the weak topology, we can assume by possibly replacing $(x_n)_n$ by a subsequence that there is $x \in L^2$ such that $x_n \rightharpoonup x$ weakly in L^2. From $T[x_n] \leq C\langle x_n\rangle \mathbb{1}$, we conclude

$$x_n \leq (\mathrm{Id} - T)[x_n] + C\langle x_n\rangle \mathbb{1}.$$

Multiplying this by $\sqrt{x_n}$ from the left and the right and applying $\langle\,\cdot\,\rangle$ yields

$$1 \leq \langle x_n\,, (\mathrm{Id} - T)[x_n]\rangle + C\langle x_n\rangle^2.$$

Taking the limit $n \to \infty$, we obtain $\langle x\rangle \geq C^{-1/2}$, due to (8.12.4). Hence, $x \neq 0$ and we can replace x by $x/\|x\|_2$ and x_n by $x_n/\|x\|_2$. For any $b \in L^2$, we have

$$\langle b, (\mathrm{Id} - T)[x]\rangle = \lim_{n\to\infty} \langle b, (\mathrm{Id} - T)[x_n]\rangle = 0$$

due to $x_n \rightharpoonup x$ and (8.12.4). Hence, $T[x] = x$. Since $\|T\|_{2\to\|\cdot\|} \leq 2C$, we have $T[b] \in \mathcal{A}$ for all $b \in L^2$ and thus $x = T[x] \in \mathcal{A}$. Owing to $x_n \rightharpoonup x$ and $x_n \in \overline{\mathcal{A}}_+$, we obtain $x \in \overline{\mathcal{A}}_+$. This completes the proof of (i).

We start the proof of (ii) by using (8.12.1) with $a = x$ which immediately yields the upper bound in (8.12.2). As $\langle x\rangle \geq C^{-1/2}$, the first inequality in (8.12.1) then yields the lower bound in (8.12.2).

In order to prove the spectral gap, (8.12.3), we remark that $\|T\|_{2\to\|\cdot\|} \leq 2C$ due to the upper bound in (8.12.1) and Lemma 8.12.2 (i). Hence, by Lemma 8.12.2 (ii), the spectrum of T as an operator on $\mathcal{A}$ is contained in the union of $\{0\}$ and the spectrum of T as an operator on L^2. Therefore, we will consider T as an operator on L^2 in the following and exclusively study its spectrum as an operator on L^2. Hence, to prove the spectral gap, it suffices to establish a lower bound on $\langle y\,, (\mathrm{Id} \pm T)[y]\rangle$ for all self-adjoint

$y \in \mathcal{A}$ satisfying $\|y\|_2 = 1$ and $\langle x\,, y\rangle = 0$. Fix such $y \in \mathcal{A}$. Since y is self-adjoint we have

$$y = \lim_{N\to\infty} y^N, \qquad y^N := \sum_{k=1}^{N} \lambda_k^N p_k^N \tag{8.12.5}$$

for some $\lambda_n^N \in \mathbb{R}$ and $p_k^N \in \mathcal{A}$ orthogonal projections such that $p_k^N p_l^N = p_k^N \delta_{k,l}$. Here, the convergence $y^N \to y$ is with respect to $\|\cdot\|$. We can assume that $\|y^N\|_2 = 1$ for all N as well as $\langle p_k^N\rangle > 0$ for all k and $\langle p_1^N + \ldots + p_N^N\rangle = 1$ for all N.

We will now reduce estimating $\langle y\,, (\mathrm{Id} \pm T)[y]\rangle$ to estimating a scalar product on $\mathbb{C}^N$. On $\mathbb{C}^N$, we consider the scalar product $\langle\,\cdot\,,\,\cdot\,\rangle_N$ induced by the probability measure $\pi(A) = \sum_{k\in A}\langle p_k^N\rangle$ on $[N]$, i.e.,

$$\langle \lambda\,, \mu\rangle_N = \sum_{k=1}^{n} \overline{\lambda_k}\mu_k \langle p_k^N\rangle$$

for $\lambda = (\lambda_k)_{k=1}^N$, $\mu = (\mu_k)_{k=1}^N \in \mathbb{C}^N$. The norm on $\mathbb{C}^N$ and the operator norm on $\mathbb{C}^{N\times N}$ induced by $\langle\,\cdot\,,\,\cdot\,\rangle_N$ are denoted by $\|\cdot\|_N$ and $\|\cdot\|$, respectively. Moreover, Id_N is the identity map on $\mathbb{C}^N$. With this notation, we obtain from (8.12.5) that

$$\langle y\,, (\mathrm{Id} \pm T)[y]\rangle = \lim_{N\to\infty} \sum_{k,l=1}^{N} \lambda_k^N \lambda_l^N \langle p_k^N\,, (\mathrm{Id} \pm T)[p_l^N]\rangle = \lim_{N\to\infty} \langle \lambda^N\,, (\mathrm{Id}_N \pm S^N)[\lambda^N]\rangle_N,$$

where we introduced $\lambda^N = (\lambda_k^N)_{k=1}^N \in \mathbb{C}^N$ and the $N \times N$ symmetric matrix S^N viewed as an integral operator on $([N], \pi)$ with the kernel s_{kl}^N given by

$$s_{kl}^N = \frac{\langle p_k^N\,, T[p_l^N]\rangle}{\langle p_k^N\rangle\langle p_l^N\rangle}.$$

Since $\|y^N\|_2 = 1$, we have $\|\lambda^N\|_N = 1$. By the flatness of T, we have

$$c \le s_{kl}^N \le C. \tag{8.12.6}$$

In the following, we will omit the N-dependence of λ_k, s_{kl} and p_k from our notation. By the definition of $\langle\cdot,\cdot\rangle_N$, we have

$$\langle \lambda\,, S\lambda\rangle_N = \sum_{k,l=1}^{N} \lambda_k \langle p_k\rangle s_{kl} \langle p_l\rangle \lambda_l = \langle y^N\,, T[y^N]\rangle.$$

Let $s \in \mathbb{C}^N$ be the Perron-Frobenius eigenvector of S satisfying $Ss = \|S\|s$, $\|s\|_N = 1$. From (8.12.6), we conclude

$$c \le \langle e\,, Se\rangle_N \le \|S\| = \langle s\,, Ss\rangle_N \le \|T\|_2 = 1, \tag{8.12.7}$$

where $e = (1, \dots, 1) \in \mathbb{C}^N$. Since $\|s\|_N = 1$ and $c \le \|S\|$, we have

$$\max_i s_i = \frac{(Ss)_i}{\|S\|} \le \frac{C}{c}\sum_{k=1}^N s_k \langle p_k\rangle \le \frac{C}{c}\left(\sum_{k=1}^N \langle p_k\rangle\right)^{1/2}\left(\sum_{k=1}^N s_k^2 \langle p_k\rangle\right)^{1/2} = \frac{C}{c}.$$

As $\inf_{k,l} s_{k,l} \ge c$ by (8.12.6), Lemma 5.7 in [4] yields

$$\mathrm{Spec}(S) \subset \left[-\|S\| + \frac{c^3}{C^2}, \|S\| - \frac{c^3}{C^2}\right] \cup \{\|S\|\}.$$

We decompose $\lambda = (1 - \|w\|_N^2)^{1/2} s + w$ with $w \perp s$ and obtain

$$|\langle \lambda\,, S\lambda\rangle_N| \le \|S\|(1 - \|w\|_N^2) + \left(\|S\| - \frac{c^3}{C^2}\right)\|w\|_N^2 \le 1 - \frac{c^3}{C^2}\|w\|_N^2, \tag{8.12.8}$$

where we used $\|S\| \le 1$ in the last step. Hence, it remains to estimate $\|w\|_N$.

Recalling $T[x] = x$, we set $\tilde{x} = (\langle xp_k\rangle / \langle p_k\rangle)_{k=1}^N$ and compute

$$\langle x\,, y^N\rangle = \sum_k \lambda_k \langle xp_k\rangle = \langle \tilde{x}\,, \lambda\rangle_N.$$

Since the left-hand side goes to $\langle x\,, y\rangle = 0$ for $N \to \infty$, we can assume that $|\langle \tilde{x}\,, \lambda\rangle_N| \le \sqrt{\varepsilon/2}$ for any fixed $\varepsilon \sim 1$ and all sufficiently large N. As $\tilde{x}_k \ge c/\sqrt{C}$ by (8.12.2), we obtain

$$\begin{aligned}(1 - \|w\|_N^2)\frac{c^2}{C}\left(\sum_k s_k \langle p_k\rangle\right)^2 &\le (1 - \|w\|_N^2)\langle \tilde{x}\,, s\rangle_N^2 = (\langle \tilde{x}\,, \lambda\rangle_N - \langle \tilde{x}\,, w\rangle_N)^2 \\ &\le 2\|\tilde{x}\|_N^2 \|w\|_N^2 + \varepsilon.\end{aligned} \tag{8.12.9}$$

Now, we use $c \le \langle s\,, Ss\rangle_N$ from (8.12.7) to get

$$c \le \langle s\,, Ss\rangle_N = \sum_{k,l} s_k s_{kl} s_l \langle p_k\rangle\langle p_l\rangle \le C\left(\sum_k s_k\langle p_k\rangle\right)^2.$$

By plugging this and $\|\tilde{x}\|_N^2 \le \|x\|^2 \sum_k \langle p_k \rangle = 1$ into (8.12.9), solving the resulting estimate for $\|w\|_N^2$ and choosing $\varepsilon = c^3/(2C^2)$, we obtain

$$\|w\|_N^2 \ge \frac{c^3}{2(c^3 + 2C^2)}.$$

Therefore, from (8.12.8), we conclude

$$|\langle \lambda\,, S\lambda \rangle_N| \le 1 - \frac{c^6}{2(c^3 + 2C^2)C^2}$$

uniformly for all sufficiently large $N \in \mathbb{N}$. We thus obtain that

$$\langle y\,, (\mathrm{Id} \pm T)[y] \rangle \ge \frac{c^6}{2(c^3 + 2C^2)C^2}$$

if $y \perp x$ and $\|y\|_2 = 1$. We conclude (8.12.3), which completes the proof of the lemma. □

Lemma 8.12.3. *If $T \colon \mathcal{A} \to \mathcal{A}$ is a positivity-preserving operator such that $\|T\|_2 < 1$ and $\|T\|_{2 \to \|\cdot\|} < \infty$ then* $\mathrm{Id} - T$ *is invertible as a bounded operator on $\mathcal{A}$ and* $(\mathrm{Id} - T)^{-1}$ *is positivity-preserving with*

$$(\mathrm{Id} - T)^{-1}[x^* x] \ge x^* x \tag{8.12.10}$$

for all $x \in \mathcal{A}$.

Proof. Since $\|T\|_2 < 1$, $\mathrm{Id} - T$ is invertible on L^2 and we conclude the invertibility of $\mathrm{Id} - T$ on $\mathcal{A}$ from Lemma 8.12.2 (ii).

Moreover, for $y \in \mathcal{A}$ with $\|y^* y\|_2 < 1$, we expand the inverse as a Neumann series using $\|T\|_2 < 1$ and obtain

$$(\mathrm{Id} - T)^{-1}[y^* y] = y^* y + \Big(\sum_{k=1}^{\infty} T^k [y^* y] \Big) \ge y^* y.$$

The series converges with respect to $\|\cdot\|_2$. In the last inequality, we used that T^k is a positivity-preserving operator for all $k \in \mathbb{N}$. Hence, by rescaling a general $x \in \mathcal{A}$, we see that $(\mathrm{Id} - T)^{-1}$ is a positivity-preserving operator on $\mathcal{A}$ which satisfies (8.12.10). □

8.13. Non-Hermitian perturbation theory

Let $B_0\colon \mathcal{A} \to \mathcal{A}$ be a bounded operator with an isolated, single eigenvalue β_0 and an associated eigenvector b_0, $\|b_0\|_2 = 1$, i.e.,

$$B_0[b_0] = \beta_0 b_0.$$

Moreover, we denote by P_0 and Q_0 the spectral projections corresponding to β_0 and $\mathrm{Spec}(B_0) \setminus \{\beta_0\}$. Note that $P_0 + Q_0 = \mathrm{Id}$ but they are not orthogonal projections in general. If l_0 is a normalized eigenvector of B_0^* associated to its eigenvalue $\overline{\beta}_0$, then we obtain

$$P_0 = \frac{\langle l_0\,,\,\cdot\,\rangle}{\langle l_0\,, b_0\rangle} b_0. \tag{8.13.1}$$

For some bounded operator $E\colon \mathcal{A} \to \mathcal{A}$, we consider the perturbation

$$B = B_0 + E.$$

We assume E to be sufficiently small such that there is an isolated, single eigenvalue β of B close to β_0 and that β and β_0 are separated from $\mathrm{Spec}(B) \setminus \{\beta\}$ and $\mathrm{Spec}(B_0) \setminus \{\beta_0\}$ by an amount $\Delta > 0$. Let P be the spectral projection of B associated to β.

Lemma 8.13.1. *We define $b := P[b_0]$ and $l := P^*[l_0]$. Then b and l are eigenvectors of B and B^* corresponding to β and $\bar{\beta}$, respectively. Moreover, we have*

$$b = b_0 + b_1 + b_2 + \mathcal{O}(\|E\|^3), \qquad l = l_0 + l_1 + l_2 + \mathcal{O}(\|E\|^3), \tag{8.13.2}$$

where we introduced

$$\begin{aligned}
b_1 &= -\,Q_0(B_0 - \beta_0\mathrm{Id})^{-1}E[b_0],\\
b_2 &= Q_0(B_0 - \beta_0\mathrm{Id})^{-1}E(B_0 - \beta_0\mathrm{Id})^{-1}Q_0E[b_0] - Q_0(B_0 - \beta_0\mathrm{Id})^{-2}EP_0E[b_0]\\
&\quad - P_0EQ_0(B_0 - \beta_0\mathrm{Id})^{-2}E[b_0],\\
l_1 &= -\,Q_0^*(B_0^* - \bar{\beta}_0\mathrm{Id})^{-1}E^*[l_0],\\
l_2 &= Q_0^*(B_0^* - \bar{\beta}_0\mathrm{Id})^{-1}E^*(B_0^* - \bar{\beta}_0\mathrm{Id})^{-1}Q_0^*E^*[l_0] - Q_0^*(B_0^* - \overline{\beta}_0\mathrm{Id})^{-2}E^*P_0^*E^*[l_0]\\
&\quad - P_0^*E^*Q_0^*(B_0^* - \overline{\beta}_0\mathrm{Id})^{-2}E^*[l_0].
\end{aligned}$$

In particular, we have $b_i, l_i = \mathcal{O}(\|E\|^i)$ for $i = 1, 2$. Furthermore, we obtain

$$\beta\langle l\,, b\rangle = \beta_0\langle l_0\,, b_0\rangle + \langle l_0\,, E[b_0]\rangle - \langle l_0\,, EB_0(B_0 - \beta_0\mathrm{Id})^{-2}Q_0E[b_0]\rangle + \mathcal{O}(\|E\|^3). \quad (8.13.3)$$

The implicit constants in the error terms depend only on the separation Δ.

Proof. In this proof, the difference $B - \omega$ with an operator B and a scalar ω is understood as $B - \omega\mathrm{Id}$. We first prove that

$$P = P_0 + P_1 + P_2 + \mathcal{O}(\|E\|^3), \quad (8.13.4)$$

where we defined

$$\begin{aligned} P_1 &:= -\frac{Q_0}{B_0 - \beta_0}EP_0 - P_0E\frac{Q_0}{B_0 - \beta_0}, \\ P_2 &:= P_0E\frac{Q_0}{B_0 - \beta_0}E\frac{Q_0}{B_0 - \beta_0} + \frac{Q_0}{B_0 - \beta_0}EP_0E\frac{Q_0}{B_0 - \beta_0} + \frac{Q_0}{B_0 - \beta_0}E\frac{Q_0}{B_0 - \beta_0}EP_0 \\ &\quad - \frac{Q_0}{(B_0 - \beta_0)^2}EP_0EP_0 - P_0E\frac{Q_0}{(B_0 - \beta_0)^2}EP_0 - P_0EP_0E\frac{Q_0}{(B_0 - \beta_0)^2}. \end{aligned}$$

The analytic functional calculus yields that

$$\begin{aligned} P &= -\frac{1}{2\pi\mathrm{i}}\oint_\Gamma \frac{1}{B - \omega}\mathrm{d}\omega \\ &= \frac{1}{2\pi\mathrm{i}}\oint_\Gamma \Big(-\frac{1}{B_0 - \omega} + \frac{1}{B_0 - \omega}E\frac{1}{B_0 - \omega} - \frac{1}{B_0 - \omega}E\frac{1}{B_0 - \omega}E\frac{1}{B_0 - \omega}\Big)\mathrm{d}\omega \quad (8.13.5) \\ &\qquad + \mathcal{O}(\|E\|^3), \end{aligned}$$

where Γ is a closed path that encloses only β and β_0 both with winding number $+1$ but no other element of the spectra of B and B_0. Integrating the first summand in the integrand of (8.13.5) yields P_0. In the second and third summand, we expand $\mathrm{Id} = P_0 + Q_0$ in the numerators. Applying an analogue of the residue theorem yields P_1 and P_2 for the second and third summand, respectively. For example, for the second summand, we obtain

$$P_1 = \frac{1}{2\pi\mathrm{i}}\oint_\Gamma \frac{1}{B_0 - \omega}E\frac{1}{B_0 - \omega}\mathrm{d}\omega = -\frac{Q_0}{B_0 - \beta_0}EP_0 - P_0E\frac{Q_0}{B_0 - \beta_0}.$$

The other two combinations of P_0, Q_0 vanish. Using a similar expansion for the third term, we get (8.13.4).

Starting from (8.13.4) as well as observing $b_i = P_i[b_0]$ and $l_i = P_i^*[l_0]$ for $i = 1, 2$, the relations (8.13.2) are a direct consequence of the definitions $b = P[b_0]$ and $l = P^*[l_0]$ and (8.13.1).

We will show below that

$$BP = B_0P_0 + B_1 + B_2 + \mathcal{O}(\|E\|^3), \tag{8.13.6}$$

where we defined

$$\begin{aligned}
B_1 &:= P_0EP_0 - \beta_0\Big(\frac{Q_0}{B_0-\beta_0}EP_0 + P_0E\frac{Q_0}{B_0-\beta_0}\Big),\\
B_2 &:= \beta_0\Big(P_0E\frac{Q_0}{B_0-\beta_0}E\frac{Q_0}{B_0-\beta_0} + \frac{Q_0}{B_0-\beta_0}EP_0E\frac{Q_0}{B_0-\beta_0} + \frac{Q_0}{B_0-\beta_0}E\frac{Q_0}{B_0-\beta_0}EP_0\Big)\\
&\quad - \frac{B_0Q_0}{(B_0-\beta_0)^2}EP_0EP_0 - P_0E\frac{B_0Q_0}{(B_0-\beta_0)^2}EP_0 - P_0EP_0E\frac{B_0Q_0}{(B_0-\beta_0)^2}.
\end{aligned}$$

Now, we obtain (8.13.3) by applying (8.13.2) as well as (8.13.6) to $\beta\langle l\,, b\rangle = \langle l\,, BPb\rangle$.

In order to prove (8.13.6), we use the analytic functional calculus with Γ as defined above to obtain

$$\begin{aligned}
BP &= -\frac{1}{2\pi\mathrm{i}}\oint_\Gamma \frac{\omega}{B-\omega}\mathrm{d}\omega\\
&= \frac{1}{2\pi\mathrm{i}}\oint_\Gamma \omega\Big(-\frac{1}{B_0-\omega} + \frac{1}{B_0-\omega}E\frac{1}{B_0-\omega} - \frac{1}{B_0-\omega}E\frac{1}{B_0-\omega}E\frac{1}{B_0-\omega}\Big)\mathrm{d}\omega\\
&\qquad + \mathcal{O}(\|E\|^3).
\end{aligned}$$

Proceeding similarly as in the proof of (8.13.4) yields (8.13.6) and thus completes the proof of Lemma 8.13.1. □

8.14. Characterization of $\operatorname{supp}\rho$

The following lemma gives equivalent characterizations of $\operatorname{supp}\rho$ in terms of m. Note $\operatorname{supp}\rho = \operatorname{supp} v$ due to the faithfulness of $\langle\,\cdot\,\rangle$. We denote the disk of radius $\varepsilon > 0$ centered at $z \in \mathbb{C}$ by $D_\varepsilon(z) := \{w \in \mathbb{C}\colon |z-w| < \varepsilon\}$.

Lemma 8.14.1 (Behaviour of m on $\mathbb{R}\setminus\operatorname{supp}\rho$). *Let m be the solution of the Dyson equation,* (8.2.3)*, for a data pair $(a, S) \in \mathcal{A}_{\mathrm{sa}} \times \Sigma$ with $\|a\| \le k_0$ and $S[x] \le k_1\langle x\rangle\mathbb{1}$ for* *all $x \in \overline{\mathcal{A}}_+$ and some $k_0, k_1 > 0$. Then, for any fixed $\tau \in \mathbb{R}$, the following statements are equivalent:*

(i) There is $c > 0$ such that

$$\limsup_{\eta\downarrow 0} \eta \|\operatorname{Im} m(\tau + \mathrm{i}\eta)\|^{-1} \geq c.$$

(ii) There are $C > 0$ and $N \subset (0, 1]$ with an accumulation point 0 such that

$$\begin{aligned} &\|m(z)\| \leq C, && \|m(z)^{-1}\| \leq C, \\ &C^{-1}\langle \operatorname{Im} m(z)\rangle \mathbb{1} \leq \operatorname{Im} m(z) \leq C\langle \operatorname{Im} m(z)\rangle \mathbb{1}, && \|F(z)\|_2 \leq 1 - C^{-1} \end{aligned} \tag{8.14.1}$$

for all $z \in \tau + \mathrm{i}N$. (The definition of F was given in (8.3.4).)

(iii) There is $m = m^* \in \mathcal{A}$ such that

$$\lim_{\eta\downarrow 0} \|m(\tau + \mathrm{i}\eta) - m\| = 0. \tag{8.14.2}$$

Moreover, there is $C > 0$ such that $\|m\| \leq C$ and $\|(\mathrm{Id} - C_m S)^{-1}\| \leq C$.

(iv) There are $\varepsilon > 0$ and an analytic function $f\colon D_\varepsilon(\tau) \to \mathcal{A}$ such that $f(z) = m(z)$ for all $z \in D_\varepsilon(\tau) \cap \mathbb{H}$ and $f(z) = f(\bar{z})^*$ for all $z \in D_\varepsilon(\tau)$. In particular, $f(z) = f(z)^*$ for $z \in D_\varepsilon(\tau) \cap \mathbb{R}$.

In other words, m can be analytically extended to a neighbourhood of τ.

(v) There is $\varepsilon > 0$ such that $\operatorname{dist}(\tau, \operatorname{supp}\rho) = \operatorname{dist}(\tau, \operatorname{supp} v) \geq \varepsilon$.

(vi) There is $c > 0$ such that

$$\liminf_{\eta\downarrow 0} \eta \|\operatorname{Im} m(\tau + \mathrm{i}\eta)\|^{-1} \geq c.$$

All constants in (i) – (vi) depend effectively on each other as well as possibly k_0, k_1 and an upper bound on $|\tau|$. For example, in the implication (iii) ⇒ (v), ε in (v) can be chosen to depend only on k_1 and C in (iii).

We remark that m in (iii) above is invertible and satisfies (8.2.3) at $z = \tau$.

As a direct consequence of the equivalence of (i) and (v), we spell out the following simple characterization of $\operatorname{supp}\rho$.

Corollary 8.14.2 (Characterization of $\operatorname{supp}\rho$)**.** *Under the conditions of Lemma 8.14.1, we have*

$$\lim_{\eta\downarrow 0}\eta\|\operatorname{Im} m(\tau+\mathrm{i}\eta)\|^{-1}=0. \tag{8.14.3}$$

if and only if $\tau\in\operatorname{supp}\rho(=\operatorname{supp}v)$.

Remark 8.14.3. In the proof of Lemma 8.14.1, the condition $S[x]\le k_1\langle x\rangle\mathbb{1}$ for all $x\in\overline{\mathcal{A}}_+$ is only used to guarantee the following two weaker consequences: First, this condition implies $\|S\|_{2\to\|\cdot\|}\le 2k_1$. Moreover, this condition yields, by Lemma 8.12.1 (i), that $F=F(\tau+\mathrm{i}\eta)$ has an eigenvector $f\in\overline{\mathcal{A}}_+$ corresponding to $\|F\|_2$, $Ff=\|F\|_2 f$, for any fixed $\tau\in\mathbb{R}\backslash\operatorname{supp}\rho$ and any $\eta\in(0,1]$. If both of these consequences are verified, then the condition $S[x]\le k_1\langle x\rangle\mathbb{1}$ may be dropped from Lemma 8.14.1 without any changes in the proof.

For the proof of Lemma 8.14.1, we need the following quantitative version of the implicit function theorem.

Lemma 8.14.4 (Quantitative implicit function theorem)**.** *Let X,Y,Z be Banach spaces, $U\subset X$ and $V\subset Y$ open subsets with $0\in U,V$. Let $\Phi\colon U\times V\to Z$ be continuously Fréchet-differentiable map such that the derivative $\partial_1\Phi(0,0)$ with respect to the first variable has a bounded inverse in the origin and $\Phi(0,0)=0$. Let $\delta>0$ such that $B_\delta^X\subset U$, $B_\delta^Y\subset V$ and*

$$\sup_{(x,y)\in B_\delta^X\times B_\delta^Y}\|\mathrm{Id}_X-(\partial_1\Phi(0,0))^{-1}\partial_1\Phi(x,y)\|\le\frac{1}{2}, \tag{8.14.4}$$

where B_δ^X and B_δ^Y denote the δ-ball around 0 in X and Y, respectively. We also assume that

$$\|(\partial_1\Phi(0,0))^{-1}\|\le C_1,\qquad \sup_{(x,y)\in B_\delta^X\times B_\delta^Y}\|\partial_2\Phi(x,y)\|\le C_2$$

for some constants C_1, C_2, where ∂_2 denotes the derivative of Φ with respect to the second variable. Then there is a constant $\varepsilon>0$, depending only on δ, C_1 and C_2, and a unique function $f\colon B_\varepsilon^Y\to B_\delta^X$ such that $\Phi(f(y),y)=0$ for all $y\in B_\varepsilon^Y$. Moreover, f is continuously Fréchet-differentiable and if $\Phi(x,y)=0$ for some $(x,y)\in B_\delta^X\times B_\varepsilon^Y$ then $x=f(y)$. If Φ is analytic then f will be analytic.

Proof. The proof is elementary and left to the reader. ☐

We will apply the implicit function theorem, Lemma 8.14.4, to the function $\Phi_x(y,\omega)$ which we introduce now. For $x, y \in \mathcal{A}$ and $\omega \in \mathbb{C}$, we define

$$\Phi_x(y,\omega) := (\mathrm{Id} - C_x S)[y] - \omega x^2 - \frac{\omega}{2}\Big(xy + yx\Big) - \frac{1}{2}\Big(xS[y]y + yS[y]x\Big). \tag{8.14.5}$$

We remark that $\Phi_{m(z)}(m(z+\omega) - m(z), \omega) = 0$ for all $z \in \mathbb{H}$ and $z + \omega \in \mathbb{H}$ (see (8.6.9)). For the function $\Phi_x(y,\omega)$, we have the following consequence of the implicit function theorem, Lemma 8.14.4.

Lemma 8.14.5. *For some $x \in \mathcal{A}$, we set $\Phi(y,\omega) := \Phi_x(y,\omega)$ for all $y \in \mathcal{A}$ and $\omega \in \mathbb{C}$. If there is $\kappa > 0$ such that*

$$\|x\| \le \kappa, \qquad \|S\| \le \kappa, \qquad \|(\mathrm{Id} - C_x S)^{-1}\| \le \kappa \tag{8.14.6}$$

then there are $\delta > 0$ and $\varepsilon > 0$, depending only on κ, and an analytic function $f\colon D_\varepsilon(0) \to B_\delta^{\mathcal{A}}$ such that

$$\Phi(f(\omega),\omega) = 0$$

for all $\omega \in D_\varepsilon(0)$, where $D_\varepsilon(0) := \{\zeta \in \mathbb{C}\colon |\zeta| < \varepsilon\}$ and $B_\delta^{\mathcal{A}} := \{\tilde{y} \in \mathcal{A}\colon \|\tilde{y}\| \le \delta\}$. Moreover, f is unique in the following strong sense: if $y \in B_\delta^{\mathcal{A}}$ satisfies $\Phi(y,\omega) = 0$ for some $\omega \in D_\varepsilon(0)$ then we have $y = f(\omega)$.

Proof. In order to prove Lemma 8.14.5, we apply Lemma 8.14.4, whose assumptions we check first. For the directional derivative $(\partial_1\Phi(y,\omega))[h]$ at (y,ω) with respect to the first variable in the direction $h \in \mathcal{A}$, we obtain

$$(\partial_1\Phi(y,\omega))[h] = (\mathrm{Id} - C_x S)[h] - \frac{\omega}{2}(xh + hx) - \frac{1}{2}(x(S[h]y + S[y]h) + (yS[h] + hS[y])x).$$

Hence, $\partial_1\Phi(0,0) = \mathrm{Id} - C_x S$ and, owing to the third assumption in (8.14.6), we can choose $C_1 = \kappa$ in Lemma 8.14.4. Moreover, we also conclude

$$\begin{aligned}(\mathrm{Id} - (\partial_1\Phi(0,0))^{-1}\partial_1\Phi(y,\omega))[h] = \frac{1}{2}(\mathrm{Id} - C_x S)^{-1}\Big[&\omega(xh + hx) + x(S[y]h + hS[y]) \\ &+ (hS[y] + yS[h])x\Big].\end{aligned}$$

We now determine how to choose δ such that (8.14.4) is satisfied. We estimate the previous expression under the assumption that $\|y\| \le \delta$ and $|\omega| \le \delta$ for some $\delta > 0$.

Under this assumption, we obtain

$$\|(\mathrm{Id} - (\partial_1\Phi(0,0))^{-1}\partial_1\Phi(y,\omega))[h]\| \le \|(\mathrm{Id} - C_x S)^{-1}\|(\delta\|x\| + 2\|x\|\|S\|\delta)\|h\|.$$

Hence, (8.14.4) is satisfied if

$$\delta < \frac{1}{2\|(\mathrm{Id} - C_x S)^{-1}\|\|x\|(1+2\|S\|)}.$$

Therefore, we can choose $\delta := (2\kappa^2(1+2\kappa))^{-1}$ in order to meet the condition (8.14.4).

From the definition of Φ in (8.14.5), we obtain that the directional derivative $\partial_2\Phi(y,\omega)$ at (y,ω) with respect to the second variable is given by

$$(\partial_2\Phi(y,\omega))[\sigma] = (-x^2 - \frac{1}{2}(xy+yx))\sigma$$

for $\sigma \in \mathbb{C}$. Hence, with the choice of δ above, we can choose $C_2 = \kappa^2 + \kappa\delta$ in Lemma 8.14.4. Therefore, δ, C_1 and C_2 in Lemma 8.14.4 can be chosen to depend only on κ due to the assumption (8.14.6). Thus, since Φ is analytic due to its definition in (8.14.5), Lemma 8.14.5 follows from the implicit function theorem, Lemma 8.14.4. □

Proof of Lemma 8.14.1. Lemma 8.12.2 (i) yields $\|S\|_{2\to\|\cdot\|} \lesssim 1$ due to $S[x] \le k_1\langle x\rangle\mathbb{1}$ for all $x \in \overline{\mathcal{A}}_+$. Therefore, $\|a\| \lesssim 1$ and $\|S\| \le \|S\|_{2\to\|\cdot\|} \lesssim 1$ imply that $\operatorname{supp} v = \operatorname{supp}\rho$ is bounded, i.e., $\sup\{|\tau| \colon \tau \in \operatorname{supp}\rho\} \lesssim 1$ by (8.2.5a).

First, we assume that (i) holds true. We set $N := \{\eta \in (0,1] \colon \eta\|\operatorname{Im} m(\tau + \mathrm{i}\eta)\|^{-1} \ge c/2\}$. By assumption, N is nonempty and has 0 as an accumulation point. In particular, we have

$$\|\operatorname{Im} m(z)\| \le \frac{2\eta}{c}, \qquad \eta\mathbb{1} \lesssim \operatorname{Im} m(z) \lesssim \frac{\eta}{c}\mathbb{1} \tag{8.14.7}$$

for all $z \in \tau + \mathrm{i}N$. The first bound is a direct consequence of the definition of N. The second bound follows from (8.2.4) and the bounded support of v. Moreover, the first bound immediately implies the third bound. By averaging the two last bounds in (8.14.7) and using $\operatorname{Im} m(\tau+\mathrm{i}\eta) \lesssim \eta$ for $\eta \in N$, we obtain the third and fourth estimates in (8.14.1). In particular, $\rho(z) \sim \|\operatorname{Im} m(z)\|$ for $z \in \tau + \mathrm{i}N$. Owing to (8.2.4), for any

$z \in \mathbb{H}$ and $x, y \in L^2$, we have

$$|\langle x\,, m(z)y\rangle| \leq \frac{1}{2}\int_{\mathbb{R}} \frac{\langle x\,, v(\mathrm{d}\tau)x\rangle + \langle y\,, v(\mathrm{d}\tau)y\rangle}{|\tau - z|} \lesssim \frac{1}{\eta}\Big(\langle x\,, \operatorname{Im} m(z)x\rangle + \langle y\,, \operatorname{Im} m(z)y\rangle\Big)$$
$$\leq \frac{2}{c}\Big(\|x\|_2^2 + \|y\|_2^2\Big).$$

Here, we used that v has a bounded support and (8.2.4) in the second step and the first bound in (8.14.7) in the last step. This proves the first bound in (8.14.1). The second estimate in (8.14.1) is a consequence of (8.2.3) as well as $\|a\| \lesssim 1$, $\|S\| \leq \|S\|_{2\to\|\cdot\|} \lesssim 1$ and the first bound in (8.14.1). We recall the definitions of $q = q(z)$ and $u = u(z)$ in (8.3.1). Owing to Lemma 8.4.8 (i), the bounds in (8.14.1) yield

$$\|q\| \lesssim 1, \qquad \|q^{-1}\| \lesssim 1, \qquad \operatorname{Im} u \sim \langle \operatorname{Im} u\rangle \mathbb{1} \sim \rho\mathbb{1} \tag{8.14.8}$$

uniformly for all $z \in \tau + \mathrm{i}N$. Thus, for all $x \in \overline{\mathcal{A}}_+$ and $z = \tau + \mathrm{i}\eta$ and $\eta \in N$, $F = F(z)$ satisfies $F[x] \lesssim \langle x\rangle\mathbb{1}$ due to $S[x] \lesssim \langle x\rangle\mathbb{1}$. Hence, Lemma 8.12.1 (i) yields the existence of an eigenvector $f \in \overline{\mathcal{A}}_+$, i.e., $Ff = \|F\|_2 f$. By taking the imaginary part of (8.3.3) and then the scalar product with f as well as using the symmetry of F, we get

$$1 - \|F\|_2 = \eta\frac{\langle f\,, qq^*\rangle}{\langle f\,, \operatorname{Im} u\rangle} \sim \eta\|\operatorname{Im} m(z)\|^{-1} \gtrsim c \tag{8.14.9}$$

for $z = \tau + \mathrm{i}\eta$ and $\eta \in N$ (compare (8.4.5)). Here, we also used $f \in \overline{\mathcal{A}}_+$, (8.14.8), $\rho(z) \sim \|\operatorname{Im} m(z)\|$ and the definition of N. This completes the proof of (i) $\Rightarrow$ (ii).

Next, let (ii) be satisfied. As before, Lemma 8.4.8 (i) implies (8.14.8) for all $z \in \tau + \mathrm{i}N$ due to the first four bounds in (8.14.1). Thus, inspecting the proofs of Lemma 8.4.8 (iii) and Proposition 8.4.1 and using $\|S\|_{2\to\|\cdot\|} \lesssim 1$ via Lemma 8.12.2 (ii) yield

$$\|(\mathrm{Id} - C_{m(z)}S)^{-1}\| \lesssim 1 \tag{8.14.10}$$

uniformly for all $z \in \tau + \mathrm{i}N$. Thus, we can apply Lemma 8.14.5, with $x = m(\tau + \mathrm{i}\eta)$ for each $\eta \in N$. For Φ as defined in (8.14.5), we set $\Psi_\eta(\Delta, \omega) := \Phi_{m(\tau+\mathrm{i}\eta)}(\Delta, \omega)$ for $\eta \in N$, $\Delta \in \mathcal{A}$ and $\omega \in \mathbb{C}$. Thus, by Lemma 8.14.5, there are $\delta > 0$, $\varepsilon > 0$ and unique analytic functions $\Delta_\eta \colon D_\varepsilon(0) \to B_\delta^{\mathcal{A}}$ such that $\Psi_\eta(\Delta_\eta(\omega), \omega) = 0$ for all $\omega \in D_\varepsilon(0)$ and all $\eta \in N$. We now explain why ε can be chosen uniformly for all $\eta \in N$. By (8.14.1) and (8.14.10), there are bounds on $m(z)$ and $(\mathrm{Id} - C_{m(z)}S)^{-1}$ which hold uniformly for $z \in \tau + \mathrm{i}N$.

Hence, there is a $\kappa > 0$, independent of η, such that (8.14.6) holds true uniformly for all $\eta \in N$. These uniform bounds yield the uniformity of ε. Since 0 is an accumulation point of N, there is $\eta_0 \in N$ such that $\eta_0 < \varepsilon$. We set $z := \tau + \mathrm{i}\eta_0$. An easy computation using (8.2.3) at spectral parameters z and $z + \omega$ shows $\Psi_{\eta_0}(m(\omega + z) - m(z), \omega) = 0$ for all $\omega \in \mathbb{C}$ such that $\omega + z \in \mathbb{H}$. Owing to the continuity of m, we find $\varepsilon' \in (0, \varepsilon)$ such that $m(\omega + z) - m(z) \in B_\delta^{\mathcal{A}}$ for all $\omega \in D_{\varepsilon'}(0)$. Thus, by the uniqueness of Δ_{η_0} (cf. Lemma 8.14.5), $\Delta_{\eta_0}(\omega) = m(\omega + z) - m(z)$ for all $\omega \in D_{\varepsilon'}(0)$. As Δ_{η_0} and $m(\cdot + z)$ are analytic, owing to the identity theorem, we obtain $\Delta_{\eta_0}(\omega) + m(z) = m(\omega + z)$ for all $\omega \in D_\varepsilon(0)$ satisfying $\omega + z \in \mathbb{H}$. Using $\eta_0 < \varepsilon$, we set $m := \Delta_{\eta_0}(-\mathrm{i}\eta_0) + m(z)$. For this choice of m, the continuity of $\Delta_{\eta_0}(\omega)$ for $\omega \to -\mathrm{i}\eta_0$ and $\Delta_{\eta_0}(\omega) + m(z) = m(\omega + z)$ yield (8.14.2). It remains to show that m is self-adjoint. Since (8.14.8) holds true under (ii) as we have shown above, we obtain

$$\eta \|\mathrm{Im}\, m(z)\|^{-1} \sim 1 - \|F\|_2 \geq C^{-1}$$

for $z = \tau + \mathrm{i}\eta$ and $\eta \in N$ as in (8.14.9). Thus, $\liminf_{\eta \downarrow 0} \|\mathrm{Im}\, m(\tau + \mathrm{i}\eta)\| \leq 0$. Hence, we obtain $\mathrm{Im}\, m = 0$, i.e., $m = m^*$. This completes the proof of (ii) $\Rightarrow$ (iii).

If (iii) holds true then $\mathrm{Id} - C_m S$ has a bounded linear inverse on $\mathcal{A}$ for m. Hence, we can apply Lemma 8.14.5 with $x = m$. Therefore, there are $\delta > 0$, $\varepsilon > 0$ and an analytic function $\Delta \colon D_\varepsilon(0) \to B_\delta^{\mathcal{A}}$ such that $\Phi_m(\Delta(\omega), \omega) = 0$ for all $\omega \in D_\varepsilon(0)$. In particular, $f \colon D_\varepsilon(\tau) \to \mathcal{A}$, $f(w) := \Delta(w - \tau) + m$ is analytic. From (8.14.2) and (8.2.3), we see that m is invertible and satisfies (8.2.3) at $z = \tau$. Thus, a straightforward computation using (8.2.3) at $z = \tau$ and at $z = \tau + \mathrm{i}\eta$ yields $\Phi_m(m(\tau + \mathrm{i}\eta) - m, \mathrm{i}\eta) = 0$ for all $\eta \in (0, \varepsilon]$. Therefore, $m(\tau + \mathrm{i}\eta) = \Delta(\mathrm{i}\eta) + m = f(\tau + \mathrm{i}\eta)$ for all $\eta \in (0, \eta_*]$ and some $\eta_* \in (0, \varepsilon]$ due to the uniqueness part of Lemma 8.14.5 and (8.14.2). Since m and f are analytic on $D_\varepsilon(\tau) \cap \mathbb{H}$, the identity theorem implies $m(z) = f(z)$ for all $z \in D_\varepsilon(\tau) \cap \mathbb{H}$. A simple computation shows $\Phi_m(\Delta(\bar{\omega})^*, \omega) = \Phi_m(\Delta(\bar{\omega}), \bar{\omega})^* = 0$ for all $\omega \in D_\varepsilon(0)$ as $m = m^*$. Hence, $\Delta(\omega) = \Delta(\bar{\omega})^*$ for all $\omega \in D_\varepsilon(0)$ by the uniqueness part of Lemma 8.14.5. Thus, $f(w) = f(\bar{w})^*$ for all $w \in D_\varepsilon(\tau)$ and $f(w) = f(w)^*$ for all $w \in D_\varepsilon(\tau) \cap \mathbb{R}$. This proves (iii) $\Rightarrow$ (iv). Clearly, (iv) implies (v) by (8.2.4).

If the statement in (v) holds true then $\operatorname{dist}(\tau, \operatorname{supp}\rho) \geq \varepsilon$. In particular, by (8.3.7), we have

$$\liminf_{\eta\downarrow 0} \eta \|\operatorname{Im} m(\tau + \mathrm{i}\eta)\|^{-1} \geq \liminf_{\eta\downarrow 0} \operatorname{dist}(\tau + \mathrm{i}\eta, \operatorname{supp}\rho)^2 \geq \varepsilon^2$$

for all $\eta > 0$. Here, we used (8.3.7) in the first step. This immediately implies (vi) with $c = \varepsilon^2$. Moreover, (i) is immediate from (vi).

Inspecting the proofs of the implications above shows the additional statement about the effective dependence of the constants in (i) – (vi). In particular, the application of Lemma 8.14.5, in the proof of (iv) shows that ε can be chosen to depend only on k_1 and C from (iii). This completes the proof of Lemma 8.14.1. □

CHAPTER 9

Correlated Random Matrices: Band Rigidity and Edge Universality

The present chapter contains the preprint [17] which was written jointly with László Erdős, Torben Krüger and Dominik Schröder. We prove edge universality for a general class of correlated real symmetric or complex Hermitian Wigner matrices with arbitrary expectation. Our theorem also applies to internal edges of the self-consistent density of states. In particular, we establish a strong form of band rigidity which excludes mismatches between location and label of eigenvalues close to internal edges in these general models.

9.1. Introduction

Spectral statistics of large random matrices exhibit a remarkably robust universality pattern; the local distribution of eigenvalues is independent of details of the matrix ensemble up to symmetry type. In the bulk of the spectrum this was first observed by Wigner and formalized by Dyson and Mehta [114] who also computed the correlation functions of the Gaussian ensembles in the 1960's. At the spectral edges the correct statistics was identified by Tracy and Widom both in the GUE and GOE ensembles [148, 149] in the mid 1990's. Subsequently, a main line of research became to extend universality to more and more general classes of ensembles with the goal of eventually approaching the grand vision that predicts GUE/GOE statistics for any sufficiently complex disordered quantum system in the delocalized phase.

Beyond Gaussian ensembles, the first actual proofs of universality for Wigner matrices took different paths in the bulk and at the edge. While in the bulk only limited progress was made until a decade ago, the first fairly general edge universality proof by Soshnikov [136] appeared shortly after the calculations of Tracy and Widom. The main reason is that edge statistics is still accessible via an ingenious but laborious extension of the classical moment method. In contrast, the bulk universality required fundamentally new tools based on resolvents and the analysis of the Dyson Brownian motion developed in a series of work [58, 59, 62, 64, 68, 71]. This method, called the *three-step strategy*, is summarized in [67]. In certain cases parallel results [144, 145] were obtained via the *four moment comparison theorem*.

Despite its initial success [74, 136], the moment method seems limited when it comes to generalizations beyond Wigner matrices with i.i.d. entries; the bookkeeping of the combinatorics is extremely complicated even in the simplest case. The resolvent approach is much more flexible. Its primary goal is to establish *local laws*, i.e., proving that the local eigenvalue density on scales slightly above the eigenvalue spacing becomes deterministic as the dimension of the matrix tends to infinity. Refined versions of the local law even identify resolvent matrix elements with a spectral parameter very close to the real axis. In contrast to the bulk, at the spectral edge this information can be boosted to detect individual eigenvalue statistics by comparison with the Gaussian ensemble. These ideas have led to the proof of the Tracy-Widom edge universality for Wigner matrices with high moment conditions [71], see also [145] with vanishing third moment. Finally, a necessary and sufficient condition on the entry distributions was found in [109] following earlier work in [125] and an almost optimal necessary condition in [21]. Direct resolvent comparison methods have been used to prove Tracy-Widom universality for *deformed Wigner matrices*, i.e., matrices with a deterministic diagonal expectation, [106], even in a certain sparse regime [107]. The extension of this approach to sample covariance matrices with a diagonal population covariance matrix at extreme edges [108] has resolved a long standing conjecture in the statistics literature. Tracy-Widom universality for general population covariance matrices, including internal edges, was established in [101].

The next level of generality is to depart from the i.i.d. case. While the resolvent method for proving local laws can handle *generalized Wigner ensemble*, i.e., matrices $H = (h_{ab})$ with merely stochastic variance profile $\sum_b \mathbf{Var}\, h_{ab} = 1$, the direct comparison becomes problematic if higher moments vary since they cannot be simultaneously matched with a GUE/GOE ensemble. The problem was resolved in [43] with a general approach that also covered invariant β-ensembles. While Dyson Brownian motion did not play a direct role in [43], the proof used the addition of a small Gaussian component and the concept of local ergodicity of the Gibbs state; ideas developed originally in [64, 65] in the context of bulk universality.

A fully dynamical approach to edge universality, following an earlier development in the bulk based on the *three-step strategy*, has recently been given in [103]. In general, the first step within any three-step strategy is the local law providing a priori bounds. The second step is the fast relaxation to equilibrium of the Dyson Brownian motion that proves universality for Gaussian divisible ensembles. The third step is a perturbative comparison argument to remove the small Gaussian component. Recent advances in the bulk have crystallized that the only model dependent step in this strategy is the first one. The other two steps have been formulated as very general "black-box" tools whose only input is the local law see [66, 103, 104, 105]. Using the three-step approach and [103], edge universality for sparse matrices was proven in [97] and for correlated Gaussian matrices with a specific two-scale correlation structure in [1]. All these edge universality results only cover the extremal edges of the spectrum, while the self-consistent (deterministic) density of states may be supported on several intervals. Multiple interval support becomes ubiquitous for *Wigner-type* matrices [7], i.e., matrices with independent entries and general expectation and variance profile. The square root singularity in the density, even at the internal edges, is a universal phenomenon for a very large class of random matrices since it is inherent to the underlying *Dyson equation*. This was demonstrated for Wigner-type matrices in [4] and more recently for correlated random matrices with a general correlation structure in Chapter 8.

In the current paper we show that the eigenvalue statistics at the spectral edges of the self-consistent density follow the Tracy-Widom distribution, assuming only a mild decay of correlation between entries, but otherwise no special structure. We can handle any internal edge as well. In the literature internal edge universality for matrices of Wigner-type has first been established for deformed GUE ensembles [129] which critically relied on contour integral methods, only available for Gaussian models in the Hermitian symmetry class. We remark that a similar method handled extreme eigenvalues of deformed GUE [48, 98]. A more general approach for internal edges has been given in [101] that could handle any deformed Wigner matrices with general expectation, as long as the variance profile is constant, by comparing it with the corresponding Gaussian model. Our method requires neither constant variance nor independence of the matrix elements.

In order to prove our general form of edge universality at all internal edges we used three key inputs in addition to [103]. First, we rely on a detailed shape analysis of the self-consistent density of states ϱ from Chapter 8. Secondly, we prove a strong version of the local law that excludes eigenvalues in the internal gaps. Thirdly, we establish a topological rigidity phenomenon for the *bands*, the connected components that constitute the support of ϱ. This *band rigidity* asserts that the number of eigenvalues within each band exactly matches the mass that ϱ predicts for that band. The topological nature of band rigidity guarantees that this mass remains constant along the deformations of the expectation and correlation structure of the entries as long as the gaps between the bands remain open. A similar rigidity (also called "exact separation of eigenvalues") has first been established for sample covariance matrices in [23] and it also played a key role in Tracy-Widom universality proof at internal edges in [101]. Note that band rigidity is a much stronger concept than the customary rigidity in random matrix theory [71] that allows for an uncertainty in the location of N^{ϵ} eigenvalues. In other words, there is no mismatch whatsoever between location and label of the eigenvalues near the internal edges along the matrix Dyson Brownian motion, the label of the eigenvalue uniquely determines to which spectral band it belongs.

Our result also highlights a key difference between Wigner-type matrix models and invariant β-ensembles. For self-consistent densities with multiple support intervals (the so-called *multi-cut* regime), the number of particles (eigenvalues) close to some support interval fluctuates for invariant ensembles with general potentials [41]. As a consequence internal edge universality results (see e.g. [30, 118]) require a stochastic relabelling of eigenvalues.

Our setup is a general $N \times N$ random matrix $H = H^*$ with a slowly decaying correlation structure and arbitrary expectation, under the very same conditions as the recent bulk universality result from [56]. Regarding strategy of proving the local law, the starting point is to find the deterministic approximation of the resolvent $G(z) = (H - z)^{-1}$

with a complex spectral parameter z in the upper half plane. This approximation is given as the solution $M = M(z)$ to the *Matrix Dyson Equation (MDE)*

$$1 + (z - A + \mathcal{S}[M])M = 0,$$

where the expectation matrix $A := \mathbb{E}H$ and the linear map $\mathcal{S}[R] := \mathbb{E}(H - A)R(H - A)$ on the space of matrices R encode the first two moments of the random matrix. The resolvent $G(z)$ approximately satisfies the MDE with an additive perturbation term which was already shown to be sufficiently small in [56]. This fact, combined with a careful stability and shape analysis of the MDE in Chapter 8 imply that G is indeed close to M. In order to prove edge universality we use a correlated Ornstein-Uhlenbeck process H_t which adds a small Gaussian component of size t to the original matrix model, while preserving expectation and covariance. We prove that the resolvent satisfies the optimal local law uniformly along the flow and appeal to the recent result from [103] to prove edge universality for H_t whenever $t \gg N^{-1/3}$. In the final step we perform a Green function comparison together with our band rigidity to show that the eigenvalue correlation functions of H_t matches those of H as long as $t \ll N^{-1/6}$ which yields the desired edge universality.

After presenting our main results in Section 9.2, we then prove the optimal local law at regular edges (and in the spectral bulk), as well as eigenvector delocalization and both types of rigidity in Section 9.3. Section 9.4 is devoted to the proof of edge universality.

Notations. We now introduce some custom notations we use throughout the paper. For non-negative functions $f(A, B), g(A, B)$ we use the notation $f \leq_A g$ if there exist constants $C(A)$ such that $f(A, B) \leq C(A)g(A, B)$ for all A, B. Similarly, we write $f \sim_A g$ if $f \leq_A g$ and $g \leq_A f$. We do not indicate the dependence of constants on basic parameters that will be called model parameters later. If the implied constants are universal, we instead write $f \lesssim g$ and $f \sim g$. We denote vectors by bold-faced lower case Roman letters $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$, and matrices by upper case Roman letters $A, B \in \mathbb{C}^{N \times N}$. The standard scalar product and Euclidean norm on $\mathbb{C}^N$ will be denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ and $\|\boldsymbol{x}\|$, while we also write $\langle A, B \rangle := N^{-1} \operatorname{Tr} A^* B$ for the scalar product of matrices, and $\langle A \rangle := N^{-1} \operatorname{Tr} A$. The usual operator norm induced by the vector norm $\|\cdot\|$ will be denoted by $\|A\|$, while the Hilbert-Schmidt (or Frobenius) norm will be denoted by $\|A\|_{\mathrm{hs}} := \sqrt{\langle A, A\rangle}$. For random variables $X, Y, \ldots$ we denote the joint cumulant by $\kappa(X, Y, \ldots)$. For integers n we define $[n] := \{1, \ldots, n\}$.

9.2. Main results

We consider correlated real symmetric and complex Hermitian random matrices of the form

$$H = A + W, \qquad \mathbb{E}W = 0$$

with deterministic $A \in \mathbb{C}^{N\times N}$ and sufficiently fast decaying correlations among the matrix elements of W. The matrix entries $w_{ab} = w_\alpha$ are often labelled by double indices $\alpha = (a, b) \in [N]^2$. The randomness W is scaled in such a way that $\sqrt{N} w_\alpha$ are random variables of order one[1]. This requirement ensures that the spectrum of H is kept of order 1, as N tends to infinity. Our first aim is to prove that, in the *bulk* and around the *regular edges* of the spectrum, the resolvent $G = G(z) = (H - z)^{-1}$ is well approximated by the solution $M = M(z)$ to the *Matrix Dyson equation (MDE)*

$$1 + (z - A + \mathcal{S}[M])M = 0, \quad \operatorname{Im} M := \frac{M - M^*}{2\mathrm{i}} > 0, \quad \mathcal{S}[R] := \mathbb{E}WRW, \tag{9.2.1}$$

with $z \in \mathbb{H} := \{\, z \in \mathbb{C} \mid \operatorname{Im} z > 0 \,\}$. We suppress the dependence of G and M, and similarly of many other quantities, on the spectral parameter z in our notation. Estimates on z-dependent quantities are always meant uniformly for z in some specified domain. From the solution M we define the *self-consistent density of states*

$$\varrho(E) := \lim_{\eta\searrow 0} \frac{\operatorname{Im}\, \langle M(E + \mathrm{i}\eta)\rangle}{\pi}, \qquad E \in \mathbb{R},$$

which approximates the density of states of H increasingly well as N tends to infinity. The support of ϱ is known to consist of several compact intervals with square root behaviour at the edges. An edge is called *regular* if it is well separated from other edges. The *spectral bulk* refers to points E where $\varrho(E) \geq \zeta$ with some fixed threshold $\zeta > 0$.

We now list our main assumptions, which are identical to those from [56]. All explicit and implicit constants in Assumptions (A)–(F) are called *model parameters*.

[1]In some previous works, as in [56], the convention $H = A + W/\sqrt{N}$ with order one w_α was used.

Assumption (A) (Bounded expectation)**.** There exists some constant C such that $\|A\| \le C$ for all N.

Assumption (B) (Finite moments)**.** For all $q \in \mathbb{N}$ there exists a constant μ_q such that $\mathbb{E}|\sqrt{N} w_\alpha|^q \le \mu_q$ for all α.

Assumption (CD) (Polynomially decaying metric correlation structure)**.** For the $k = 2$ point correlation we assume a decay of the type

$$\Big|\kappa\Big(f_1(\sqrt{N}W), f_2(\sqrt{N}W)\Big)\Big| \lesssim \frac{\sqrt{\mathbb{E}|f_1(\sqrt{N}W)|^2}\sqrt{\mathbb{E}|f_2(\sqrt{N}W)|^2}}{1 + d(\operatorname{supp} f_1, \operatorname{supp} f_2)^s}, \tag{9.2.2a}$$

for some $s > 12$ and all square integrable functions f_1, f_2. For $k \ge 3$ we assume a decay condition of the form

$$\Big|\kappa\Big(f_1(\sqrt{N}W), \dots, f_k(\sqrt{N}W)\Big)\Big| \lesssim \prod_{e \in E(T_{\min})} |\kappa(e)|\,, \tag{9.2.2b}$$

where $T_{\min}$ is the minimal spanning tree in the complete graph on the vertices $1, \dots, k$ with respect to the edge length $\operatorname{dist}(\{i,j\}) = d(\operatorname{supp} f_i, \operatorname{supp} f_j)$, i.e., the tree for which the sum of the lengths $\operatorname{dist}(e)$ is minimal, and $\kappa(\{i,j\}) = \kappa(f_i, f_j)$. Here d is the standard Euclidean metric on the index space $[N]^2$ and $\operatorname{supp} f \subset [N]^2$ denotes the set indexing all entries in $\sqrt{N}W$ that f genuinely depends on.

Remark 9.2.1. All results in this paper and their proofs hold verbatim if Assumption (CD) is replaced by the more general assumptions (C), (D) from [56]. In particular, the metric structure imposed on the index space $[N]^2$ is not essential. For details the reader is referred to [56, Section 2.1].

Assumption (E) (Flatness)**.** There exist constants $0 < c < C$ such that $c\,\langle T\rangle \le \mathcal{S}[T] \le C\,\langle T\rangle$ for any positive semi-definite matrix T.

Assumption (F) (Fullness)**.** There exists a constant $\lambda > 0$ such that $N\mathbb{E}\,|\operatorname{Tr} BW|^2 \ge \lambda \operatorname{Tr} B^2$ for any deterministic matrix B of the same symmetry class (either real symmetric or complex Hermitian) as H.

Our main technical result is an optimal local law in the spectral bulk and at regular edges. According to the shape analysis from Chapter 8 it follows that ϱ can also feature almost-cusp points which we have to exclude from our spectral domain. For $E \in \mathbb{R} \setminus \operatorname{supp} \varrho$ we define $\Delta(E) = \Delta^{\varrho}(E)$ as the length of the largest interval around E in $\mathbb{R} \setminus \operatorname{supp} \varrho$. Accordingly, we define the set of *almost-cusp points* $P_{\text{cusp}} = P_{\text{cusp}}^{\zeta}$ for small ζ as

$$\begin{aligned} P_{\text{cusp}} := \{ \, E \in \operatorname{supp} \varrho \setminus \partial \operatorname{supp} \varrho \mid E \text{ is a local minimum of } \varrho, \ \varrho(E) \leq \zeta \, \} \\ \cup \{ \, E \in \mathbb{R} \setminus \operatorname{supp} \varrho \mid \Delta(E) \leq \zeta \, \} \, , \end{aligned}$$

and $d_{\text{cusp}}(z) = d_{\text{cusp}}^{\zeta}(z) := \operatorname{dist}(z, P_{\text{cusp}})$ denotes the distance from the almost-cusps. We will always work with spectral parameters z such that the solution M to (9.2.1) remains bounded in a neighbourhood of z. To define this condition precisely, we fix a large constant M_* and define the set $P_{\text{M}} = P_{\text{M}}^{M_*}$ as

$$P_{\text{M}}^{M_*} := \left\{ \tau \in \mathbb{R} \;\middle|\; \sup_{\eta > 0} \| M(\tau + \mathrm{i}\eta) \| > M_* \right\} ,$$

and the distance $d_{\text{M}}(z) = d_{\text{M}}^{M_*}(z) := \operatorname{dist}(z, P_{\text{M}})$ from this set. For $\zeta, \delta, M_* > 0$ we then define the spectral domain $\mathbb{D} = \mathbb{D}^{\zeta, \delta, M_*}$ away from almost-cusp and large $\|M\|$ points by

$$\mathbb{D} := \left\{ \, z \in \mathbb{C} \;\middle|\; d_{\text{cusp}}(z) \geq \delta, \ d_{\text{M}}(z) \geq \delta, \ |z| \leq N^{C_0} \, \right\}$$

for some arbitrary fixed C_0. We remark that the boundedness of $\|M\|$ in a small interval around the spectral parameter is automatically satisfied in the spectral bulk. At regular edges, however, the boundedness cannot be guaranteed under our general assumptions but has to be checked for each concrete model (see Section 8.9 in Chapter 8 for a large class of models for which $\|M\|$ is guaranteed to be bounded). Our goal is to establish an optimal local law for those spectral parameters $z = E + \mathrm{i}\eta$ whose imaginary part $\eta = \operatorname{Im} z$ is slightly larger than $1/N$, i.e., in the spectral domain

$$\mathbb{D}_{\gamma} := \mathbb{D} \cap \left\{ \, z \in \mathbb{C} \;\middle|\; \operatorname{Im} z \geq N^{-1+\gamma} \, \right\}$$

for some $\gamma > 0$.

Theorem 9.2.2 (Bulk and edge local law)**.** *Under Assumptions (A)–(E) and for any* $D, M_*, \gamma, \epsilon, \delta, \zeta > 0$*, there exists some* $C < \infty$ *depending only on these and the model* *parameters such that we have the isotropic local law,*

$$\mathbb{P}\left(|\langle \boldsymbol{x}, (G-M)\boldsymbol{y}\rangle| \le N^{\epsilon}\, \|\boldsymbol{x}\|\, \|\boldsymbol{y}\| \left(\sqrt{\frac{\varrho}{N \mathrm{Im}\, z}} + \frac{1}{N\mathrm{Im}\, z}\right) \quad in \quad \mathbb{D}_{\gamma}\right) \ge 1 - CN^{-D} \tag{9.2.3a}$$

for all deterministic vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$ *and the averaged local law,*

$$\mathbb{P}\left(|\langle B(G-M)\rangle| \le N^{\epsilon}\, \|B\|\, \frac{1}{N\mathrm{Im}\, z} \quad in \quad \mathbb{D}_{\gamma}\right) \ge 1 - CN^{-D} \tag{9.2.3b}$$

for all deterministic matrices $B \in \mathbb{C}^{N\times N}$*. Moreover, outside the spectrum at a distance*[2] *of* $\kappa(z) := \mathrm{dist}(\mathrm{Re}\, z, \partial\, \mathrm{supp}\, \varrho)$ *we have the improved averaged local law for any* $\omega > 0$

$$\mathbb{P}\left(|\langle B(G-M)\rangle| \le \frac{N^{\epsilon}\, \|B\|}{N(\kappa + \mathrm{Im}\, z)(1+|z|)} \ \ in\ \mathbb{D}_{\mathrm{out}}\right) \ge 1 - CN^{-D}, \tag{9.2.3c}$$

with C *also depending on* ω*, where we introduced*

$$\mathbb{D}_{\mathrm{out}} := \left\{\, z \in \mathbb{D} \,\middle|\, \mathrm{dist}(\mathrm{Re}\, z, \mathrm{supp}\, \varrho) \ge N^{-2/3+\omega} \,\right\}.$$

We remark that in the spectral bulk Theorem 9.2.2 is identical to the local law in [56]. The novelty of the present paper is the optimal local law and its corollaries at the regular edges.

Corollary 9.2.3 (No eigenvalues outside the support of the self-consistent density)**.** *Under the assumptions of Theorem 9.2.2 we have for any* $\omega, \zeta, \delta, D, M_* > 0$

$$\mathbb{P}\left(\exists \lambda \in \mathrm{Spec}\, H,\ \mathrm{dist}(\lambda, \mathrm{supp}\, \varrho) \ge N^{-2/3+\omega},\ d_{\mathrm{cusp}}(\lambda) \ge \delta,\ d_{\mathrm{M}}(\lambda) \ge \delta\right) \le_{\omega,\zeta,\delta,D,M_*} N^{-D}.$$

Corollary 9.2.4 (Delocalization)**.** *Under the assumptions of Theorem 9.2.2 it holds for an* ℓ^2*-normalized eigenvector* $\boldsymbol{u}$ *corresponding to a non-cusp eigenvalue* λ *of* H *that*

$$\sup_{\|\boldsymbol{x}\|=1} \mathbb{P}\left(|\langle \boldsymbol{x}, \boldsymbol{u}\rangle| \ge \frac{N^{\epsilon}}{\sqrt{N}},\ H\boldsymbol{u} = \lambda \boldsymbol{u},\ \|\boldsymbol{u}\| = 1,\ d_{\mathrm{cusp}}(\lambda) \ge \delta,\ d_{\mathrm{M}}(\lambda) \ge \delta\right) \le_{\epsilon,\zeta,\delta,D} N^{-D}$$

for any $\epsilon, \zeta, \delta, D > 0$*.*

Corollary 9.2.5 (Band rigidity and eigenvalue rigidity)**.** *Under the assumptions of Theorem 9.2.2 the following holds. For any* $\epsilon, D > 0$ *there exists some* $C < \infty$ *such that for*

[2] We warn the reader that cumulants and the distance to the boundary of the spectrum are both denoted by κ. Because cumulants are usually written with explicit random variables in the argument, this should not create any confusions.

any $E \in \mathbb{R} \setminus \operatorname{supp} \varrho$ *with* $\operatorname{dist}(E, \operatorname{supp} \varrho) \geq \epsilon$ *the number of eigenvalues less than* E *is with high probability deterministic, i.e., that*

$$\mathbb{P}\Big(|\operatorname{Spec} H \cap (-\infty, E)| = N \int_{-\infty}^{E} \varrho(x) \mathrm{d}x \Big) \geq 1 - CN^{-D}. \tag{9.2.4a}$$

We also have the following strong form of eigenvalue rigidity. Let $\lambda_1 \leq \cdots \leq \lambda_N$ *be the ordered eigenvalues of* H *and denote the classical position of the eigenvalue close to energy* $E \in \operatorname{supp} \varrho$ *by* $k(E) := \lceil N \int_{-\infty}^{E} \varrho(x) \mathrm{d}x \rceil$. *It then holds that*

$$\mathbb{P}\left(\sup_{E} \left| \lambda_{k(E)} - E \right| \geq \min \left\{ \frac{N^{\epsilon}}{N \operatorname{dist}(E, \partial \operatorname{supp} \varrho)^{1/2}}, \frac{N^{\epsilon}}{N^{2/3}} \right\} \right) \leq_{\epsilon, \zeta, \delta, D} N^{-D} \tag{9.2.4b}$$

for any $\epsilon, \zeta, \delta, D > 0$, *where the supremum is taken over all* $E \in \operatorname{supp} \varrho$ *such that* $d_{\mathrm{cusp}}(E) \geq \delta$ *and* $d_{\mathrm{M}}(E) \geq \delta$.

Remark 9.2.6 (Integer mass)**.** Note that (9.2.4a) entails the non trivial fact that for $E \notin \operatorname{supp} \varrho$, $N \int_{-\infty}^{E} \varrho(x) \mathrm{d}x$ is always an integer, see Proposition 8.2.6 in Chapter 8. Moreover, it then trivially implies that $N \int_{[a,b]} \varrho(x) \mathrm{d}x$ is an integer for each *band* $[a, b]$, i.e., connected component of $\operatorname{supp} \varrho$. Finally, (9.2.4a) also shows that the number of eigenvalues in each band is given by this integer with overwhelming probability. This is in sharp contrast to invariant β-ensembles where no such mechanism is present. For example, for an odd number of particles in a symmetric double-well potential, $N \int_{-\infty}^{0} \varrho(x) \mathrm{d}x = N/2$ is a half integer.

The main application of the optimal local law from Theorem 9.2.2 is edge universality, as stated in the following theorem, generalising several previous edge universality results listed in the introduction. For definiteness we only state and prove the result for regular right-edges. The corresponding statement for left-edges can be proven along the same lines.

Theorem 9.2.7 (Edge Universality)**.** *Under the Assumptions (A)–(F) the following statement holds true. Assume that* $E \in \mathbb{R}$ *is a regular right-edge of* ϱ *with a gap of size* c *for some* $c > 0$, *i.e.,* $\varrho([E, E + c]) = \{0\}$. *Then we have a square root edge of the form* $\varrho(x) = \gamma^{3/2} \sqrt{(E - x)_+}/\pi + \mathcal{O}(|E - x|)$ *for some* $\gamma > 0$. *The integer (see Remark 9.2.6)* $i_0 := N \int_{-\infty}^{E} \varrho(x) \mathrm{d}x$ *labels the largest eigenvalue* λ_{i_0} *close to the band* *edge E with high probability. Furthermore, for test functions $F\colon \mathbb{R}^{k+1} \to \mathbb{R}$ such that $\|F\|_\infty + \|\nabla F\|_\infty \le C < \infty$ we have*

$$\Big| \mathbb{E}\left[F\Big(\gamma N^{2/3}(\lambda_{i_0} - E), \dots, \gamma N^{2/3}(\lambda_{i_0-k} - E)\Big)\right] - \mathbb{E}\left[F\Big(N^{2/3}(\mu_N - 2), \dots, N^{2/3}(\mu_{N-k} - 2)\Big)\right] \Big| \lesssim N^{-\epsilon}$$

for some $\epsilon > 0$. Here $\mu_1, \dots, \mu_N$ are the eigenvalues of a standard GUE/GOE matrix, depending on the symmetry class of H.

From Theorem 9.2.7 we can immediately conclude that the eigenvalues of H near the regular edges follow the Tracy-Widom distribution. We remark that the direct analogue of Theorem 9.2.7 does not hold true for invariant β-ensembles with a *multi-cut* density. This is due to the fact that the number of particles close to a band of the self-consistent density, commonly known as the *filling fraction*, is known to be a fluctuating quantity for general classes of potentials. We refer the reader to [37] for a description of this phenomenon, to [117, 127] for non-Gaussian linear statistics in the multi-cut regime and to [41] for results on the fluctuations of filling fractions. Variants of Theorem 9.2.7 which allow for a relabelling of eigenvalues for invariant β-ensembles can be found in [30, 118].

9.3. Proof of the local law

The proof of a local law consists of three largely separate arguments. The first part concerns the analysis of the stability operator $\mathcal{B} := 1 - M\mathcal{S}[\cdot]M$ and shape analysis of the solution M to (9.2.1). The second part is proving that the resolvent G is indeed an approximate solution to (9.2.1) in the sense that

$$D := 1 + (z - A + \mathcal{S}[G])G = WG + \mathcal{S}[G]G \tag{9.3.1}$$

is small. Finally, the third part consists of a bootstrap argument starting far away from the real axis and iteratively lowering the imaginary part $\eta = \operatorname{Im} z$ of the spectral parameter while maintaining the desired bound on $G - M$.

For brevity we will carry out the proof of Theorem 9.2.2 for $|z| \lesssim 1$. Following the very same steps also proves the general result but requires carrying correction terms for the large z regime in many estimates. Since the large z-regime is already covered by the results from [56] we focus on the $|z| \lesssim 1$ regime in the present paper.

9.3.1. Stability. We denote the right-eigenmatrix corresponding to an, in absolute value, smallest eigenvalue β of $\mathcal{B}$ by B, i.e., $\mathcal{B}[B] = \beta B$, and the corresponding left-eigenmatrix and spectral projections by P and $\mathcal{P} = \langle P, \cdot \rangle B$, $\mathcal{Q} := 1 - \mathcal{P}$ with $\langle P, B \rangle = 1$. From (9.2.1) and (9.3.1) we have

$$\mathcal{B}[G - M] = -MD + M\mathcal{S}[G - M](G - M). \tag{9.3.2}$$

We note that $\mathcal{B}^{-1}$ is unstable in some particular direction near the edge, which is why we separate this unstable direction and establish bounds in terms of $\Theta := \langle P, G - M \rangle$ and D from (9.3.2). This separation is not necessary away the edge, but to keep our presentation shorter, we refrain from distinguishing these two cases and we just mimic the edge proof for the bulk as well. We begin by collecting some qualitative [96] and quantitative (cf. Chapter 8 and [6]) information about the MDE. We recall the definition of $\kappa = \kappa(z)$ in Theorem 9.2.2 as the distance of $\operatorname{Re} z$ to $\partial \operatorname{supp} \varrho$.

Proposition 9.3.1 (Stability of MDE and properties of the solution)**.** *The following hold true under Assumption (A)–(E).*

(i) The MDE (9.2.1) *has a unique solution $M = M(z)$ for all $z \in \mathbb{H}$ and moreover the map $z \mapsto M(z)$ is holomorphic.*
(ii) The holomorphic function $\langle M \rangle : \mathbb{H} \to \mathbb{H}$ is the Stieltjes transform of a compactly supported probability measure μ on $\mathbb{R}$.
(iii) The measure μ from (ii) is absolutely continuous with respect to the Lebesgue measure and has a continuous density $\varrho\colon \mathbb{R} \to [0, \infty)$, called the self-consistent density of states, which is also real analytic on the open set $\{\, \varrho > 0 \,\}$.
(iv) If $d_{\mathrm{cusp}} \ge \delta$ and $d_{\mathrm{M}} \ge \delta$ for some $\delta > 0$ and $|z| \lesssim 1$, then $\varrho(z) \sim_\delta \sqrt{\kappa + \eta}$ for $\operatorname{Re} z \in \operatorname{supp} \varrho$*, and $\varrho(z) \sim_\delta \eta/\sqrt{\kappa + \eta}$ for* $\operatorname{Re} z \notin \operatorname{supp} \varrho$*.*

(v) If $d_{\mathrm{cusp}} \geq \delta$ and $d_{\mathrm{M}} \geq \delta$ for some $\delta > 0$ and $|z| \lesssim 1$, there exist P, B such that we have the bounds on the stability operator and its unstable direction

$$\left\|\mathcal{B}^{-1}\right\|_{hs\to hs} \lesssim 1/\sqrt{\kappa+\eta}, \qquad \left\|\mathcal{B}^{-1}\mathcal{Q}\right\|_{hs\to hs} + \|B\| + \|P\| \leq_\delta 1,$$
$$|\langle P, M\mathcal{S}[B]B\rangle| + |\beta| \sim_\delta 1, \qquad |\beta| \sim_\delta \sqrt{\kappa+\eta}.$$

Proof. Claims (i)–(iii) follow directly from [96] and [6]. In order to conclude (iv)–(v) from Chapter 8, we specialize its setup by choosing $\mathcal{A} = \mathbb{C}^{N\times N}$ and $\langle\,\cdot\,\rangle = N^{-1}\operatorname{Tr}$ in Chapter 8. Moreover, we note that ϱ, $\mathcal{S}$, P, B, $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{B}$ and $\|\cdot\|_{\mathrm{hs}\to\mathrm{hs}}$ are denoted by ρ, S, l, b, P, Q, B and $\|\cdot\|_2$, respectively, in Chapter 8. We also observe that $d_{\mathrm{cusp}} \geq \delta$, $d_{\mathrm{M}} \geq \delta$ implies that $\operatorname{Re} z$ is either in the spectral bulk, close to a regular edge or well away from $\operatorname{supp}\varrho$. Thus, (iv) follows from (8.7.71a) in Chapter 8. Furthermore, whenever $\sqrt{\kappa+\eta} \ll 1$, then it follows that $|\langle P, M\mathcal{S}[B]B\rangle| \sim 1$ from (8.7.72) in Chapter 8 by the normalization from Corollary 8.5.2 in Chapter 8. This yields the third bound in (v). The first and the last bound in (v) are shown in (8.7.73) in Chapter 8. The second bound in (v) is a consequence of (8.5.15) and (8.5.16) in Chapter 8. We note that if $\sqrt{\kappa+\eta} \gtrsim 1$ then the choice of P, B is of no particular importance as then already $\|\mathcal{B}^{-1}\| \lesssim 1$. □

We now design a suitable norm following [56]. For cumulants of matrix elements $\kappa(w_{ab}, w_{cd})$ we use the short-hand notation $\kappa(ab, cd)$. We also use the short-hand notation $\kappa(\boldsymbol{x}b, cd)$ for the $\boldsymbol{x} = (x_a)_{a\in[N]}$-weighted linear combination $\sum_a x_a \kappa(ab, cd)$ of such cumulants. We use the notation that replacing an index in a scalar quantity by a dot $(\cdot)$ refers to the corresponding vector, e.g. $A_{a\cdot}$ is a short-hand notation for the vector $(A_{ab})_{b\in[N]}$. We fix two vectors $\boldsymbol{x}, \boldsymbol{y}$ and some large integer K and define the sets

$$I_0 := \{\,\boldsymbol{x}, \boldsymbol{y}\,\} \cup \{\, e_a, P_{a\cdot}^* \mid a \in [N]\,\},$$
$$I_{k+1} := I_k \cup \{\, M\boldsymbol{u} \mid \boldsymbol{u} \in I_k\,\} \cup \{\, \kappa_c((M\boldsymbol{u})a, b\cdot), \kappa_d((M\boldsymbol{u})a, \cdot b) \mid \boldsymbol{u} \in I_k,\ a, b \in [N]\,\},$$

where $\kappa_c + \kappa_d = \kappa$ is a decomposition of κ according to the Hermitian symmetry. Due to Assumption (CD) such a decomposition exists in a way that the operator norms of the matrices $\|\kappa_d(\boldsymbol{x}a, \cdot b)\|$ and $\|\kappa_c(\boldsymbol{x}a, b\cdot)\|$, indexed by (a, b), are bounded uniformly in $\boldsymbol{x}$

with $\|\boldsymbol{x}\| \le 1$. We now define the norm

$$\|R\|_* = \|R\|_*^{K,\boldsymbol{x},\boldsymbol{y}} := \sum_{0\le k<K} N^{-k/2K} \|R\|_{I_k} + N^{-1/2} \max_{\boldsymbol{u}\in I_K} \frac{\|R_{\cdot\boldsymbol{u}}\|}{\|\boldsymbol{u}\|}, \qquad \|R\|_I := \max_{\boldsymbol{u},\boldsymbol{v}\in I} \frac{|R_{\boldsymbol{u}\boldsymbol{v}}|}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|}.$$

Remark 9.3.2. We remark that compared to [56], the sets I_k contain some additional vectors generated by the vectors of the form $P_{a\cdot}^*$ in I_0. This addition is necessary to control the spectral projection $\mathcal{P}$ in the $\|\cdot\|_*$-norm. We note, however, that the precise form of the sets I_k were not important for the proofs in [56]. It was only used that the sets contain deterministic vectors, and that their cardinality grows at most as some finite power $|I_k| \lesssim N^{C_k}$ of N.

In terms of this norm we obtain the following easy estimate on $G - M$ in terms of its projection $\Theta = \langle P, G-M\rangle$ onto the unstable direction of the stability operator $\mathcal{B}$. We remark that if the, in absolute value, smallest eigenvalue of $\mathcal{B}$ is of order 1, then this projection onto the corresponding direction is not necessary.

Proposition 9.3.3. *For fixed z such that $\|G-M\|_* \lesssim N^{-3/K}$ there are deterministic matrices R_1, R_2 with norm $\lesssim 1$ such that*

$$G - M = \Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD] + \mathcal{E}, \qquad \|\mathcal{E}\|_* \lesssim N^{2/K}(|\Theta|^2 + \|D\|_*^2), \tag{9.3.3a}$$

with an error term $\mathcal{E}$, where Θ satisfies the approximate quadratic equation

$$\xi_1\Theta + \xi_2\Theta^2 = \mathcal{O}\left(N^{2/K}\|D\|_*^2 + |\langle R_1 D\rangle| + |\langle R_2 D\rangle|\right) \tag{9.3.3b}$$

with

$$|\xi_1| \sim \sqrt{\eta+\kappa}, \qquad |\xi_1| + |\xi_2| \sim 1$$

and any implied constants are uniform in $\boldsymbol{x}, \boldsymbol{y}$ and $z \in \mathbb{D}$.

Proof. We begin with an auxiliary lemma about the $\|\cdot\|_*$-norm of some important quantities, the proof of which we defer to Section 9.5 below.

Lemma 9.3.4. *Depending only on the model parameters we have the estimates for any* $R \in \mathbb{C}^{N\times N}$,

$$\|M\mathcal{S}[R]R\|_* \lesssim N^{1/2K}\|R\|_*^2, \qquad \|MR\|_* \lesssim N^{1/2K}\|R\|_*,$$
$$\|\mathcal{Q}\|_{*\to*} \lesssim 1, \qquad \left\|\mathcal{B}^{-1}\mathcal{Q}\right\|_{*\to*} \lesssim 1.$$

Decomposing $G-M = \mathcal{P}[G-M]+\mathcal{Q}[G-M]$ and inverting $\mathcal{B}$ in (9.3.2) on the range of $\mathcal{Q}$ yields

$$\begin{aligned} G-M &= \Theta B + \mathcal{Q}[G-M] = \Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD] + \mathcal{O}\left(N^{1/2K}\|G-M\|_*^2\right) \\ &= \Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD] + \mathcal{O}\left(N^{3/2K}(|\Theta|^2+\|D\|_*^2)\right), \end{aligned}$$

where $\mathcal{O}(\cdot)$ is meant with respect to the $\|\cdot\|_*$-norm and the second equality followed by iteration, Lemma 9.3.4 and the assumption on $\|G-M\|_*$. Going back to the original equation (9.3.2) we find

$$\begin{aligned} \beta\Theta B + \mathcal{B}\mathcal{Q}[G-M] = &- MD + M\mathcal{S}[\Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD]](\Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD]) \\ &+ \mathcal{O}\left(N^{2/K}(|\Theta|^3+\|D\|_*^3)\right) \end{aligned}$$

and thus by projecting with $\mathcal{P}$ we arrive at the quadratic equation

$$\begin{aligned} \mu_0 - \mu_1\Theta + \mu_2\Theta^2 &= \mathcal{O}\left(N^{2/K}(|\Theta|^3+\|D\|_*^3)\right), \\ \mu_0 &= \langle P, M\mathcal{S}[\mathcal{B}^{-1}\mathcal{Q}[MD]]\mathcal{B}^{-1}\mathcal{Q}[MD] - MD\rangle, \\ \mu_1 &= \langle P, M\mathcal{S}[B]\mathcal{B}^{-1}\mathcal{Q}[MD] + M\mathcal{S}[\mathcal{B}^{-1}\mathcal{Q}[MD]]B\rangle + \beta, \\ \mu_2 &= \langle P, M\mathcal{S}[B]B\rangle. \end{aligned}$$

We now proceed by analysing the coefficients in this quadratic equation. We estimate the quadratic term in μ_0 directly by $N^{2/K}\|D\|_*^2$, while we write the linear term as $\langle R_1 D\rangle$ for the deterministic $R_1 := -M^*P$ with $\|R_1\| \lesssim 1$. For the linear coefficient μ_1 we similarly find a deterministic matrix R_2 such that $\|R_2\| \lesssim 1$ and $\mu_1 = \langle R_2 D\rangle + \beta$. Finally, we find from Proposition 9.3.1(v) that $|\mu_2| + |\beta| \sim 1$ and $|\beta| \sim \sqrt{\kappa+\eta}$. By incorporating the $|\Theta|\,N^{2/K}$ term into ξ_2 we obtain (9.3.3b). □

9.3.2. Probabilistic bound. We now collect the averaged and isotropic bound on D from [56]. We first introduce a commonly used (see, e.g., [60]) notion of high-probability bound.

Definition 9.3.5 (Stochastic Domination)**.** If

$$X = \left(X^{(N)}(u) \,|\, N \in \mathbb{N}, u \in U^{(N)}\right) \quad \text{and} \quad Y = \left(Y^{(N)}(u) \,|\, N \in \mathbb{N}, u \in U^{(N)}\right)$$

are families of non-negative random variables indexed by N, and possibly some parameter u, then we say that X is stochastically dominated by Y, if for all $\epsilon, D > 0$ we have

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[X^{(N)}(u) > N^{\epsilon} Y^{(N)}(u)\right] \leq N^{-D}$$

for large enough $N \geq N_0(\epsilon, D)$. In this case we use the notation $X \prec Y$.

It can be checked (see [60, Lemma 4.4]) that $\prec$ satisfies the usual arithmetic properties, e.g. if $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then also $X_1 + X_2 \prec Y_1 + Y_2$ and $X_1 X_2 \prec Y_1 Y_2$. To formulate the result compactly we also introduce the notations

$$\begin{aligned} |R| \prec \Lambda \text{ in } \mathbb{D} \quad &\Longleftrightarrow \quad \|R\|_*^{K,\boldsymbol{x},\boldsymbol{y}} \prec \Lambda \text{ unif. in } \boldsymbol{x}, \boldsymbol{y} \text{ and } z \in \mathbb{D}, \\ |R|_{\mathrm{av}} \prec \Lambda \text{ in } \mathbb{D} \quad &\Longleftrightarrow \quad \frac{|\langle BR \rangle|}{\|B\|} \prec \Lambda \text{ unif. in } B \text{ and } z \in \mathbb{D} \end{aligned} \tag{9.3.4}$$

for random matrices R and a deterministic control parameter $\Lambda = \Lambda(z)$, where $B, \boldsymbol{x}, \boldsymbol{y}$ are deterministic matrices and vectors. We also define an isotropic high-moment norm, already used in [56], for $p \geq 1$ and a random matrix R,

$$\|R\|_p := \sup_{\boldsymbol{x},\boldsymbol{y}} \frac{\left(\mathbb{E}\, |\langle \boldsymbol{x}, R\boldsymbol{y} \rangle|^p\,\right)^{1/p}}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|}.$$

Proposition 9.3.6 (Bound on the Error)**.** *Under the Assumptions (A)–(E) there exists a constant C such that for any fixed vectors $\boldsymbol{x}, \boldsymbol{y}$ and matrices B and spectral parameters $|z| \lesssim 1$, and any $p \geq 1$, $\epsilon > 0$,*

$$\frac{\|\langle \boldsymbol{x}, D\boldsymbol{y} \rangle\|_p}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|} \leq_{\epsilon,p} N^{\epsilon} \sqrt{\frac{\|\mathrm{Im}\, G\|_q}{N \mathrm{Im}\, z}} \left(1 + \|G\|_q\right)^C \left(1 + \frac{\|G\|_q}{N^{\mu}}\right)^{Cp} \tag{9.3.5a}$$

$$\frac{\|\langle BD \rangle\|_p}{\|B\|} \leq_{\epsilon,p} N^{\epsilon} \frac{\|\mathrm{Im}\, G\|_q}{N \mathrm{Im}\, z} \left(1 + \|G\|_q\right)^C \left(1 + \frac{\|G\|_q}{N^{\mu}}\right)^{Cp}, \tag{9.3.5b}$$

where $q := Cp^4/\epsilon$. *Here* $\mu > 0$ *depends on* s *in Assumption (CD). In particular, if* $|G - M| \prec \Lambda \lesssim 1$, *then*

$$|D| \prec \sqrt{\frac{\varrho + \Lambda}{N\eta}}, \qquad |D|_{av} \prec \frac{\varrho + \Lambda}{N\eta}. \tag{9.3.5c}$$

Proof. This follows from combining [56, Theorem 3.1], the following lemma[3] from [56, Lemma 4.4] and $\|M\| \le M_*$. □

Lemma 9.3.7. *Let* R *be a random matrix and* Φ *a deterministic control parameter. Then the following implications hold:*

(i) If $\Phi \ge N^{-C}$, $\|R\| \le N^C$ *and* $|R_{\boldsymbol{xy}}| \prec \Phi \|\boldsymbol{x}\| \|\boldsymbol{y}\|$ *for all* $\boldsymbol{x}, \boldsymbol{y}$ *and some* C, *then* $\|R\|_p \le_{p,\epsilon} N^\epsilon \Phi$ *for all* $\epsilon > 0, p \ge 1$.

(ii) Conversely, if $\|R\|_p \le_{p,\epsilon} N^\epsilon \Phi$ *for all* $\epsilon > 0, p \ge 1$, *then* $\|R\|_*^{K,\boldsymbol{x},\boldsymbol{y}} \prec \Phi$ *for any fixed* $K \in \mathbb{N}$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$.

9.3.3. Bootstrapping. We now fix $\gamma > 0$ and start with the proof of Theorem 9.2.2. Phrased in terms of the $\|\cdot\|_*$-norm we will prove

$$\begin{aligned} |G - M| &\prec N^{2/K} \left(\sqrt{\frac{\varrho}{N\eta}} + \frac{1}{N\eta} \right), \\ |G - M|_{\text{av}} &\prec N^{2/K} \begin{cases} \frac{1}{N\eta} & \operatorname{Re} z \in \operatorname{supp} \varrho \\ \frac{1}{N(\kappa+\eta)} + \frac{N^{2/K}}{(N\eta)^2\sqrt{\kappa+\eta}} & \operatorname{Re} z \notin \operatorname{supp} \varrho \end{cases} \quad \text{in} \quad \mathbb{D}, \end{aligned} \tag{9.3.6}$$

for $\mathbb{D} = \mathbb{D}_\gamma$ and $K \gg 1/\gamma$, i.e., for $K\gamma$ sufficiently large. In order to prove (9.3.6) we use the following iteration procedure.

Proposition 9.3.8. *There exists a constant* $\gamma_s > 0$ *depending only on* K *and* γ *such that* (9.3.6) *for* $\mathbb{D} = \mathbb{D}_{\gamma_0}$ *with* $\gamma_0 > \gamma$ *implies* (9.3.6) *also for* $\mathbb{D} = \mathbb{D}_{\gamma_1}$ *with* $\gamma_1 := \max\{\gamma, \gamma_0 - \gamma_s\}$.

[3]C.f. Remark 9.3.2, where we argue that the proof of [56] about $\|\cdot\|_*$ hold true verbatim in the present case despite the slightly larger sets I_k.

Proof of (9.3.6) for $\mathbb{D} = \mathbb{D}_\gamma$, assuming Proposition 9.3.8. For $\mathbb{D} = \mathbb{D}_1$ we have (9.3.6) by [56, Theorem 2.1][4]. We then iteratively apply Proposition 9.3.8 finitely many times until we have shown (9.3.6) for $\mathbb{D} = \mathbb{D}_\gamma$. □

Proof of Proposition 9.3.8. We now suppose that (9.3.6) has been proven for some $\mathbb{D} = \mathbb{D}_{\gamma_0}$ and aim at proving (9.3.6) for $\mathbb{D} = \mathbb{D}_{\gamma_1}$ for some $\gamma_1 = \gamma_0 - \gamma_s$, $0 < \gamma_s \ll \gamma$. The proof has two stages. Firstly, we will establish the rough bounds

$$|\Theta| \prec N^{-5/K} \quad \text{and} \quad |G - M| \prec N^{-5/K} \quad \text{in} \quad \mathbb{D}_{\gamma_1}, \tag{9.3.7}$$

and then in the second stage improve upon this bound iteratively until we reach (9.3.6) for $\mathbb{D} = \mathbb{D}_{\gamma_1}$.

Rough bound. By (9.3.6), Lemma 9.3.7 and monotonicity of the map $(0,\infty) \to \mathbb{R}, \eta \mapsto \eta \left\| G(E + \mathrm{i}\eta) \right\|_p$ (see e.g. (77) in [56]) we find $\|G\|_p \le_{\epsilon,p} N^{\epsilon+\gamma_s} \le N^{2\gamma_s}$ in $\mathbb{D}_{\gamma_1}$. As long as $2\gamma_s < \mu$ we thus have

$$\|D\|_p \le_{\epsilon,p} \frac{N^{\epsilon+2C\gamma_s+\gamma_s}}{\sqrt{N\eta}} \le \frac{N^{\gamma_s(2+2C)}}{\sqrt{N\eta}}, \qquad \|\langle BD\rangle\|_p \le_{\epsilon,p} \|B\| \frac{N^{\epsilon+2\gamma_s+2\gamma_s C}}{N\eta} \le \|B\| \frac{N^{\gamma_s(3+2C)}}{N\eta}.$$

We now fix $\boldsymbol{x}, \boldsymbol{y}$ and it follows from (9.3.3b) that

$$\left|\xi_1\Theta + \xi_2\Theta^2\right| \prec \frac{N^{2\gamma_s(3+2C)+2/K}}{N\eta} \quad \text{in} \quad \mathbb{D}_{\gamma_1}$$

and consequently by Lipschitz continuity of the lhs. with a Lipschitz constant of $\eta^{-2} \le N^2$, and choosing K, γ_s large and respectively small enough depending on γ we find that with high probability $|\xi_1\Theta + \xi_2\Theta^2| \le N^{-10/K}$ in all of $\mathbb{D}_{\gamma_1}$. The following lemma translates the bound on $|\xi_1\Theta + \xi_2\Theta^2|$ into a bound on $|\Theta|$.

Lemma 9.3.9. *Let $d = d(\eta)$ be a monotonically decreasing function in $\eta \ge 1/N$ and assume $0 \le d \lesssim N^{-\epsilon}$ for some $\epsilon > 0$. Suppose that*

$$\left|\xi_1\Theta + \xi_2\Theta^2\right| \lesssim d \quad \textit{for all } z \in \mathbb{D}, \qquad \textit{and} \qquad |\Theta| \lesssim \min\left\{\frac{d}{\sqrt{\kappa+\eta}}, \sqrt{d}\right\} \quad \textit{for some } z_0,$$

then also $|\Theta| \lesssim \min\{d/\sqrt{\kappa+\eta}, \sqrt{d}\}$ for all $z' \in \mathbb{D}$ with $\operatorname{Re} z' = \operatorname{Re} z_0$ and $\operatorname{Im} z' < \operatorname{Im} z_0$.

[4]We remark referring to [56] for the initial bound is purely a matter of brevity and convenience. Equally well we could also prove (9.3.6) in some initial domain, say, $\mathbb{D}_2$ from scratch, where we have the trivial bound $\|G - M\| \le \frac{2}{N}$. Using this rough bound we could then iteratively improve the bound as detailed in the paragraph *Strong bound* below, until (9.3.6) follows in $\mathbb{D}_2$.

Proof. This proof is basically identical to the analysis of the solutions to the same approximate quadratic equation, as appeared in various previous works, see e.g. [67]. In the spectral bulk this is trivial since then $|\xi_1| \sim \sqrt{\kappa+\eta} \sim 1$. Near a spectral edge we observe that $(\kappa+\eta)/d$ is monotonically increasing in η. First suppose that $(\kappa+\eta)/d \gg 1$ from which it follows that $|\Theta| \lesssim d/\sqrt{\kappa+\eta} \lesssim \sqrt{d}$ in the relevant branch determined by the given estimate on Θ at z_0. Now suppose that below some η-threshold we have $(\kappa+\eta)/d \lesssim 1$. Then we find $|\Theta| \lesssim \sqrt{\kappa+\eta}+\sqrt{d} \lesssim \sqrt{d} \lesssim d/\sqrt{\kappa+\eta}$ and the claim follows also in this regime. □

Since (9.3.7) holds in $\mathbb{D}_{\gamma_0}$ and $1/N\eta \le N^{-100/K}$, we know

$$|\Theta| \le \min\{N^{-10/K}/\sqrt{\kappa+\eta}, N^{-5/K}\}$$

and therefore can conclude the rough bound $|\Theta| \prec N^{-5/K}$ in all of $\mathbb{D}_{\gamma_1}$ by Lemma 9.3.9 with $d = N^{-10/K}$. Consequently we have also that

$$\|G-M\|_* \mathbf{1}(\|G-M\|_* \le N^{-3/K}) \prec N^{-5/K} \quad \text{in} \quad \mathbb{D}_{\gamma_1}.$$

Due to this gap in the possible values for $\|G-M\|_*$ it follows from a standard continuity argument that $\|G-M\|_* \prec N^{-5/K}$ and therefore since $\boldsymbol{x}, \boldsymbol{y}$ were arbitrary, $|\Theta| \prec N^{-5/K}$ and $|G-M| \prec N^{-5/K}$ in all of $\mathbb{D}_{\gamma_1}$.

Strong bound. All of the following bounds hold uniformly in the domain $\mathbb{D}_{\gamma_1}$ which is why we suppress this qualifier. By combining Proposition 9.3.3 and Proposition 9.3.6 we find for deterministic $0 \le \theta \le \Lambda \le N^{-3/K}$ under the assumptions $|\Theta| \prec \theta$, $|G-M| \prec \Lambda$, that

$$|G-M| \prec \theta + N^{2/K}\sqrt{\frac{\varrho+\Lambda}{N\eta}}, \qquad \left|\xi_1\Theta + \xi_2\Theta^2\right| \prec N^{2/K}\frac{\varrho+\Lambda}{N\eta}. \tag{9.3.8}$$

The bound on $|G-M|$ in (9.3.8) is a self-improving bound and we find after iteration that

$$|G-M| \prec \theta + N^{2/K}\left(\frac{1}{N\eta} + \sqrt{\frac{\varrho+\theta}{N\eta}}\right).$$

Hence, we have

$$\left|\xi_1\Theta + \xi_2\Theta^2\right| \prec N^{2/K}\frac{\varrho+\theta}{N\eta} + N^{4/K}\frac{1}{(N\eta)^2}.$$

We now distinguish whether $\operatorname{Re} z$ is inside or outside the spectrum. Inside we have $\varrho \sim \sqrt{\kappa+\eta}$, so we fix θ and use Lemma 9.3.9 with $d = N^{2/K}(\sqrt{\kappa+\eta}+\theta)/(N\eta)+N^{4/K}/(N\eta)^2$ to conclude $|\Theta| \prec \min\{d/\sqrt{\kappa+\eta}, \sqrt{d}\}$ from the input assumption $|\Theta| \prec N^{2/K}/N\eta$ in $\mathbb{D}_{\gamma_0}$. Iterating this bound, we obtain

$$|\Theta| \prec N^{2/K}\frac{1}{N\eta}, \quad \text{hence} \quad |G-M| \prec N^{2/K}\left(\sqrt{\frac{\varrho}{N\eta}}+\frac{1}{N\eta}\right).$$

By an analogous argument, outside of the spectrum we have an improved bound on Θ

$$|\Theta| \prec N^{2/K}\frac{1}{N(\kappa+\eta)} + N^{4/K}\frac{1}{(N\eta)^2\sqrt{\kappa+\eta}},$$

because $\varrho \sim \eta/\sqrt{\kappa+\eta}$. Finally, for the claimed bound on $|G-M|_{\mathrm{av}}$ we use (9.3.3a) in order to obtain a bound on $|G-M|_{\mathrm{av}}$ in terms of a bound on Θ. □

Due to (9.3.6), we now have all the ingredients to prove the local law, as well as delocalization of eigenvectors, and the absence of eigenvalues away from the support of ϱ.

Proof of Theorem 9.2.2, Corollary 9.2.3 and Corollary 9.2.4. The local law inside the spectrum (9.2.3a)–(9.2.3b) follows immediately from (9.3.6). Now we prove Corollary 9.2.3. If there exists an eigenvalue λ with $\operatorname{dist}(\lambda, \operatorname{supp}\varrho) > N^{-2/3+\omega}$, then at, say, $z = \lambda + \mathrm{i}N^{-4/5}$ we have $|\langle G-M\rangle| \ge cN^{-1/5}$. On the other hand we know from the improved local law (9.3.6) that with high probability $|\langle G-M\rangle| \le N^{-1/4}$ and we obtain the claim.

We now turn to the proof of Corollary 9.2.4. For the eigenvectors $\boldsymbol{u}_k$ and eigenvalues λ_k of H we find from the spectral decomposition and the local law with high probability

$$1 \gtrsim \operatorname{Im}\langle \boldsymbol{x}, G\boldsymbol{x}\rangle = \eta\sum_k \frac{|\langle \boldsymbol{x}, \boldsymbol{u}_k\rangle|^2}{(E-\lambda_k)^2+\eta^2} \ge \frac{|\langle \boldsymbol{x}, \boldsymbol{u}_k\rangle|^2}{\eta}$$

for $z = E + \mathrm{i}\eta$ and any normalised $\boldsymbol{x} \in \mathbb{C}^N$, where the last inequality followed assuming that E is chosen η-close to λ_k. With the choice $\eta = N^{-1+\gamma}$ for arbitrarily small $\gamma > 0$ the claim follows. Note that for this proof only (9.2.3a) of Theorem 9.2.2 was used.

Finally, we establish (9.2.3c) and consider z with $|z| \lesssim 1$, $\operatorname{dist}(\operatorname{Re} z, \operatorname{supp}\varrho) \ge N^{-2/3+\omega}$, $d_{\mathrm{cusp}} \ge \delta$, $d_{\mathrm{M}} \ge \delta$ and $\boldsymbol{x}, \boldsymbol{y}, B$ fixed. We note that the regime $|z| \not\lesssim 1$ was already covered in [56] and we therefore do not have to track the large $|z|$-dependence again in the present paper. As in the proof of [7, Corollary 1.11], the optimal local law (9.3.6) implies rigidity up to the edge as formulated in Corollary 9.2.5. The only difference is that this standard argument proves (9.2.4b) only if the supremum is restricted to $E \in \operatorname{supp} \varrho$ with $\operatorname{dist}(E, \partial \operatorname{supp} \varrho) \geq N^{-2/3+\epsilon}$. The cause for this restriction is a possible mismatch of the labelling of the edge eigenvalues, in other words the precise location of N^ϵ eigenvalues near an internal gap is not established yet; they may belong to either band adjacent to this gap. This shortcoming will be remedied by the band rigidity in the proof of Corollary 9.2.5 below. However, for the current argument, the imprecise location of N^ϵ eigenvalues does not matter. In fact, already from this version of rigidity, together with the delocalisation of eigenvectors (Corollary 9.2.4) and the absence of eigenvalues outside of the spectrum by Corollary 9.2.3 we have, at $z = E + \mathrm{i}\eta$ (recall that we consider z with $d_{\mathrm{cusp}} \geq \delta$, $d_{\mathrm{M}} \geq \delta$ and $\operatorname{dist}(\operatorname{Re} z, \operatorname{supp} \varrho) \geq N^{-2/3+\omega}$),

$$\operatorname{Im} \langle \boldsymbol{x}, G(z)\boldsymbol{x} \rangle = \eta \sum_k \frac{|\langle \boldsymbol{x}, \boldsymbol{u}_k \rangle|^2}{(E-\lambda_k)^2 + \eta^2} \prec \frac{1}{N} \sum_k \frac{\eta}{(E-\lambda_k)^2 + \eta^2} \prec \int_{\mathbb{R}} \frac{\eta\, \varrho(x)\mathrm{d}x}{|E-x|^2 + \eta^2}$$

for any normalised vector $\boldsymbol{x}$. From the square root behaviour of ϱ at the edge and $\kappa(z) \geq N^{-2/3+\omega}$ we can easily infer $\|\operatorname{Im} G\|_* \prec \eta/\sqrt{\kappa+\eta}$. Therefore it follows from Proposition 9.3.6 that $\|D\|_*^2 + |\langle RD \rangle| \prec 1/(N\sqrt{\kappa+\eta})$ and from (9.3.3b) and Lemma 9.3.9 that $|\Theta| \prec N^{2/K-1}/(\kappa+\eta)$. Finally, we thus obtain,

$$|G - M|_{\mathrm{av}} \prec \frac{N^{2/K}}{N(\kappa+\eta)} + \frac{N^{2/K}}{N\sqrt{\kappa+\eta}} \lesssim N^{2/K} \frac{1}{N(\kappa+\eta)}$$

from (9.3.3a) and (9.2.3c) follows. □

Proof of Corollary 9.2.5. We begin with the proof of (9.2.4a) and consider a flow that interpolates between $H = H_0$ and a deterministic matrix H_1. Fix $E \notin \operatorname{supp} \varrho$ with $\operatorname{dist}(E, \operatorname{supp} \varrho) \geq \epsilon$. We set

$$H_t := \sqrt{1-t}W + A_t, \quad A_t := A - t\mathcal{S}[M(E)], \quad \mathcal{S}_t := (1-t)\mathcal{S}, \tag{9.3.9}$$

for any $t \in [0,1]$. The MDE corresponding to H_t is $1 + (z - A_t + \mathcal{S}_t[M_t])M_t = 0$ and is designed in such a way that $M_t(E)$, the solution evaluated in E, is kept constant. The flow of solutions M_t was considered in the proof of Proposition 8.2.6 in Chapter 8, where it was shown that the self-consistent spectrum $\operatorname{supp} \varrho_t$ stays away from E uniformly along the flow, i.e., that $\operatorname{dist}(E, \operatorname{supp} \varrho_t) \geq_\epsilon 1$, see Lemma 8.8.1 (ii) in Chapter 8. We will now show that along the flow, with overwhelming probability, no eigenvalue crosses the spectral parameter E. More precisely we claim that

$$\mathbb{P}\Big(E \in \operatorname{Spec} H_t \text{ for some } t \in [0,1]\Big) \leq_\epsilon N^{-D} \tag{9.3.10}$$

for any $D > 0$. Since $H_0 = H$ and $H_1 = A - \mathcal{S}[M(E)]$, (9.3.10) implies that with overwhelming probability

$$|\operatorname{Spec} H \cap (-\infty, E)| = |\operatorname{Spec}(A - \mathcal{S}[M(E)]) \cap (-\infty, E)| = N \left\langle \mathbf{1}_{(-\infty,0)}(M(E)) \right\rangle,$$

where the last identity used the fact that

$$M(E) = (A - \mathcal{S}[M(E)] - E)^{-1}, \tag{9.3.11}$$

i.e., that $M(E)$ is the resolvent of $A - \mathcal{S}[M(E)]$ at spectral parameter E (see Lemma 8.8.1 (i) in Chapter 8). Now (9.2.4a) follows from Proposition 8.2.6 in Chapter 8, i.e., from

$$\langle \mathbf{1}_{(-\infty,0)}(M(E)) \rangle = \int_{-\infty}^{E} \varrho(\lambda) \mathrm{d}\lambda.$$

It remains to show (9.3.10). We first consider the regime of values t close to 1. Since E is separated away from $\operatorname{supp} \varrho$, and $M(E)$ is bounded we conclude from (9.3.11) that the spectrum of $A - \mathcal{S}[M(E)]$ is also separated away from E. Moreover, applying Corollary 9.2.3 to $H = W$ yields $\|W\| \leq C$ with overwhelming probability as the corresponding self-consistent density of states has compact support by Proposition 9.3.1 (ii). Since therefore H_t is a small perturbation of $A - \mathcal{S}[M(E)]$ as long as t is close to 1, we conclude that the spectrum of H_t is bounded away from E as well for every fixed $t \geq 1 - c$ for some small enough constant $c > 0$. We are thus left with the regime $t \leq 1 - c$, where the flatness condition from Assumption (E) is satisfied. In this regime we use Corollary 9.2.3 with $H = H_t$. Since $\operatorname{dist}(E, \operatorname{supp} \rho_t) \geq_\epsilon 1$ this corollary implies that the spectrum of H_t is bounded away from E with overwhelming probability for every fixed $t \leq 1 - c$. Applying a discrete union bound in t together with the Lipschitz continuity of the eigenvalues in t for the flow (9.3.9) on the set $\|W\| \leq C$ yields (9.3.10).

Finally, (9.2.4b) follows from the optimal local law as in the proof of Theorem 9.2.2 and Corollary 9.2.3 above. This time, however, (9.2.4a) ensures that there is no mismatch between location and label of eigenvalues close to internal edges. In the spectral bulk this potential discrepancy between label and location does not matter as (9.2.4b) allows for an N^{ϵ}-uncertainty. At the spectral edge, however, neighbouring eigenvalues can lie on opposite sides of a spectral gap and we need (9.2.4a) to make sure that each eigenvalue has, with high probability, a definite location with respect to the spectral gap. □

9.4. Proof of Universality

In order to prove Theorem 9.2.7, we define the Ornstein Uhlenbeck (OU) process starting from $H = H_0$ by

$$\mathrm{d}H_t = -\frac{1}{2}(H_t - A)\mathrm{d}t + \Sigma^{1/2}[\mathrm{d}B_t], \qquad \Sigma[R] := \mathbb{E}W\,\mathrm{Tr}(WR), \tag{9.4.1}$$

where B_t is a matrix of, up to symmetry, independent (real or complex, depending on the symmetry class of H) Brownian motions and $\Sigma^{1/2}$ is the square root of the positive definite operator $\Sigma : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N}$. We note that the same process has already been used in [6, 49, 56] to prove bulk universality. The proof now has two steps: Firstly, we will prove edge universality for H_t if $t \gg N^{-1/3}$ and then we will prove that for $t \ll N^{-1/6}$, the eigenvalues of H_t have the same k-point correlation functions as those of $H = H_0$.

9.4.1. Dyson Brownian Motion. The process (9.4.1) can be integrated, and we have

$$H_t - A = e^{-t/2}(H_0 - A) + \int_0^t e^{(s-t)/2}\Sigma^{1/2}[\mathrm{d}B_s], \qquad \int_0^t e^{(s-t)/2}\Sigma^{1/2}[\mathrm{d}B_s] \sim \mathcal{N}(0, (1-e^{-t})\Sigma).$$

The process is designed in such a way that it preserves expectation $\mathbb{E}H_t = A$ and covariances $\mathbf{Cov}\,(h^t_{ab}, h^t_{cd}) = \mathbf{Cov}\,(h_{ab}, h_{cd})$ along the flow. Due to the fullness Assumption (F) there exists a constant $c > 0$ such that $(1 - e^{-t})\Sigma - ct\Sigma^{\mathrm{GUE/GOE}} \geq 0$ for $t \leq 1$, where $\Sigma^{\mathrm{GOE/GUE}}$ denotes the covariance operator of the GOE/GUE ensembles. It follows that we can write

$$H_t = \widetilde{H}_t + \sqrt{ct}U, \qquad \kappa_t = \kappa - ct\kappa^{\mathrm{GOE/GUE}}, \qquad \mathbb{E}\widetilde{H}_t = A, \qquad U \sim \mathrm{GOE/GUE},$$

where κ_t here denotes the cumulants of $\widetilde{H}_t$, and U is chosen to be independent of $\widetilde{H}_t$. Due to the fact that Gaussian cumulants of degree more than 2 vanish, it is easy to check that $H_t, \widetilde{H}_t$ satisfy the assumptions of Theorem 9.2.2 uniformly in, say, $t \leq N^{-1/10}$. From now on we fix $t = N^{-1/3+\epsilon}$ with some small $\epsilon > 0$.

Since the MDE is purely determined by the first two moments of the corresponding random matrix, it follows that $G_t := (H_t - z)^{-1}$ is close to the same M in the sense of a local law for all t. For $\widetilde{G}_t := (\widetilde{H}_t - z)^{-1}$ we have the MDE

$$1 + (z - A + \mathcal{S}_t[M_t])M_t = 0, \qquad \mathcal{S}_t := \mathcal{S} - ct\mathcal{S}^{\mathrm{GOE/GUE}} \tag{9.4.2}$$

that can be viewed as a perturbation of the original MDE with $t = 0$. The corresponding self-consistent density of states is $\varrho_t(E) := \lim_{\eta \searrow 0} \operatorname{Im} \langle M_t(E + \mathrm{i}\eta) \rangle / \pi$. The fact that M_t remains bounded uniformly in $t \leq N^{-1/10}$ follows from the MDE perturbation result in Proposition 8.10.1 in Chapter 8 with $a_t := A$ and $S_t := \mathcal{S}_t$ as S_t is positivity-preserving and the condition on S_t in (8.10.1) in Chapter 8 is obviously satisfied for this choice of S_t due to $\left\| \mathcal{S}^{\mathrm{GOE/GUE}}[R] \right\| \lesssim \langle R \rangle$ for all positive definite R. In particular the shape analysis from Chapter 8 also applies to M_t.

The free convolutions of the empirical spectral density of $\widetilde{H}_t$ and ϱ_t with the semicircular distribution generated by $\sqrt{ct}U$ are given implicitly as the unique solutions to the equations

$$\widetilde{m}^t_{\mathrm{fc}}(z) = \langle \widetilde{G}_t(z + ct\widetilde{m}^t_{\mathrm{fc}}(z)) \rangle , \qquad m^t_{\mathrm{fc}}(z) = \langle M_t(z + ctm^t_{\mathrm{fc}}(z)) \rangle .$$

We denote the corresponding right-edges close to E by $\widetilde{E}_t$ and E_t. By differentiating the defining equations for m^t_{fc} and $\widetilde{m}^t_{\mathrm{fc}}$ we find

$$\begin{gathered} \frac{(m^t_{\mathrm{fc}})'(z)}{1 + ct(m^t_{\mathrm{fc}})'(z)} = \langle M_t'(\xi_t(z)) \rangle , \qquad \frac{(\widetilde{m}^t_{\mathrm{fc}})'(z)}{1 + ct(\widetilde{m}^t_{\mathrm{fc}})'(z)} = \langle \widetilde{G}_t'(\widetilde{\xi}_t(z)) \rangle , \\ \frac{(m^t_{\mathrm{fc}})''(z)}{(1 + ct(m^t_{\mathrm{fc}})'(z))^3} = \langle M_t''(\xi_t(z)) \rangle , \end{gathered} \tag{9.4.3a}$$

where $\xi_t(z) := z + ctm^t_{\mathrm{fc}}(z)$ and $\widetilde{\xi}_t(z) := z + ct\widetilde{m}^t_{\mathrm{fc}}(z)$. From the first two equalities in (9.4.3a) we conclude

$$1 = ct \langle M_t'(\xi_t(E_t)) \rangle , \qquad 1 = ct \langle \widetilde{G}_t'(\widetilde{\xi}_t(\widetilde{E}_t)) \rangle , \tag{9.4.3b}$$

by considering the $z \to E_t$ and $z \to \widetilde{E}_t$ limits and that $(m_{\rm fc}^t)'$, $(\widetilde{m}_{\rm fc}^t)'$ blow up at the edge due to the well-known square root behaviour of the density along the semicircular flow. We now compare the edge location and edge slope of the densities $\varrho_{\rm fc}^t$ and $\widetilde{\varrho}_{\rm fc}^t$ corresponding to $m_{\rm fc}^t$ and $\widetilde{m}_{\rm fc}^t$ with that of M. Very similar estimates for deformed Wigner ensembles have been used in [97]. We split the analysis into four claims.

Claim 1. $|E_t - E| \lesssim t/N$. Using that $\mathcal{S}^{\rm GUE}[R] = \langle R\rangle$, $\mathcal{S}^{\rm GOE}[R] = \langle R\rangle + R^t/N$ and (9.4.2) evaluated at $\xi_t(z)$, we find using the boundedness of M_t,

$$\begin{aligned}1 + (z - A + \mathcal{S}[M_t(\xi_t(z))])M_t(\xi_t(z)) &= ct\Big(\mathcal{S}^{\rm GOE/GUE}[M_t(\xi_t(z))] - \langle M_t(\xi_t(z))\rangle\Big)M_t(\xi_t(z))\\ &= \mathcal{O}\left(\frac{t}{N}\right).\end{aligned}$$

It thus follows that $M_t(\xi_t(z))$ approximately satisfies the MDE for M at z. By using the first bound in Proposition 9.3.1(v) expressing the stability of the MDE against small additive perturbations it follows that

$$\begin{aligned}\Big|m_{\rm fc}^t(z) - \langle M(z)\rangle\Big| = |\langle M_t(\xi_t(z)) - M(z)\rangle| &\lesssim \frac{t}{N\sqrt{\eta + \operatorname{dist}(\operatorname{Re} z, \partial \operatorname{supp} \varrho)}}\\ &\le \frac{t}{N\sqrt{\operatorname{dist}(\operatorname{Re} z, \partial \operatorname{supp} \varrho)}}.\end{aligned} \tag{9.4.4}$$

Suppose first that $E = E_t + \delta$ for some positive $\delta > 0$. Then $\sqrt{\delta} \lesssim \operatorname{Im} \langle M(E_t + \delta/2)\rangle \lesssim t/N\sqrt{\delta}$, where the first bound follows from the square root behaviour of ϱ at the edge E, while the second bound comes from (9.4.4) at $z = E_t + \delta/2$ and $\operatorname{Im} m_{\rm fc}^t(E_t + \delta/2) = 0$. We thus conclude $\delta \lesssim t/N$. If on the contrary $E = E_t - \delta$ for some $\delta > 0$, then with a similar argument $\sqrt{\delta} \lesssim \operatorname{Im} m_{\rm fc}^t(E + \delta/2) \lesssim t/N$ and we have $\delta \lesssim t/N$ also in this case and the claim follows.

Claim 2. $|\gamma_t - \gamma| \lesssim (t/N)^{1/4}$. From the third equality in (9.4.3a) we can relate the edge-slope of $m_{\rm fc}^t$ to M_t''. Indeed, if $\gamma_t^{3/2}$ denotes the slope, i.e., $\varrho_{\rm fc}^t(x) = \gamma_t^{3/2}\sqrt{(E_t - x)_+}/\pi + \scriptstyle\mathcal{O}\displaystyle(E_t - x)$, then using the elementary integrals

$$\lim_{\eta\to 0} \eta^{1/2} \int_0^\infty \frac{\sqrt{x}/\pi}{(x - {\rm i}\eta)^2}{\rm d}x = \frac{{\rm i}^{1/2}}{2}, \qquad \lim_{\eta\to 0} \eta^{3/2} \int_0^\infty \frac{\sqrt{x}/\pi}{(x - {\rm i}\eta)^3}{\rm d}x = \frac{{\rm i}^{3/2}}{8}$$

we obtain the precise divergence asymptotics of the derivatives $(m_{\mathrm{fc}}^t)'(z)$ and $(m_{\mathrm{fc}}^t)''(z)$ as $z = E_t + \mathrm{i}\eta \to E_t$ and conclude

$$\frac{2}{\gamma_t^3} = \lim_{z\to E_t} \frac{(ct)^3 (m_{\mathrm{fc}}^t)''(z)}{(1+ct(m_{\mathrm{fc}}^t)'(z))^3} = (ct)^3 \left\langle M_t''(\xi_t(E_t))\right\rangle, \quad \text{i.e,} \quad \gamma_t = \frac{(\left\langle M_t''(\xi_t(E_t))\right\rangle/2)^{-1/3}}{ct}.$$

We now use (9.4.4) at, say, $z = x := E - \sqrt{t/N}$. By Claim 1 we have $E_t - x \sim \sqrt{t/N}$ and thus

$$\begin{aligned}\gamma_t^{3/2} &= \frac{\operatorname{Im} m_{\mathrm{fc}}^t(x)}{\sqrt{E_t - x}} + \mathcal{O}\left((t/N)^{1/4}\right) = \frac{\operatorname{Im}\left\langle M(x)\right\rangle}{\sqrt{E_t - x}} + \mathcal{O}\left((t/N)^{1/4}\right) \\ &= \frac{\operatorname{Im}\left\langle M(x)\right\rangle}{\sqrt{E - x}} + \mathcal{O}\left((t/N)^{1/4}\right) = \gamma^{3/2} + \mathcal{O}\left((t/N)^{1/4}\right),\end{aligned}$$

where we used Claim 1 again in the third equality. This completes the proof of the claim.

Claim 3. $|\widetilde{E}_t - E_t| \prec 1/Nt$. Since M_t has a square root edge at some $\widehat{E}_t$, it follows from the first equality in (9.4.3b) that $\xi_t(E_t) - \widehat{E}_t \sim t^2$. Using rigidity in the form of Corollary 9.2.5 for the matrix $\widetilde{H}_t$ to estimate $\widetilde{G}_t'$ from below at a spectral parameter outside of the support, we have the bound

$$ct = |\left\langle \widetilde{G}_t'(\widetilde{\xi}_t(\widetilde{E}_t))\right\rangle|^{-1} \prec |\widetilde{\xi}_t(\widetilde{E}_t) - \widehat{E}_t|^{1/2}.$$

Consequently using the local law in the form of Lemma 9.5.1 it follows that

$$|\left\langle M_t'(\widetilde{\xi}_t(\widetilde{E}_t))\right\rangle| = 1/ct + \mathcal{O}_\prec(1/Nt^4) \sim 1/t,$$

whence $\widetilde{\xi}_t(\widetilde{E}_t) - \widehat{E}_t \sim t^2$ where we again used the square root singularity of $\langle M_t\rangle$ at $\widehat{E}_t$. We can conclude, starting from (9.4.3b), that

$$\begin{aligned}0 &= \langle M_t'(\xi_t(E_t))\rangle - \langle \widetilde{G}_t'(\widetilde{\xi}_t(\widetilde{E}_t))\rangle = \langle M_t'(\xi_t(E_t))\rangle - \langle M_t'(\widetilde{\xi}_t(\widetilde{E}_t))\rangle + \langle (M_t' - \widetilde{G}_t')(\widetilde{\xi}_t(\widetilde{E}_t))\rangle \\ &\sim |\xi_t(E_t) - \widetilde{\xi}_t(\widetilde{E}_t)|/t^3 + \mathcal{O}_\prec(1/Nt^4),\end{aligned}$$

where we used that $|\langle M_t''(\widehat{E}_t + rt^2)\rangle| \sim t^{-3}$ for $c < r < C$ and the improved local law $\langle G' - M'\rangle \prec 1/N\kappa^2$ at a distance $\kappa \sim t^2$ away from the spectrum, as stated in Lemma 9.5.1. We thus find that $|\xi_t(E_t) - \widetilde{\xi}_t(\widetilde{E}_t)| \prec 1/Nt$. It remains to relate this to an estimate on $|E_t - \widetilde{E}_t|$. We have

$$|E_t - \widetilde{E}_t| \lesssim |\xi_t(E_t) - \widetilde{\xi}_t(\widetilde{E}_t)| + t|m_{\mathrm{fc}}^t(E_t) - m_{\mathrm{fc}}^t(\widetilde{E}_t)| + t|(m_{\mathrm{fc}}^t - \widetilde{m}_{\mathrm{fc}}^t)(\widetilde{E}_t)|,$$

where we bounded the second term by $t|\langle M_t(\xi_t(E_t)) - M_t(\widetilde{\xi}_t(\widetilde{E}_t))\rangle| \prec 1/Nt$ using the bounds in $|\langle M_t'(\widehat{E}_t + rt^2)\rangle| \sim 1/t$ and the third term by $t|\langle (M_t - \widetilde{G}_t)(\widetilde{\xi}_t(\widetilde{E}_t))\rangle| \prec 1/Nt$ using the local law t^2 away from $\operatorname{supp}\varrho_t$. Thus we can conclude that $|E_t - \widetilde{E}_t| \prec 1/Nt$.

Claim 4. $|\gamma_t - \widetilde{\gamma}_t| \prec 1/Nt^3$. We first note that $\gamma_t \sim 1$ follows from $|\langle M_t''(\xi_t(E_t))\rangle| \sim t^{-3}$. Therefore it suffices to estimate

$$\begin{aligned} t^3|\langle M_t''(\xi_t(E_t)) - \widetilde{G}_t''(\widetilde{\xi}_t(\widetilde{E}_t))\rangle| &\leq t^3|\langle M_t''(\xi_t(E_t)) - M_t''(\widetilde{\xi}_t(\widetilde{E}_t))\rangle| \\ &\qquad + t^3|\langle M_t''(\widetilde{\xi}_t(\widetilde{E}_t)) - \widetilde{G}_t''(\widetilde{\xi}_t(\widetilde{E}_t))\rangle| \\ &\prec \frac{1}{Nt^3}, \end{aligned}$$

as follows from $\langle M_t'''(\widehat{E}_t + rt^2)\rangle \sim t^{-5}$ for $c < r < C$ and the local law from Lemma 9.5.1 at a distance of $\kappa \sim t^2$ away from the spectrum. Thus we have $|\gamma_t - \widetilde{\gamma}_t| \prec 1/Nt^3$.

We now check that $\widetilde{H}_t$ is η_*-regular in the sense of [103, Definition 2.1] for $\eta_* := N^{-2/3+\epsilon}$. It follows from the local law that $c\varrho_t(z) \prec \operatorname{Im}\langle \widetilde{G}_t(z)\rangle \prec C\varrho_t(z)$ for some constants c, C, whenever $\operatorname{Im} z \geq \eta_*$. Now (2.4)–(2.5) in [103] follow in high probability from the assumption that ϱ_t has a regular edge at E_t . Furthermore, the absence of eigenvalues in the interval $[E_t + \eta_*, E_t + c/2]$ with high probability follows directly from Corollary 9.2.3. Finally, $\|\widetilde{H}_t\| \leq N$ with high probability follows directly from $\|\widetilde{H}_t\| \leq (\operatorname{Tr}|\widetilde{H}_t|^2)^{1/2}$. We can thus conclude that with high probability, $\widetilde{H}_t$ is $\eta_* = N^{-2/3+\epsilon}$ regular for any positive $\epsilon > 0$.

We denote the eigenvalues of $H_t = \widetilde{H}_t + c\sqrt{t}U$ by $\lambda_1^t \geq \cdots \geq \lambda_N^t$. Then it follows from [103, Theorem 2.2] that for $N^{-\epsilon} \geq t \geq N^{-2/3+\epsilon}$ with high probability for test functions $F\colon \mathbb{R}^{k+1} \to \mathbb{R}$ with $\|F\|_\infty + \|\nabla F\|_\infty \lesssim 1$ there exists some $c > 0$ such that

$$\begin{aligned} \Big|\mathbb{E}\Big[F\Big(\widetilde{\gamma}_t N^{2/3}(\lambda_{i_0}^t - \widetilde{E}_t), \dots, \widetilde{\gamma}_t N^{2/3}(\lambda_{i_0+k}^t - \widetilde{E}_t)\Big)|\widetilde{H}_t\Big] \\ - \mathbb{E}\Big[F\Big(N^{2/3}(\mu_1 - 2), \dots, N^{2/3}(\mu_{k+1} - 2)\Big)\Big]\Big| \leq N^{-c}. \end{aligned} \tag{9.4.5}$$

By combining (9.4.5) with $|E - \widetilde{E}_t| \prec N^{-2/3-\epsilon}$, $|\gamma - \widetilde{\gamma}_t| \prec N^{-\epsilon}$ from Claims 1–4, we obtain

$$\begin{aligned} \Big|\mathbb{E}\Big[F\Big(\gamma N^{2/3}(\lambda_{i_0}^t - E), \dots, \gamma N^{2/3}(\lambda_{i_0+k}^t - E)\Big)\Big] \\ - \mathbb{E}\Big[F\Big(N^{2/3}(\mu_1 - 2), \dots, N^{2/3}(\mu_{k+1} - 2)\Big)\Big]\Big| \lesssim N^{-c} + N^{-\epsilon} \end{aligned} \tag{9.4.6}$$

for our choice of $t = N^{-1/3+\epsilon}$.

9.4.2. Green's Function Comparison. It remains to prove that the local correlation functions of H_t agree with those of H. We will prove that for any fixed $x_i \in \mathbb{R}$,

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3}(\lambda_{i_0+i}^t - E) \geq x_i,\, i = 0, \ldots, k\right)$$

is independent of t as long as, say, $t \leq N^{-1/3+\epsilon}$. We first note that the local law holds uniformly in t also for H_t. This follows easily from the fact that the assumptions stay uniformly satisfied along the flow because expectation and covariance are preserved while higher order cumulants also remain unchanged up to a multiplication with a t-dependent constant. For $l = N^{-2/3-\epsilon/3}$, $\eta = N^{-2/3-\epsilon}$, and smooth monotonous cut-off functions K_i with $K_i(x) = 0$ for $x \leq i-1$ and $K_i(x) = 1$ for $x \geq i$ we have

$$\begin{aligned} \mathbb{E}\prod_{i=0}^k K_{i_0+i}\left(\frac{\operatorname{Im}}{\pi}\int_{x_i N^{-2/3}+l}^{N^{-2/3+\epsilon}} \operatorname{Tr} G_t(x+E+\mathrm{i}\eta)\mathrm{d}x\right) - \mathcal{O}\left(N^{-\epsilon/9}\right) \\ \leq \mathbb{P}\left(N^{2/3}(\lambda_{i_0+i}^t - E) \geq x_i,\, i = 0, \ldots, k\right) \\ \leq \mathbb{E}\prod_{i=0}^k K_{i_0+i}\left(\frac{\operatorname{Im}}{\pi}\int_{x_i N^{-2/3}-l}^{N^{-2/3+\epsilon}} \operatorname{Tr} G_t(x+E+\mathrm{i}\eta)\mathrm{d}x\right) + \mathcal{O}\left(N^{-\epsilon/9}\right). \end{aligned} \tag{9.4.7}$$

We note that the strategy of expressing k-point correlation functions of edge-eigenvalues through a regularized expression involving the resolvent has already been used in [71, 97, 102, 106] for proving edge universality. The precise formula (9.4.7) has already been used, for example, in [97, Eq. (4.8)].

In order to compare the expectations in (9.4.7) at times $t = 0$ and $t = N^{-1/3+\epsilon}$, we claim that we have the bound

$$X_y := \operatorname{Im}\int_{yN^{-2/3}\pm l}^{N^{-2/3+\epsilon}} \operatorname{Tr} G_t(E+x+\mathrm{i}\eta)\mathrm{d}x, \qquad \left|\mathbb{E}\frac{\mathrm{d}X_y}{\mathrm{d}t}\right| \lesssim N^{1/6+3\epsilon}. \tag{9.4.8}$$

Proof of (9.4.8). We consider general functions f of the random matrix $f(H_t)$ and find from Itô's lemma that

$$\mathbb{E}\frac{\mathrm{d}f(H)}{\mathrm{d}t} = \mathbb{E}\left[-\frac{1}{2}\sum_\alpha w_\alpha(\partial_\alpha f)(H) + \frac{1}{2}\sum_{\alpha,\beta}\kappa(\alpha,\beta)(\partial_\alpha\partial_\beta f)(H)\right].$$

For the second term we use the general neighbourhood cumulant expansion from [56, Proposition 3.5] to obtain

$$\begin{aligned}\mathbb{E}\frac{\mathrm{d}f(H)}{\mathrm{d}t} =& \mathbb{E}\Big[-\frac{1}{2}\sum_{2\le m<R}\sum_{\alpha}\sum_{\beta_1,\dots,\beta_m\in\mathcal{N}}\frac{\kappa(\alpha,\boldsymbol{\beta})}{m!}(\partial_\alpha\partial_{\boldsymbol{\beta}}f)(H)-\frac{1}{2}\Omega((\partial_\alpha f)(H),\alpha,\mathcal{N})\\ &-\frac{1}{2}\sum_{m<R}\sum_{\alpha}\sum_{\beta_1,\dots,\beta_m\in\mathcal{N}}\frac{K(w_\alpha;w_{\boldsymbol{\beta}})-\kappa(\alpha,\boldsymbol{\beta})}{m!}(\partial_\alpha\partial_{\boldsymbol{\beta}}f)(H)\Big|_{W_{\mathcal{N}}=0}\\ &+\frac{1}{2}\sum_{\alpha}\sum_{\beta\in\mathcal{N}^c}\kappa(\alpha,\beta)(\partial_\alpha\partial_\beta f)(H)\Big].\end{aligned} \tag{9.4.9}$$

Eq. (9.4.9) requires some explanations. The neighbourhood $\mathcal{N}(\alpha)\ni\alpha$ is a neighbourhood of α of size $|\mathcal{N}|\le N^{1/2-\mu}$ for some constant $\mu>0$ which is guaranteed to exist by Assumptions (C), (D) in [56], and thereby by Assumption (CD) in the present paper. The random variable $K(w_\alpha;w_{\boldsymbol{\beta}})$, as defined in [56, Section 3.1], is called the *pre-cumulant* which is justified by the fact that $\mathbb{E}K=\kappa$. In (9.4.9), Ω is an irrelevant error term, defined in [56, Proposition 3.5]. The central assumption on the correlation decay is that there exist some nested neighbourhoods $\mathcal{N}_1\subset\cdots\subset\mathcal{N}_R=\mathcal{N}$ such that the covariance of f supported in $\mathcal{N}_k$ and g supported in $\mathcal{N}_{k+1}^c$ is of size N^{-3}. The pre-cumulants K have the property that $\mathbf{Cov}\,(K,f)\lesssim N^{-3}$ whenever f is supported outside $\mathcal{N}$ and $w_\alpha,w_{\boldsymbol{\beta}}$ split into two groups contained in $\mathcal{N}_k$ and $\mathcal{N}_{k+1}^c$. Due to the pigeon-hole principle such a splitting always occurs. The large integer R is chosen in such a way that $R\gg 1/\mu$ in which case the second term in (9.4.9) becomes negligible small. For more details the reader is referred to [56].

We now apply (9.4.9) to X_t. We consider the first term in (9.4.9) as the leading order term and will first work out the desired bound for

$$\mathbb{E}\mathrm{Im}\int_{xN^{-2/3}\pm l}^{N^{-2/3+\epsilon}}\left[\sum_{2\le m<R}\sum_{\alpha_1}\sum_{\alpha_2,\dots,\alpha_{m+1}\in\mathcal{N}}\frac{(m+1)\kappa(\boldsymbol{\alpha})}{2}\operatorname{Tr}G_t\Delta^{\alpha_1}G_t\Delta^{\alpha_2}\dots G_t\Delta^{\alpha_{m+1}}G_t\right]\mathrm{d}x, \tag{9.4.10}$$

where $G_t=G_t(x+E+\mathrm{i}\eta)$. For $m\ge 4$, we can trivially estimate the corresponding term from (9.4.10) by

$$N^{-2/3+\epsilon}N^{2-(m+1)/2}N\sup_{|x|\le N^{-2/3+\epsilon}}\|G_t\|_{m+2}^{m+2}\lesssim N^{-1/6+\epsilon}$$

where we used the local law in the last step to obtain $\|G_t\|_p \lesssim \|M_t\| \lesssim 1$ and the summability of cumulants in the form $\sum_{\alpha_2,\dots,\alpha_k} |\kappa(\alpha_1,\dots,\alpha_k)| \lesssim 1$. For $m = 2$ we write out and use the local law in the form $\|\operatorname{Im} G_t\|_p \lesssim \varrho_t + \|G_t - M_t\|_p \lesssim N^{-1/3+\epsilon}$ to obtain

$$\sum_{a_i,b_i,c} \kappa(a_1b_1,a_2b_2,a_3b_3)\mathbb{E}\,|(G_t)_{ca_1}(G_t)_{b_1a_2}(G_t)_{b_2a_3}(G_t)_{b_3c}|$$

$$\le \sum_{a_i,b_i} \kappa(a_1b_1,a_2b_2,a_3b_3)\mathbb{E}\frac{\sqrt{(\operatorname{Im} G_t)_{a_1a_1}}\sqrt{(\operatorname{Im} G_t)_{b_3b_3}}}{\eta}\,|(G_t)_{b_1a_2}(G_t)_{b_2a_3}| \lesssim N^{2-3/2+1/3+2\epsilon}$$

and consequently can bound the corresponding term by $N^{1/6+3\epsilon}$. The case $m = 3$ is very similar and we obtain a bound of $N^{-1/3+3\epsilon}$.

We now consider the neighbourhood induced error terms in (9.4.9), i.e., the second, third and fourth term. The treatment of these error terms is rather easy and closely resembles the argument in [56, Proof of Corollary 2.6]. For the convenience of the reader we briefly sketch the bounds for all remaining terms but leave the details to the reader. For the last term we use $|\kappa(\alpha,\beta)| \lesssim N^{-4}$ for $\beta \in \mathcal{N}^c$ to obtain

$$\mathbb{E}\sum_{\alpha}\sum_{\beta\in\mathcal{N}^c} \kappa(\alpha,\beta)\left|\operatorname{Tr} G_t\Delta^\alpha G_t\Delta^\beta G_t\right| \lesssim N^{-4}\sum_{abcde}\mathbb{E}\,|(G_t)_{ab}(G_t)_{cd}(G_t)_{ea}|$$

$$\lesssim N\left(\frac{\varrho_t}{N\eta}\right)^{3/2} \lesssim N^{3\epsilon}$$

for the integrand and can conclude that the term is bounded by $N^{-2/3+4\epsilon}$ due to the integration length. For the third term in (9.4.9) we bound the derivative trivially by N (coming from the trace), while the cumulant is of size $N^{-(R+1)/2}$, which compensates for the summation of size $N^2|\mathcal{N}|^R \le N^{2+R/2-\mu R}$ and we can choose $R = 2/\mu$ large to obtain a bound of $N^{-1/6+\epsilon}$ for the term after integration. Finally, for the fourth term in (9.4.9) we have a naive bound of size $N^{-2/3+5/2+\epsilon}$, which we can improve to $N^{-7/6+\epsilon}$ using the pigeon-hole principle and the covariance bound (as in [56, Eq. (27)]). □

For the case of general k and smooth functions K_j's in (9.4.7) we can easily generalise (9.4.8) to

$$\left|\mathbb{E}g(X_{x_0},\dots,X_{x_k})\frac{\mathrm{d}X_{x_j}}{\mathrm{d}t}\right| \lesssim N^{1/6+3\epsilon}$$

for any $0 \leq j \leq k$ and any smooth function g. Then by a routine power counting argument and Taylor expanding the K_j's it follows that for any $t \lesssim N^{-1/3+\epsilon}$ we have

$$\Big|\mathbb{E}\prod_{i=0}^{k} K_{i_0+i}\left(\frac{\mathrm{Im}}{\pi}\int_{x_i N^{-2/3}\pm l}^{N^{-2/3+\epsilon}} \mathrm{Tr}\, G_t(x+E+\mathrm{i}\eta)\mathrm{d}x\right) \\ - \mathbb{E}\prod_{i=0}^{k} K_{i_0+i}\left(\frac{\mathrm{Im}}{\pi}\int_{x_i N^{-2/3}\pm l}^{N^{-2/3+\epsilon}} \mathrm{Tr}\, G_0(x+E+\mathrm{i}\eta)\mathrm{d}x\right)\Big| \lesssim \frac{1}{N^{1/6-4\epsilon}}.$$

Together with (9.4.7) we obtain for any k, x_i

$$\mathbb{P}\left(N^{2/3}(\lambda_{i_0+i}^t - E) \geq x_i,\, i\in[k]\right) = \mathbb{P}\left(N^{2/3}(\lambda_{i_0+i}^0 - E) \geq x_i,\, i\in[k]\right) + \mathcal{O}\left(N^{-\epsilon/9}\right). \tag{9.4.11}$$

Proof of Theorem 9.2.7. The theorem follows directly from (9.4.6) and (9.4.11). □

9.5. Auxiliary results

Proof of Lemma 9.3.4. From (70a)–(70b) in [56] we have[5]

$$\|M\mathcal{S}[R]R\|_* \lesssim N^{1/2K}\|R\|_*^2, \qquad \|MR\|_* \lesssim N^{1/2K}\|R\|_* \tag{9.5.1a}$$

and furthermore by a three term geometric expansion also

$$\left\|\mathcal{B}^{-1}\mathcal{Q}\right\|_{*\to*} \leq (1+\|\mathcal{Q}\|_{*\to*})\Big(1+\|\mathcal{C}_M\mathcal{S}\|_{*\to*} + \|\mathcal{C}_M\mathcal{S}\|_{*\to\mathrm{hs}}\left\|\mathcal{B}^{-1}\mathcal{Q}\right\|_{\mathrm{hs}\to\mathrm{hs}}\|\mathcal{C}_M\mathcal{S}\|_{\mathrm{hs}\to*}\Big). \tag{9.5.1b}$$

Since

$$\|\mathcal{P}[R]\|_* = |\langle P,R\rangle|\,\|B\|_* \leq \frac{\|B\|}{N}\sum_a |R_{P_{a\cdot}^* a}| \leq \frac{\|B\|\,\|R\|_*}{N}\sum_a \|P_{a\cdot}^*\| \leq \|P\|\,\|B\|\,\|R\|_*$$

it follows that $\|\mathcal{P}\|_{*\to*} \lesssim 1$ and therefore also $\|\mathcal{Q}\|_{*\to*} \lesssim 1$. Now, since $\|R\|_{\max} \leq \|R\|_* \leq \|R\|$ and according to (73) in [56] also $\max\{\|\mathcal{S}\|_{\max\to\|\cdot\|}, \|\mathcal{S}\|_{\mathrm{hs}\to\|\cdot\|}\} \lesssim 1$, the lemma follows together with $\|\mathcal{B}^{-1}\mathcal{Q}\|_{\mathrm{hs}\to\mathrm{hs}} \lesssim 1$ from Proposition 9.3.1(v). □

[5] C.f. Remark 9.3.2 for the applicability of these bounds in the present setup.

Lemma 9.5.1. *Fix any $\epsilon, \delta > 0$ and an integer $k \geq 0$. Under the assumptions of Theorem 9.2.2, for the k-th derivatives of M and G we have the bound*

$$\Big|\langle G^{(k)}(z) - M^{(k)}(z)\rangle\Big| \prec \frac{1}{N\kappa^{k+1}}. \tag{9.5.2}$$

uniformly in $z \in \mathbb{D}$ with $\kappa = \operatorname{dist}(z, \operatorname{supp}\varrho) \geq N^{-2/3+\epsilon}$, $d_{\text{cusp}} \geq \delta$, $d_{\text{M}} \geq \delta$.

Proof. We will fix $z = x + \mathrm{i}\eta$ throughout the proof. Let $\chi\colon \mathbb{R} \to \mathbb{R}$ be a smooth cut-off function such that $\chi(x') = 1$ for $\kappa' = \operatorname{dist}(x', \operatorname{supp}\varrho) \leq \kappa/3$ and $\chi(x') = 0$ for $\kappa' \geq 2\kappa/3$ and let $\widetilde{\chi}$ be a cut-off function such that $\widetilde{\chi}(\eta') = 1$ for $\eta' \leq 1$ and $\widetilde{\chi}(\eta') = 0$ for $\eta' \geq 2$. We also assume that the cut-off functions have bounded derivatives in the sense $\|\chi'\|_\infty \lesssim 1/\kappa$, $\|\chi''\|_\infty \lesssim 1/\kappa^2$ and $\|\widetilde{\chi}'\|_\infty \lesssim 1$. We now define $f(x') := (x'-z)^{-k}\chi(x')$ and the almost analytic extension

$$f^{\mathbb{C}}(z') = f^{\mathbb{C}}(x' + \mathrm{i}\eta') := \widetilde{\chi}(\eta')\Big[f(x') + \mathrm{i}\eta' f'(x')\Big],$$
$$\partial_{\overline{z}} f^{\mathbb{C}}(z') = \frac{\mathrm{i}\eta'}{2}\widetilde{\chi}(\eta') f''(x') + \frac{\mathrm{i}}{2}\widetilde{\chi}'(\eta')\Big[f(x') + \mathrm{i}\eta' f'(x')\Big].$$

It follows from the Cauchy Theorem and the absence of eigenvalues outside $\{\,\chi = 1\,\}$ in the sense of Corollary 9.2.3 that with high probability

$$\langle G^{(k)}(z) - M^{(k)}(z)\rangle = \frac{2}{\pi}\mathrm{Re}\int_{\mathbb{R}}\int_{\mathbb{R}_+} \partial_{\overline{z}} f^{\mathbb{C}}(z')\,\langle G(z') - M(z')\rangle\,\mathrm{d}\eta'\mathrm{d}x'.$$

Due to the fact that $\widetilde{\chi}' = 0$ for $\eta' \leq 1$ the second term in $\partial_{\overline{z}} f^{\mathbb{C}}$ only gives a contribution of $1/N\kappa^{k+1}$ even by the local law and the $\|\cdot\|_\infty$ bound for $\partial_{\overline{z}} f^{\mathbb{C}}$ and we now concentrate on the first term. First, we exclude the integration regime $\eta' \lesssim N^{-1+\gamma}$ in which we cannot use the local law but only the trivial bound $\langle G - M\rangle \lesssim 1/\eta'$. For the contribution of this regime to (9.5.2) we thus have to estimate

$$\begin{aligned} N^{-1+\gamma}\int_{\mathbb{R}} |f''(x')|\,\mathrm{d}x' &\lesssim \frac{1}{N}\int_{|x-x'|\geq 2\kappa/3}\left[\frac{1}{\kappa^2\,|x-x'|^k} + \frac{1}{\kappa\,|x-x'|^{k+1}} + \frac{1}{|x-x'|^{k+2}}\right]\mathrm{d}x' \\ &\lesssim \frac{N^\gamma}{N\kappa^{k+1}} \end{aligned}$$

and we have shown that

$$\begin{aligned}&\Big|\langle G^{(k)}(z) - M^{(k)}(z)\rangle\Big|\\&\prec \frac{N^\gamma}{N\kappa^{k+1}} + \int_{\mathbb{R}}\int_{N^{-1+\gamma}}^{2} \eta'\Big[\frac{\chi(x')}{|x'-z|^{k+2}} + \frac{\chi'(x')}{|x'-z|^{k+1}} + \frac{\chi''(x')}{|x'-z|^{k}}\Big]\,|\langle G(z') - M(z')\rangle|\,\mathrm{d}\eta'\mathrm{d}x'.\end{aligned}$$

We now use the local law of the form $|\langle G - M\rangle| \prec 1/N(\kappa + \eta')$ and that in the second and third term the integration regime is only of order κ to obtain the final bound of $N^\gamma/N\kappa^{k+1}$ for any $\gamma > 0$. □

Bibliography

[1] A. Adhikari and Z. Che, *The edge universality of correlated matrices*, preprint (2017), `arXiv:1712.04889`.

[2] Y. Ahmadian, F. Fumarola, and K. D. Miller, *Properties of networks with partially structured and partially random connectivity*, Phys. Rev. E **91** (2015), 012820.

[3] O. H. Ajanki, L. Erdős, and T. Krüger, *Local semicircle law with imprimitive variance matrix*, Electron. Commun. Probab. **19** (2014), Paper no. 33, 9 pp.

[4] O. H. Ajanki, L. Erdős, and T. Krüger, *Quadratic vector equations on complex upper half-plane*, to appear in Mem. Amer. Math. Soc. (2015), `arXiv:1506.05095`.

[5] O. H. Ajanki, L. Erdős, and T. Krüger, *Singularities of solutions to quadratic vector equations on the complex upper half-plane*, Comm. Pure Appl. Math. **70** (2017), no. 9, 1672–1705.

[6] O. H. Ajanki, L. Erdős, and T. Krüger, *Stability of the matrix Dyson equation and random matrices with correlations*, Probab. Theory Related Fields (2018).

[7] O. H. Ajanki, L. Erdős, and T. Krüger, *Universality for general Wigner-type matrices*, Probab. Theory Related Fields **169** (2017), no. 3-4, 667–727.

[8] A. B. Aleksandrov and V. V. Peller, *Operator Lipschitz functions*, Russian Mathematical Surveys **71** (2016), no. 4, 605.

[9] J. Aljadeff, D. Renfrew, and M. Stern, *Eigenvalues of block structured asymmetric random matrices*, J. Math. Phys. **56** (2015), no. 10, 103502, 14.

[10] J. Aljadeff, M. Stern, and T. Sharpee, *Transition to chaos in random networks with cell-type-specific connectivity*, Phys. Rev. Lett. **114** (2015), 088101.

[11] J. Alt, *Singularities of the density of states of random Gram matrices*, Electron. Commun. Probab. **22** (2017), Paper no. 63, 13 pp.

[12] J. Alt, *The local semicircle law for random matrices with a fourfold symmetry*, J. Math. Phys. **56** (2015), no. 10, 103301, 20.

[13] J. Alt, L. Erdős, and T. Krüger, *Local inhomogeneous circular law*, Ann. Appl. Probab. **28** (2018), no. 1, 148–203.

[14] J. Alt, L. Erdős, and T. Krüger, *Local law for random Gram matrices*, Electron. J. Probab. **22** (2017), Paper no. 25, 41 pp.

[15] J. Alt, L. Erdős, and T. Krüger, *The Dyson equation with linear self-energy: spectral bands, edges and cusps*, preprint (2018), `arXiv:1804.07752v2`.

[16] J. Alt, L. Erdős, T. Krüger, and Yu. Nemish, *Location of the spectrum of Kronecker random matrices*, to appear in Ann. Inst. Henri Poincaré Probab. Stat. (2017), `arXiv:1706.08343`.

[17] J. Alt, L. Erdős, T. Krüger, and D. Schröder, *Correlated random matrices: band rigidity and edge universality*, preprint (2018), `arXiv:1804.07744v2`.

[18] G. W. Anderson, *Convergence of the largest singular value of a polynomial in independent Wigner matrices*, Ann. Probab. **41** (2013), no. 3B, 2103–2181.

[19] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Vol. 118, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 2010), pp. xiv+492.

[20] G. W. Anderson and O. Zeitouni, *A CLT for a band matrix model*, Probab. Theory Related Fields **134** (2006), no. 2, 283–338.

[21] A. Auffinger, G. Ben Arous, and S. Péché, *Poisson convergence for the largest eigenvalues of heavy tailed random matrices*, Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), no. 3, 589–610.

[22] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), no. 1, 494–529.

[23] Z. D. Bai and J. W. Silverstein, *Exact separation of eigenvalues of large-dimensional sample covariance matrices*, Ann. Probab. **27** (1999), no. 3, 1536–1555.

[24] Z. D. Bai and Y. Q. Yin, *Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang*, Probab. Theory Related Fields **73** (1986), no. 4, 555–569.

[25] 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** (1988), no. 4, 1729–1741.

[26] Z. D. Bai and J. W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices*, Random Matrices Theory Appl. **1** (2012), no. 1, 1150004, 44.

[27] Z. Bao, L. Erdős, and K. Schnelli, *Local single ring theorem on optimal scale*, preprint (2016), `arXiv:1612.05920`.

[28] Z. Bao, G. Pan, and W. Zhou, *Tracy-Widom law for the extreme eigenvalues of sample correlation matrices*, Electron. J. Probab. **17** (2012), Paper no. 88, 32 pp.

[29] R. B. Bapat and T. E. S. Raghavan, *Nonnegative matrices and applications*, Vol. 64, Encyclopedia of Mathematics and its Applications (Cambridge University Press, Cambridge, 1997), pp. xiv+336.

[30] F. Bekerman, *Transport maps for β-matrix models in the multi-cut regime*, Random Matrices Theory Appl. **7** (2018), no. 1, 1750013, 36.

[31] S. T. Belinschi and M. Capitaine, *Spectral properties of polynomials in independent Wigner and deterministic matrices*, J. Funct. Anal. **273** (2017), no. 12, 3901–3963.

[32] J. Bellissard, J. Magnen, and V. Rivasseau, *Supersymmetric analysis of a simplified two-dimensional Anderson model at small disorder*, Markov Process. Related Fields **9** (2003), no. 2, Inhomogeneous random systems (Cergy-Pontoise, 2002), 261–278.

[33] F. Benaych-Georges and R. Couillet, *Spectral analysis of the Gram matrix of mixture models*, ESAIM Probab. Stat. **20** (2016), 217–237.

[34] F. A. Berezin, *Some remarks on the Wigner distribution*, Teoret. Mat. Fiz. **17** (1973), 305–318.

[35] R. Bhatia, *Matrix analysis*, Vol. 169, Graduate Texts in Mathematics (Springer-Verlag, New York, 1997), pp. xii+347.

[36] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Isotropic local laws for sample covariance and generalized Wigner matrices*, Electron. J. Probab. **19** (2014), Paper no. 33, 53 pp.

[37] G. Bonnet, F. David, and B. Eynard, *Breakdown of universality in multi-cut matrix models*, J. Phys. A **33** (2000), no. 38, 6739–6768.

[38] C. Bordenave and M. Capitaine, *Outlier eigenvalues for deformed i.i.d. random matrices*, Comm. Pure Appl. Math. **69** (2016), no. 11, 2131–2194.

[39] C. Bordenave, P. Caputo, D. Chafaï, and K. Tikhomirov, *On the spectral radius of a random matrix*, preprint (2016), `arXiv:1607.05484`.

[40] C. Bordenave and D. Chafai, *Around the circular law*, Probab. Surv. **9** (2012), 1–89.

[41] G. Borot and A. Guionnet, *Asymptotic expansion of beta matrix models in the multi-cut regime*, preprint (2013), `arXiv:1303.1045`.

[42] P. Bourgade, *Random band matrices*, preprint (2018), `arXiv:1807.03031`.

[43] P. Bourgade, L. Erdős, and H.-T. Yau, *Edge universality of beta ensembles*, Comm. Math. Phys. **332** (2014), no. 1, 261–353.

[44] P. Bourgade, H.-T. Yau, and J. Yin, *Local circular law for random matrices*, Probab. Theory Related Fields **159** (2014), no. 3-4, 545–595.

[45] P. Bourgade, H.-T. Yau, and J. Yin, *The local circular law II: the edge case*, Probab. Theory Related Fields **159** (2014), no. 3-4, 619–660.

[46] C. Cacciapuoti, A. Maltsev, and B. Schlein, *Local Marchenko-Pastur law at the hard edge of sample covariance matrices*, J. Math. Phys. **54** (2013), no. 4, 043302, 13.

[47] M. Capitaine and C. Donati-Martin, *Strong asymptotic freeness for Wigner and Wishart matrices*, Indiana Univ. Math. J. **56** (2007), no. 2, 767–803.

[48] M. Capitaine and S. Péché, *Fluctuations at the edges of the spectrum of the full rank deformed GUE*, Probab. Theory Related Fields **165** (2016), no. 1-2, 117–161.

[49] Z. Che, *Universality of random matrices with correlated entries*, Electron. J. Probab. **22** (2017), Paper no. 30, 38 pp.

[50] N. Cook, *Lower bounds for the smallest singular value of structured random matrices*, preprint (2016), `arXiv:1608.07347`.

[51] N. Cook, W. Hachem, J. Najim, and D. Renfrew, *Limiting spectral distribution for non-Hermitian random matrices with a variance profile*, preprint (2016), `arXiv:1612.04428`.

[52] R. Couillet and M. Debbah, *Random matrix methods for wireless communications* (Cambridge University Press, Cambridge, 2011), pp. xxii+539.

[53] R. Couillet and W. Hachem, *Analysis of the limiting spectral measure of large random matrices of the separable covariance type*, Random Matrices Theory Appl. **3** (2014), no. 4, 1450016, 23.

[54] M. Disertori and V. Rivasseau, *Random matrices and the Anderson model*, in *Random Schrödinger operators*, Vol. 25, Panor. Synthèses (Soc. Math. France, Paris, 2008), pp. 161–213.

[55] R. B. Dozier and J. W. Silverstein, *On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices*, J. Multivariate Anal. **98** (2007), no. 4, 678–694.

[56] L. Erdős, T. Krüger, and D. Schröder, *Random matrices with slow correlation decay*, preprint (2017), `arXiv:1705.10661`.

[57] L. Erdős, *Universality for random matrices and log-gases*, Curr. Dev. Math. (2012), 59–132.

[58] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi graphs I: Local semicircle law*, Ann. Probab. **41** (2013), no. 3B, 2279–2375.

[59] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. **314** (2012), no. 3, 587–640.

[60] 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** (2013), Paper no. 59, 58 pp.

[61] L. Erdős, T. Krüger, and Yu. Nemish, *Local laws for polynomials of Wigner matrices*, preprint (2018), `arXiv:1804.11340`.

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

[63] L. Erdős, B. Schlein, and H.-T. Yau, *Local semicircle law and complete delocalization for Wigner random matrices*, Comm. Math. Phys. **287** (2009), no. 2, 641–655.

[64] L. Erdős, B. Schlein, and H.-T. Yau, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119.

[65] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics for random matrices*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 1, 1–46.

[66] L. Erdős and K. Schnelli, *Universality for random matrix flows with time-dependent density*, Ann. Inst. Henri Poincaré Probab. Stat. **53** (2017), no. 4, 1606–1656.

[67] L. Erdős and H.-T. Yau, *A dynamical approach to random matrix theory*, Vol. 28, Courant Lecture Notes in Mathematics (Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017), pp. ix+226.

[68] L. Erdős and H.-T. Yau, *Gap universality of generalized Wigner and β-ensembles*, J. Eur. Math. Soc. (JEMS) **17** (2015), no. 8, 1927–2036.

[69] L. Erdős and H.-T. Yau, *Universality of local spectral statistics of random matrices*, Bull. Amer. Math. Soc. (N.S.) **49** (2012), no. 3, 377–414.

[70] L. Erdős, H.-T. Yau, and J. Yin, *Bulk universality for generalized Wigner matrices*, Probab. Theory Related Fields **154** (2012), no. 1-2, 341–407.

[71] L. Erdős, H.-T. Yau, and J. Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. **229** (2012), no. 3, 1435–1515.

[72] L. Erdős, H.-T. Yau, and J. Yin, *Universality for generalized Wigner matrices with Bernoulli distribution*, J. Comb. **2** (2011), no. 1, 15–81.

[73] O. N. Feldheim and S. Sodin, *A universality result for the smallest eigenvalues of certain sample covariance matrices*, Geom. Funct. Anal. **20** (2010), no. 1, 88–123.

[74] D. Féral and S. Péché,
The largest eigenvalue of rank one deformation of large Wigner matrices,
Comm. Math. Phys. **272** (2007), no. 1, 185–228.

[75] A. Fey, R. van der Hofstad, and M. J. Klok, *Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems*,
Adv. in Appl. Probab. **40** (2008), no. 4, 1048–1071.

[76] G. J. Foschini, *Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas*,
Bell Labs Technical Journal **1** (1996), no. 2, 41–59.

[77] G. Foschini and M. Gans, *On limits of wireless communications in a fading environment when using multiple antennas*,
Wireless Per. Commun. **6** (1998), no. 3, 311–335.

[78] Y. V. Fyodorov and A. D. Mirlin,
Scaling properties of localization in random band matrices: a σ-model approach,
Phys. Rev. Lett. **67** (1991), no. 18, 2405–2409.

[79] A. Garg, L. Gurvits, R. Oliveira, and A. Wigderson, *A deterministic polynomial time algorithm for non-commutative rational identity testing*,
in *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016* (IEEE Computer Soc., Los Alamitos, CA, 2016), pp. 109–117.

[80] S. Geman, *The spectral radius of large random matrices*,
Ann. Probab. **14** (1986), no. 4, 1318–1328.

[81] V. L. Girko, *The circular law*,
Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 669–679.

[82] V. L. Girko, *Theory of stochastic canonical equations. Vol. I*, Vol. 535,
Mathematics and its Applications (Kluwer Academic Publishers, Dordrecht, 2001), pp. xxiv+497.

[83] F. Götze and A. Tikhomirov, *The circular law for random matrices*,
Ann. Probab. **38** (2010), no. 4, 1444–1491.

[84] A. Guionnet, *Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices*,
Ann. Inst. H. Poincaré Probab. Statist. **38** (2002), no. 3, 341–384.

[85] A. Guionnet, M. Krishnapur, and O. Zeitouni, *The single ring theorem*,
Ann. of Math. (2) **174** (2011), no. 2, 1189–1217.

[86] 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** (2006), no. 1, 1–83.

[87] U. Haagerup and S. Thorbjørnsen,
A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(\mathbb{F}_2))$ *is not a group*,
Ann. of Math. (2) **162** (2005), no. 2, 711–775.

[88] W. Hachem, P. Loubaton, and J. Najim, *The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile*,
Ann. Inst. H. Poincaré Probab. Statist. **42** (2006), no. 6, 649–670.

[89] W. Hachem, M. Kharouf, J. Najim, and J. W. Silverstein,
A CLT for information-theoretic statistics of non-centered Gram random matrices,
Random Matrices Theory Appl. **1** (2012), no. 2, 1150010, 50.

[90] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, *A new approach for mutual information analysis of large dimensional multi-antenna channels*,
IEEE Trans. Inform. Theory **54** (2008), no. 9, 3987–4004.

[91] W. Hachem, P. Loubaton, and J. Najim, *A CLT for information-theoretic statistics of Gram random matrices with a given variance profile*,
Ann. Appl. Probab. **18** (2008), no. 6, 2071–2130.

[92] W. Hachem, P. Loubaton, and J. Najim,
Deterministic equivalents for certain functionals of large random matrices,
Ann. Appl. Probab. **17** (2007), no. 3, 875–930.

[93] H. M. Hastings, F. Juhasz, and M. A. Schreiber,
Stability of structured random matrices,
Proceedings: Biological Sciences **249** (1992), no. 1326, 223–225.

[94] Y. He, A. Knowles, and R. Rosenthal,
Isotropic self-consistent equations for mean-field random matrices,
Probab. Theory Related Fields (2017), 1–47.

[95] J. W. Helton, T. Mai, and R. Speicher,
Applications of realizations (aka linearizations) to free probability,
J. Funct. Anal. **274** (2018), no. 1, 1–79.

[96] 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), no. 22, Art. ID rnm086, 15.

[97] J. Huang, B. Landon, and H.-T. Yau, *Transition from Tracy-Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős-Rényi graphs*, preprint (2017), arXiv:1712.03936.

[98] K. Johansson, *From Gumbel to Tracy-Widom*, Probab. Theory Related Fields **138** (2007), no. 1-2, 75–112.

[99] A. M. Khorunzhy and L. A. Pastur, *On the eigenvalue distribution of the deformed Wigner ensemble of random matrices*, in *Spectral operator theory and related topics*, Vol. 19, Adv. Soviet Math. (Amer. Math. Soc., Providence, RI, 1994), pp. 97–127.

[100] B. Khoruzhenko, *Large-N eigenvalue distribution of randomly perturbed asymmetric matrices*, J. Phys. A **29** (1996), no. 7, L165–L169.

[101] A. Knowles and J. Yin, *Anisotropic local laws for random matrices*, Probab. Theory Related Fields **169** (2017), no. 1-2, 257–352.

[102] A. Knowles and J. Yin, *Eigenvector distribution of Wigner matrices*, Probab. Theory Related Fields **155** (2013), no. 3-4, 543–582.

[103] B. Landon and H.-T. Yau, *Edge statistics of Dyson Brownian motion*, preprint (2017), arXiv:1712.03881.

[104] B. Landon, P. Sosoe, and H.-T. Yau, *Fixed energy universality for Dyson Brownian motion*, preprint (2016), arXiv:1609.09011.

[105] B. Landon and H.-T. Yau, *Convergence of local statistics of Dyson Brownian motion*, Comm. Math. Phys. **355** (2017), no. 3, 949–1000.

[106] J. O. Lee and K. Schnelli, *Edge universality for deformed Wigner matrices*, Rev. Math. Phys. **27** (2015), no. 8, 1550018, 94.

[107] J. O. Lee and K. Schnelli, *Local law and Tracy-Widom limit for sparse random matrices*, Probab. Theory Related Fields (2017).

[108] J. O. Lee and K. Schnelli, *Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population*, Ann. Appl. Probab. **26** (2016), no. 6, 3786–3839.

[109] J. O. Lee and J. Yin, *A necessary and sufficient condition for edge universality of Wigner matrices*, Duke Math. J. **163** (2014), no. 1, 117–173.

[110] 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** (2017), 1080–1107.

[111] T. Mai, R. Speicher, and S. Yin, *The free field: zero divisors, Atiyah property and realizations via unbounded operators*, preprint (2018), `arXiv:1805.04150`.

[112] V. A. Marchenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Mat. Sb. **1** (1967), no. 4, 457–483.

[113] R. M. May, *Will a large complex system be stable?*, Nature **238** (1972), no. 5364, 413–414.

[114] M. L. Mehta, *Random matrices and the statistical theory of energy levels* (Academic Press, New York-London, 1967), pp. x+259.

[115] J. A. Mingo and R. Speicher, *Free probability and random matrices*, Vol. 35, Fields Institute Monographs (Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017), pp. xiv+336.

[116] G. Pan and W. Zhou, *Circular law, extreme singular values and potential theory*, J. Multivariate Anal. **101** (2010), no. 3, 645–656.

[117] L. Pastur, *Limiting laws of linear eigenvalue statistics for Hermitian matrix models*, J. Math. Phys. **47** (2006), no. 10, 103303, 22.

[118] L. Pastur and M. Shcherbina, *On the edge universality of the local eigenvalue statistics of matrix models*, Mat. Fiz. Anal. Geom. **10** (2003), no. 3, 335–365.

[119] L. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, Vol. 171, Mathematical Surveys and Monographs (American Mathematical Society, Providence, RI, 2011), pp. xiv+632.

[120] V. Paulsen, *Completely bounded maps and operator algebras*, Vol. 78, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 2002), pp. xii+300.

[121] R. Peled, J. Schenker, M. Shamis, and S. Sodin, *On the Wegner orbital model*, (2016), arXiv:1608.02922.

[122] N. S. Pillai and J. Yin, *Universality of covariance matrices*, Ann. Appl. Probab. **24** (2014), no. 3, 935–1001.

[123] K. Rajan and L. F. Abbott, *Eigenvalue spectra of random matrices for neural networks*, Phys. Rev. Lett. **97** (2006), 188104.

[124] M. Rudelson and R. Vershynin, *Delocalization of eigenvectors of random matrices with independent entries*, Duke Math. J. **164** (2015), no. 13, 2507–2538.

[125] A. Ruzmaikina, *Universality of the edge distribution of eigenvalues of Wigner random matrices with polynomially decaying distributions of entries*, Comm. Math. Phys. **261** (2006), no. 2, 277–296.

[126] J. Schenker, *Eigenvector localization for random band matrices with power law band width*, Comm. Math. Phys. **290** (2009), no. 3, 1065–1097.

[127] M. Shcherbina, *Fluctuations of linear eigenvalue statistics of β matrix models in the multi-cut regime*, J. Stat. Phys. **151** (2013), no. 6, 1004–1034.

[128] M. Shcherbina and T. Shcherbina, *Characteristic polynomials for 1D random band matrices from the localization side*, Comm. Math. Phys. **351** (2017), no. 3, 1009–1044.

[129] T. Shcherbina, *On universality of local edge regime for the deformed Gaussian unitary ensemble*, J. Stat. Phys. **143** (2011), no. 3, 455–481.

[130] T. Shcherbina, *On the second mixed moment of the characteristic polynomials of 1D band matrices*, Comm. Math. Phys. **328** (2014), no. 1, 45–82.

[131] D. Shlyakhtenko, *Random Gaussian band matrices and freeness with amalgamation*, Internat. Math. Res. Notices (1996), no. 20, 1013–1025.

[132] D. Shlyakhtenko and P. Skoufranis, *Freely independent random variables with non-atomic distributions*, Trans. Amer. Math. Soc. **367** (2015), no. 9, 6267–6291.

[133] J. W. Silverstein, *The smallest eigenvalue of a large-dimensional Wishart matrix*, Ann. Probab. **13** (1985), no. 4, 1364–1368.

[134] J. W. Silverstein and Z. D. Bai, *On the empirical distribution of eigenvalues of a class of large-dimensional random matrices*, J. Multivariate Anal. **54** (1995), no. 2, 175–192.

[135] H. Sompolinsky, A. Crisanti, and H. J. Sommers, *Chaos in random neural networks*, Phys. Rev. Lett. **61** (1988), 259–262.

[136] A. Soshnikov, *Universality at the edge of the spectrum in Wigner random matrices*, Comm. Math. Phys. **207** (1999), no. 3, 697–733.

[137] R. Speicher, *Combinatorial theory of the free product with amalgamation and operator-valued free probability theory*, Mem. Amer. Math. Soc. **132** (1998), no. 627, x+88.

[138] M. Takesaki, *Theory of operator algebras. I* (Springer-Verlag, New York-Heidelberg, 1979), pp. vii+415.

[139] T. Tao, *Erratum to: Outliers in the spectrum of iid matrices with bounded rank perturbations [mr3010398]*, Probab. Theory Related Fields **157** (2013), no. 1-2, 511–514.

[140] T. Tao, *Outliers in the spectrum of iid matrices with bounded rank perturbations*, Probab. Theory Related Fields **155** (2013), no. 1-2, 231–263.

[141] T. Tao and V. Vu, *Random covariance matrices: universality of local statistics of eigenvalues*, Ann. Probab. **40** (2012), no. 3, 1285–1315.

[142] T. Tao and V. Vu, *Random matrices: the circular law*, Commun. Contemp. Math. **10** (2008), no. 2, 261–307.

[143] T. Tao and V. Vu, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. **38** (2010), no. 5, With an appendix by Manjunath Krishnapur, 2023–2065.

[144] T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1, 127–204.

[145] T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics up to the edge*, Comm. Math. Phys. **298** (2010), no. 2, 549–572.

[146] T. Tao and V. Vu, *Random matrices: universality of local spectral statistics of non-Hermitian matrices*, Ann. Probab. **43** (2015), no. 2, 782–874.

[147] E. Telatar, *Capacity of multi-antenna Gaussian channels*, Eur. Trans. Telecommun. **10** (1999), no. 6, 585–595.

[148] C. A. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys. **159** (1994), no. 1, 151–174.

[149] C. A. Tracy and H. Widom, *On orthogonal and symplectic matrix ensembles*, Comm. Math. Phys. **177** (1996), no. 3, 727–754.

[150] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*, Found. Trends Commun. Inf. Theory **1** (2004), no. 1, 1–182.

[151] D. Voiculescu, *The analogues of entropy and of Fisher's information measure in free probability theory. II*, Invent. Math. **118** (1994), no. 3, 411–440.

[152] D. Voiculescu, *The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms*, Invent. Math. **132** (1998), no. 1, 189–227.

[153] D. Voiculescu, *Free entropy*, Bull. London Math. Soc. **34** (2002), no. 3, 257–278.

[154] D. Voiculescu, *Operations on certain non-commutative operator-valued random variables*, Astérisque (1995), no. 232, Recent advances in operator algebras (Orléans, 1992), 243–275.

[155] D. Voiculescu, *The coalgebra of the free difference quotient and free probability*, Internat. Math. Res. Notices (2000), no. 2, 79–106.

[156] F. J. Wegner, *Disordered system with n orbitals per site: $n = \infty$ limit*, Phys. Rev. B **19** (1979), 783–792.

[157] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564.

[158] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327.

[159] E. P. Wigner, *Results and theory of resonance absorption*, in Gatlinburg conference on neutron physics by time-of-flight, oak ridge natl. lab. rept. no. ornl-2309 (1956), pp. 59–70.

[160] J. Wishart, *The generalised product moment distribution in samples from a normal multivariate population*, Biometrika **20A** (1928), no. 1-2, 32–52.

[161] H. Xi, F. Yang, and J. Yin, *Local circular law for the product of a deterministic matrix with a random matrix*, Electron. J. Probab. **22** (2017), Paper no. 60, 77 pp.

[162] J. Yin, *The local circular law III: general case*, Probab. Theory Related Fields **160** (2014), no. 3-4, 679–732.