Dissertation
an der Fakultät für Mathematik, Informatik und Statistik der
Ludwig-Maximilians-Universität München

Local spectral universality for random matrices with independent entries

by

Torben Krüger

19th August 2015

Erstgutachter: László Erdős

Zweitgutachter: Antti Knowles

Drittgutachter: Sven Bachmann

Disputation:
6th November 2015

Abstract

We consider the local eigenvalue statistics of large self-adjoint $N \times N$ - random matrices, $\mathbf{H} = \mathbf{H}^*$, with centred independent entries. In contrast to previous works the matrix of variances, $s_{ij} = \mathbb{E}\,|h_{ij}|^2$, is not assumed to be stochastic. Hence the density of states is not the Wigner semicircle law. In this work we prove that as N tends to infinity the k - point correlation function of finitely many eigenvalues becomes universal, i.e., it depends only on the symmetry class of the underlying random matrix ensemble and not on the distributions of its entries. The proof consists of three major steps. In the first step we analyse the solution, $\mathbf{m}(z) = (m_1(z), \dots, m_N(z))$, of the quadratic vector equation (QVE), $-1/m_i(z) = z + \sum_j s_{ij} m_j(z)$, for any complex number z. We show that the entries, m_i, can be represented as Stieltjes transforms of probability densities on the real line. We characterise these densities in terms of their singularities, which are algebraic of degree at most three. We present a complete stability analysis of the QVE everywhere, including the vicinity of the singularities. This stability analysis is used in the second step. Here we prove that the diagonal elements of the resolvent, $\mathbf{G} = (\mathbf{H} - z)^{-1}$, satisfy the perturbed QVE, $-1/G_{ii}(z) = z + \sum_j s_{ij} G_{jj}(z) + d_i(z)$, with a random noise vector $\mathbf{d}$. We show that as N grows the noise vanishes and the resolvent is close to the deterministic diagonal matrix $\mathrm{diag}(m_1, \dots, m_N)$. This result is shown with a precision down to the finest spectral scale, just above the typical eigenvalue spacing. It thus implies the local law and rigidity of the eigenvalue positions for this random matrix model. In the third and final step, we use the Dyson-Brownian-motion to establish universality of the local eigenvalue statistics.

Zusammenfassung

Wir analysieren die lokale Eigenwertstatistik großer selbstadjungierter $N \times N$ - Zufallsmatrizen, $\mathbf{H} = \mathbf{H}^*$, mit unabhängigen und zentrierten Einträgen. Anders als in vorangegangenen Arbeiten nehmen wir nicht an, dass die Matrix der Varianzen, $s_{ij} = \mathbb{E}\,|h_{ij}|^2$, stochastisch ist. Insbesondere ist somit auch die globale Eigenwertdichte nicht durch Wigners Halbkreisverteilung gegeben. Wir beweisen in dieser Arbeit, dass mit wachsender Größe N der Zufallsmatrix die k-Punktfunktion endlich vieler Eigenwerte einem universellen Limes entgegen strebt. Dieser ist ausschließlich durch die Symmetrieklasse des zugrundeliegenden Matrixensembles bestimmt und von den Details der Verteilung der individuellen Einträge unabhängig. Der Beweis wird in drei Schritten geführt. Im ersten Schritt analysieren wir die Lösung, $\mathbf{m}(z) = (m_1(z), \ldots, m_N(z))$, der quadratischen Vektorgleichung (QVE), $-1/m_i(z) = z + \sum_j s_{ij} m_j(z)$, in der z eine komplexe Zahl ist. Wir zeigen, dass die Komponenten, m_i, der Lösung als Stieltjes-Transformation gewisser Wahrscheinlichkeitsdichten auf der reellen Achse dargestellt werden können. Wir charakterisieren diese Dichten anhand ihres Singularitätsverhaltens und zeigen dass dieses höchstens von algebraischer Ordnung drei ist. Wir führen eine vollständige Stabilitätsanalyse der QVE durch, welche auch die Umgebung der Singularitäten einschließt. Diese wird im zweiten Schritt des Beweises verwendet, in welchem wir zeigen, dass die Diagonaleinträge der Resolvente, $\mathbf{G} = (\mathbf{H} - z)^{-1}$, die gestörte QVE, $-1/G_{ii}(z) = z + \sum_j s_{ij} G_{jj}(z) + d_i(z)$, mit einer zufälligen vektorwertigen Störung, $\mathbf{d}$, erfüllen. Da mit wachsendem N die Störung gegen Null konvergiert, nähert sich die Resolvente im Limes der deterministischen Diagonalmatrix $\mathrm{diag}(m_1, \ldots, m_N)$ an. Dieses Resultat wird mit einer optimalen spektralen Auflösung gezeigt, welche knapp über dem typischen Abstand der Eigenwerte liegt. Als Konsequenz sehen wir, dass die Fluktuation der Eigenwerte die durch diese Auflösung gegebene Größenordung nicht übersteigt. Im dritten und letzten Schritt nutzen wir den von Dyson eingeführten Prozess der Dyson-Brownschen Bewegung der Eigenwerte und die Kürze seine lokalen Relaxationszeit um die Universalität der lokale Eigenwertstatistik zu beweisen.

To Evi and Lijan

Contents

Acknowledgements

At this point I want to express my deep thanks to all those who over the years helped develop and improve my intuition, knowledge and most of all fascination for physics and mathematics. A special thanks goes to Detlef Dürr, Erwin Frey and Heinz Siedentop who introduced me to many concepts of mathematical physics. I thank my colleagues and friends for countless inspiring discussions. Among these I am very glad to count Johannes Knebel, with whom I have a fruitful collaboration, and Oskari Ajanki, who contributed to this work. I consider myself extremely lucky to have had both Franz Merkl and László Erdős as my advisors. I owe most of my mathematics background to Franz. He has been the best teacher I could have wished for and I want to thank him for all the long discussions, his patience and his support. László is my advisor for this thesis and it has been an absolute pleasure to work with him during the last few years. I cannot thank him enough for his continuing support and the many many days and nights he spend with me in front of a board, calculating, explaining, arguing and teaching. Furthermore, I thank my family, my brothers, Eiko and Hendrik, and my parents, Anne and Bernd, for always being there for me when I need them. Finally, all of this would not be worth much without Evi, but here words will not suffice...

Structure of this work

In this work we prove the local law and the universality conjecture for random matrices with independent entries. The work is split into three parts. In Part I we present a pedagogical introduction into the problem and outline the strategy of the proofs by restricting ourselves to a simplified set-up. In Part II and Part III, we present the new scientific results of this thesis and provide the complete proofs. Apart from minor modifications, Part II and III coincide both in content and writing with [1] and [2], respectively. Certain paragraphs concerning the background of the problem in Section 1 of Part I can be found in [1] and [2] as well. The main statements in Part I, Theorems 2.1, 2.2 and 2.4, are simplified versions of Theorems 6.2 and 6.4 in Part II, as well as Theorems 15.6 and 15.14 in Part III, and thus of the corresponding results from [1] and [2]. In Part I we give an outline of the proofs of these simplified theorems that follow the same ideas as the proofs of their more general counter parts. Reading the presentation in Part I, which cannot be found in [1] and [2], is recommended for an overview of the relevant mechanisms without attention to technical details. In Part II we investigate the *quadratic vector equation (QVE)*. This equation naturally arises in the resolvent expansion method and is satisfied by the diagonal entries of the resolvent of the random matrix in the limit as the size of the matrix tends to infinity. Very detailed knowledge about the solution of this equation and its stability against perturbations is a prerequisite for the analysis carried out in Part III. Here we prove the local law and bulk universality. The papers [1] and [2] are a joint work with László Erdős and Oskari Ajanki.

Part I

1 Introduction

In his seminar paper [64] Wigner introduced random self-adjoint matrices, $\mathbf{H} = \mathbf{H}^*$, with centred, identically distributed and independent entries (subject to the symmetry constraint). He proved that as the size of the matrix grows the empirical density of the eigenvalues converges to the semicircle distribution and he conjectured that the distribution of the distance between consecutive eigenvalues (*gap statistics*) is universal, hence it is the same as in the Gaussian model (GOE/GUE/GSE) with the same symmetry class.

In the Gaussian case all entries are (up to symmetry constraints) i.i.d. standard Gaussian random variables. The invariance of these ensembles under their large symmetry groups allows one to compute the common eigenvalue distribution explicitly. Its density with respect to the N-dimensional Lebesgue-measure has the form

$$\rho^{(N)}(\lambda_1, \dots, \lambda_N) = c_{N,\beta} \prod_{i \neq j} |\lambda_i - \lambda_j|^{\beta} \mathrm{e}^{-N \frac{\beta}{2} \sum_{i=1}^{N} \lambda_i^2},$$

where β equals $1, 2$ or 4, depending on whether the symmetry class is real symmetric (GOE), complex Hermitian (GUE) or symplectic (GSE), N denotes the size of the matrix and $c_{N,\beta}$ is a normalisation constant. The density of states (or 1-point function) for a N-particle distribution is the integral of $\rho^{(N)}$ over $N-1$ variables, $\lambda_2, \dots, \lambda_N$. For the standard Gaussian ensembles this density is Wigner's famous semicircle law,

$$\rho_{\mathrm{sc}}(\lambda) := \frac{1}{2\pi} \sqrt{(4 - \lambda^2)_+} \,. \tag{1.1}$$

In the case of $\beta = 2$ the properly normalised k-point function (with all but k variables integrated out from $\rho^{(N)}$) can be written as a determinant,

$$\rho_k^{(N)}(\lambda_1, \dots, \lambda_k) = \frac{(N-k)!}{N!} \det(K^{(N)}(\lambda_i, \lambda_j))_{i,j=1}^{k} \,,$$

where the kernel $K^{(N)}$ is explicitly expressible in terms of orthogonal polynomials. Studying the asymptotics of these polynomials reveals that in the limit as N tends to infinity the local eigenvalue statistics of the GUE is identical to a determinantal point process characterised by the Dyson sine kernel,

$$\frac{1}{N\rho(\lambda)} K^{(N)}\Big(\lambda + \frac{x_1}{N\rho(\lambda)}, \lambda + \frac{x_2}{N\rho(\lambda)}\Big) \to \frac{\sin \pi(x_1 - x_2)}{\pi(x_1 - x_2)} \,, \qquad N \to \infty \,.$$

This kernel is universal in the sense hat it does not depend on the position λ in the spectrum as long as the density of states does not vanish, $\rho(\lambda) > 0$, i.e., when λ lies inside the *bulk* of the spectrum. An analogous procedure for $\beta = 1$ and $\beta = 4$ uses skew orthogonal polynomials and leads to local spectral universality in the form of a determinantal point process with an explicit kernel. Results of this type on eigenvalue statistics of the Gaussian ensembles in the bulk spectrum were rigorously proven first by Dyson, Mehta and Gaudin in the 60's.

Wigner's revolutionary observation was that these universality phenomena hold for much larger classes of physical systems and that only the basic symmetry type determines local spectral statistics. It is generally believed, but mathematically unproven, that random matrix theory (RMT), among many other examples, also describes the local statistics of random Schrödinger operators in the delocalised regime and quantisation of chaotic classical Hamiltonians. Eigenvalue statistics predicted by RMT are observed in areas as diverse as the distribution of zeros of the Riemann-ζ-function [6], low energy vibration in large molecules [20], statistics of neutron resonances in heavy nuclei [49] and eigenvalues of the Dirac-operator in QCD [63]. For none of these examples has universality been proven with mathematical rigour. Nevertheless, there have been significant improvements in the understanding of the mechanisms that lead to this phenomenon. These improvements made it possible to establish universality for a wide class of random matrices, including Wigner's original model.

According to Wigner's universality hypothesis, universality of eigenvalue statistics should hold for random matrices with i.i.d. entries independently of the law of the matrix elements.

This conjecture, also known as the Wigner-Dyson-Mehta conjecture, was resolved recently in a series of works. The strongest result on Wigner matrices in the bulk spectrum is Theorem 7.2 in [24], see [35] and [59] for a summary of the history and related results. In fact, the *three-step* approach developed in [33, 36, 26] also applies for *generalised Wigner matrices* that allow for non-identically distributed matrix elements as long as the variance matrix $s_{ij} := \mathbb{E}|h_{ij}|^2$ is stochastic, i.e. $\sum_j s_{ij} = 1$ (in particular, independent of i). The stochasticity of $\mathbf{S}$ guarantees that the eigenvalue density is given by the semicircle law and the diagonal elements G_{ii} of the resolvent $\mathbf{G} = (\mathbf{H} - z)^{-1}$ with $\mathrm{Im}\, z > 0$ become not only deterministic but also independent of i as the matrix size N goes to infinity. Second order perturbation theory indicates that they asymptotically satisfy a system of self-consistent equations

$$-\frac{1}{G_{ii}} \approx z + \sum_{j=1}^{N} s_{ij} G_{jj} \,. \tag{1.2}$$

In the case of a stochastic variance matrix this becomes a particularly simple scalar equation

$$-\frac{1}{m_{\mathrm{sc}}} = z + m_{\mathrm{sc}} \,, \tag{1.3}$$

for the common value $m_{\mathrm{sc}} \approx G_{ii}$ for all i as $N \to \infty$. The solution of (1.3) is the Stieltjes transform of the Wigner semicircle law,

$$m_{\mathrm{sc}}(z) = \int_{\mathbb{R}} \frac{\rho_{\mathrm{sc}}(\tau)\mathrm{d}\tau}{\tau - z} \,. \tag{1.4}$$

In this work we consider a general variance matrix $\mathbf{S}$ without stochasticity condition. The corresponding general random matrix with independent entries is said to be of *Wigner-type*. We show that the approximate self-consistent equation (1.2) still holds, but it does not simplify to a scalar equation. In fact, for any complex number z in the upper half plane

$$\mathbb{H} := \{z \in \mathbb{C} : \ \mathrm{Im}\, z > 0\} \,,$$

the diagonal resolvent entries $G_{ii}(z)$ remain i-dependent even as N tends to infinity and are close to the solution $m_i = m_i(z)$ of the *Quadratic Vector Equation (QVE)*

$$-\frac{1}{m_i} = z + \sum_{j=1}^{N} s_{ij} m_j \,, \qquad i = 1, \dots, N \,, \tag{1.5}$$

for N numbers $m_1, \dots, m_N \in \mathbb{H}$. In this context the importance of this equation has been realised by Girko [41], see also the work of Helton, Far, and Speicher [43], and Anderson and Zeitouni [5], but no detailed study has been initiated that would allow for establishing universality.

The main goal of Part II of this work is to give a detailed analysis of the solution of this system of non-linear equations. Several qualitative and quantitative aspects may be considered, but we are especially interested in three issues: (i) regularity up to the real axis, apart from a few singular points; (ii) classification of these singularities; (iii) stability of the solution of (1.5) under small perturbations. We show that $\langle \mathbf{m} \rangle := \frac{1}{N}\sum_i m_i$, which is close to $N^{-1}\,\mathrm{Tr}\,\mathbf{G}$ in the limit, is the Stieltjes transform of a density ρ on the real line that is not the semicircle law in general, but it is still a real analytic function on the interior of its support. This function, ρ, describes the asymptotic density of states. We also classify its asymptotic behaviour near the edges of its support. It features only square root or cubic root (cusp) singularities and an explicit one parameter family of shape functions interpolating between them as a gap in the support closes.

The main result of Part III of this work is the universality of the local eigenvalue statistics in the bulk for Wigner-type matrices with a general variance matrix. This extends Wigner's vision towards full universality by considering a much larger class of matrix ensembles than previously studied. In particular, we demonstrate that local statistics, as expected, are fully independent of the global density. This fact has already been established for very general β-ensembles in [14] (see also [11] and [54]) and for additively deformed Wigner ensembles having a density with a single interval support [48]. Our class admits a general variance matrix and allows for densities with several intervals (we do not, however, consider non-centred distributions here, apart from an extension to matrices with non-centred entries on the diagonal in Appendix A.1).

This is the main and novel part of our analysis. The previous proofs (see [26] for a pedagogical presentation) heavily relied on properties of the semicircle law, especially on its square root edge singularity. Following the three-step approach, we first prove *local laws* for $\mathbf{G}$ on the scale $\eta = \mathrm{Im}\, z \gg N^{-1}$, i.e. down to the optimal scale just slightly above the eigenvalue spacing. With possible cubic root singularities and small gaps in the support of ρ an additional scale appears which needs to be controlled. The second step is to prove universality for Wigner-type matrices with a tiny Gaussian component via *Dyson Brownian motion (DBM)*. The method of *local relaxation flow*, introduced first in [32, 33], also heavily relies on the semicircle law since it requires that the global density remain unchanged along the DBM. In [34], and independently in [47], a new method was developed to localise the DBM that proves universality of the gap statistics around a fixed energy τ in the bulk, assuming that the local law holds near τ. Since Wigner-type matrices were one of the main motivations for [34], it was formulated such that it could be directly applied once the local laws are available. Finally, the third step is a perturbation result to remove the tiny Gaussian component using the *Green's function comparison* method that first appeared in [36] and can be applied to our case basically without any modifications.

Within Part III we also apply our results to Gaussian random matrices with *correlated entries*. Most rigorous works on random matrix ensembles concern either Wigner matrices with independent entries (up to the symmetry constraint $h_{ij} = \bar{h}_{ji}$), or invariant ensembles where the correlation structure of the matrix elements is very specific, namely the probability measure on the space of self-adjoint matrices has the form

$$\mathrm{P}^{(N)}(\mathrm{d}\mathbf{H}) \,=\, c_N \, \mathrm{e}^{-\operatorname{Tr} V(\mathbf{H})} \mathrm{d}\mathbf{H} \,.$$

Since the existing methods to study Wigner matrices heavily rely on independence, only very few results are available on ensembles with correlated entries, see [46, 16, 19, 15] for the Gaussian case. The global semicircle law in the non Gaussian case with (appropriately) weakly dependent entries has been established via moment method in [53] and via resolvent method in [42]. A similar result for sample covariance matrices was given in [51]. All these works establish limiting spectral density only on the macroscopic scale and in models where the dependence is sufficiently weak so that the limiting density of states coincides with that of the independent case. A more general correlation structure was explored in [5] with a nontrivial limit density, but still only on the global scale, see also [50]. We also mention the very recent proof of the local semicircle law and bulk universality for the adjacency graph of the d-regular graphs [10, 9] which has a completely different specific correlation (due to the requirement that every row contains the same number of ones).

In our work we explore the simple fact that the (discrete) Fourier transform of correlated Gaussian random matrices with a certain translation invariant correlation structure have almost independent entries (up to an additional symmetry). Since the variance matrix in Fourier space is typically not stochastic, previous results on generalised Wigner matrices are not applicable, but our results on general variance matrices yield local laws. Additionally, we find that the off diagonal resolvent matrix elements G_{ij} are not negligible (unlike in the independent case) and in fact they inherit their decay from the correlation of the matrix elements. As a simple consequence we also get bulk universality.

2 Main results for simplified model

In this section we present the main results of our work in a simplified setting. The setting is chosen in such a way that the main results are easy to state but they and their proofs still reflect the spirit of the general results, Theorem 6.2 and Theorem 6.4 in Part II, as well as Theorem 15.6 and Theorem 15.14 in Part III. Throughout this section and the next we will point out the simplifications made in this set-up and how the results and proofs differ from the more general setting of Part II and III. Part I is meant as a pedagogical introduction without attention to technical details. In particular, the results presented in this part are simple consequences of the theorems stated in Part II and III. Their proofs are only sketched in this part in order to give an overview of the main ideas.

The first simplification is that we state our result only for the real symmetric symmetry class and we assume that the values of the variances of the matrix entries stem from a profile function (cf. assumption 4. below). Let $\mathbf{H} \in \mathbb{R}^{N\times N}$ be a symmetric random matrix whose entries, h_{ij}, satisfy the following assumptions:

1. The entries h_{ij} are independent for $1 \le i \le j \le N$.
2. The random matrix is centred, $\mathbb{E}\, h_{ij} = 0$.
3. All moments are bounded in terms of the variance, i.e., for all $k \in \mathbb{N}$ there is a positive constant μ_k such that
$$\mathbb{E}\, h_{ij}^{2k} \,\le\, \mu_k (\mathbb{E}\, h_{ij}^2)^k .$$
4. The variances converge to a positive, Hölder-continuous *profile function*, i.e. there is a symmetric, $s(x,y) = s(y,x)$, Hölder-continuous function $s : [0,1]^2 \to (0,\infty)$ with Hölder-exponent $1/2$ such that
$$\mathbb{E}\, h_{ij}^2 \,=\, \frac{1}{N}\, s\Big(\frac{i}{N}, \frac{j}{N}\Big) .$$

For every nonnegative symmetric function, $s : [0,1]^2 \to \mathbb{R}$, we consider the *quadratic vector equation (QVE)*

$$-\frac{1}{m(x;z)} \,=\, z + \int_0^1 s(x,y) m(y;z) \mathrm{d}y \,, \qquad x \in [0,1]\,,\ z \in \mathbb{H}\,, \tag{2.1}$$

for a function $m : [0,1] \times \mathbb{H} \to \mathbb{H}$. By Theorem 6.1 in Part II this equation has a unique solution. This qualitative result has been established prior to our work in [43, 5].

The following theorem is a simplified version of the main result of our work. It states that as the size of the random matrix, $\mathbf{H}$, grows, its resolvent converges to a diagonal matrix. While the off-diagonal entries all approach zero, the diagonal entries of the resolvent converge to a deterministic value that is given in terms of the solution, m, of equation (2.1).

THEOREM 2.1 (Local law for simplified model). *Suppose* $\mathbf{H}$ *satisfies assumptions 1. - 4. above. For every (small)* $\varepsilon > 0$ *and (large)* $D > 0$ *there exists a threshold* $N_0 \in \mathbb{N}$*, depending only on the bound on the moments,* $\mu = (\mu_k)_{k\in\mathbb{N}}$*, the profile of the variances,* s*, and on* ε *and* D*, such that for every* $z \in \mathbb{C}$ *with* $\operatorname{Im} z \ge N^{\varepsilon-1}$ *the resolvent,* $\mathbf{G}(z) = (G_{ij}(z))_{i,j=1}^N := (\mathbf{H} - z)^{-1}$ *is close* *to a diagonal matrix, whose entries are given in terms of the solution m of the QVE* (2.1)*, in the following sense,*

$$\mathbb{P}\biggl[\max_{i,j=1}^{N}\bigl|G_{ij}(z)-m(i/N;z)\delta_{ij}\bigr| \ge \frac{N^{\varepsilon}}{\sqrt{N\operatorname{Im} z}}\biggr] \le N^{-D}, \qquad N \ge N_0\,. \tag{2.2}$$

The average diagonal resolvent entries satisfy the improved bound

$$\mathbb{P}\biggl[\Bigl|\frac{1}{N}\operatorname{Tr}\mathbf{G}(z)-\int_0^1 m(x;z)\mathrm{d}x\Bigr| \ge \frac{N^{\varepsilon}}{N\operatorname{Im} z}\biggr] \le N^{-D}, \qquad N \ge N_0\,. \tag{2.3}$$

Theorem 2.1 provides control on the entries of the resolvent on the optimal spectral scale, just above the typical eigenvalue spacing $\sim N^{-1}$. The imaginary part of the spectral parameter z is a measure for the scale on which the information about the spectrum of $\mathbf{H}$ is resolved. This can be seen from

$$\frac{1}{N}\operatorname{Im}\operatorname{Tr}\mathbf{G}(\tau+\mathrm{i}\eta) \;=\; \frac{1}{N}\sum_{i=1}^{N}\frac{\eta}{\eta^2+(\tau-\lambda_i)^2}\,, \tag{2.4}$$

where $(\lambda_i)_{i=1}^N$ denote the eigenvalues of $\mathbf{H}$. The expression on the right hand side is a smoothed out version of the spectral measure $\frac{1}{N}\sum_i\delta_{\lambda_i}$ of the eigenvalue process. The scale on which this measure is regularised is the given by the parameter η in the approximate delta functions on the right hand side in the form of a Cauchy kernel.

The local law provides information about the number of eigenvalues in an interval $[\alpha_N,\beta_N]$, provided the size of the interval stays above the typical eigenvalue spacing in the bulk of the spectrum, $\beta_N-\alpha_N \gg N^{-1}$. In particular, it implies that in the bulk the eigenvalues are not further away from their expected positions than $N^{\varepsilon-1}$. This *rigidity* is an indication of the strong correlation between the eigenvalues. If the eigenvalue process consisted of completely independent points, the local universality class would be the poisson point process. The points of this process have a typical fluctuation on the scale $N^{-1/2}$, which is much larger than N^{-1}.

Another immediate consequence of Theorem 2.1 is the *complete delocalisation* of the eigenvectors of $\mathbf{H}$. As a measure of localisation we consider the maximum norm, $\|\mathbf{u}\|_\infty=\max_i|u_i|$, of a ℓ^2-normalised eigenvector $\mathbf{u}$ of the random matrix. If $\mathbf{u}$ were localised then its ℓ^2-mass would be concentrated on a few entries and $\|\mathbf{u}\|_\infty\sim 1$. On the other hand, if all the entries of the eigenvector are roughly of the same size, then $\mathbf{u}$ is completely delocalised and $\|\mathbf{u}\|_\infty\sim N^{-1/2}$. That such a delocalisation result can be inferred from Theorem 2.1 can be seen from the simple calculation,

$$\eta\operatorname{Im}G_{kk}(\lambda_j+\mathrm{i}\eta) \;=\; \eta\operatorname{Im}\sum_{i=1}^{N}\frac{\bigl|u_k^{(i)}\bigr|^2}{\lambda_i-\lambda_j-\mathrm{i}\eta} \;\ge\; \bigl|u_k^{(j)}\bigr|^2,$$

where $\mathbf{u}^{(i)}=(u_j^{(i)})_{j=1}^N$ denotes the normalised eigenvector corresponding to the eigenvalue λ_i. Since by the local law $\operatorname{Im}G_{kk}$ remains bounded for the choice $\eta:=N^{\varepsilon-1}$ with high probability this implies the bound

$$\|\mathbf{u}^{(i)}\|_\infty^2 \;\le\; CN^{\varepsilon-1},$$

proving the complete delocalisation of eigenvectors with high probability.

The proof of the local law requires a good understanding of the deterministic limit, $m(x;z)$, of the diagonal resolvent entries. Furthermore, having detailed knowledge about the solution of the QVE means having detailed knowledge about spectral properties of $\mathbf{H}$ in the large N limit. For example, the eigenvalue density of $\mathbf{H}$ converges in probability to the x-average of the imaginary part of $m(x;z)$, i.e.,

$$N^{-1}\#\{i:\lambda_i\le\alpha\} \;\overset{\mathbb{P}}{\to}\; \lim_{\eta\downarrow 0}\frac{1}{\pi}\int_{-\infty}^{\alpha}\int_0^1\operatorname{Im} m(x;\tau+\mathrm{i}\eta)\mathrm{d}x\,\mathrm{d}\tau\,. \tag{2.5}$$

This can be deduced from the identity (2.4) and the averaged local law, (2.3). The following theorem shows that the solution, $m(x;\cdot)$, of the QVE admits a representation as the Stieltjes transform of x-dependent densities that are all supported on a common interval with a square root growth at the edge of the support. The edge growth resembles the behaviour of the famous semicircle law and is a feature of our simplified setting. In the general case, treated in Part II, the support of the densities may consist of several disjoint intervals with edge singularities ranging from a square root to a cubic root growth.

THEOREM 2.2 (Solution of QVE for simplified model). *Let m be the solution of the QVE* (2.1). *Then there exists a positive constant β and a positive, continuous function, $h : [0,1]\times[-\beta,\beta] \to (0,\infty)$, which is analytic and even in its second variable on the open interval $(-\beta,\beta)$, such that m admits the Stieltjes transform representation*

$$m(x;z) \;=\; \int_{-\beta}^{\beta} \frac{h(x;\tau)\sqrt{\beta^2-\tau^2}}{\tau - z}\,\mathrm{d}\tau\,, \qquad x\in[0,1]\,,\; z\in\mathbb{H}\,.$$

Theorem 2.2 shows that m can be completely recovered from the values of its imaginary part, $\operatorname{Im} m$, for z close to the real axis. This justifies the position, taken also in Theorem 6.4 of Part II, that understanding m is equivalent to understanding $\lim_{\eta\downarrow 0}\operatorname{Im} m(x;\tau+\mathrm{i}\eta)$. The following definition is motivated by (2.5).

DEFINITION 2.3 (Density of states). *We define the density of states, $\rho:\mathbb{R}\to[0,\infty)$, by*

$$\rho(\tau) \;:=\; \lim_{\eta\downarrow 0}\frac{1}{\pi}\int_0^1 \operatorname{Im} m(x;\tau+\mathrm{i}\eta)\mathrm{d}x\,.$$

By Theorem 2.2 the density of states has a square root growth at the edges $-\beta$ and β. In fact, ρ satisfies

$$\rho(\tau) \;=\; \frac{1}{\pi}\int_0^1 h(x;\tau)\mathrm{d}x\,\sqrt{\beta^2-\tau^2}\;\mathbb{1}(|\tau|\le\beta)\,,$$

where the function $\tau\mapsto\int h(x;\tau)\mathrm{d}x$ is continuous and positive.

Using the method of Dyson Brownian motion developed in [32, 33], and tailored to our specific situation in [34, 47], we infer bulk universality from the rigidity of the eigenvalues, which itself is a consequence of the local law, Theorem 2.1 (cf. Section 3 below for more details).

THEOREM 2.4 (Universality for simplified model). *Let $\mathbf{H}$ satisfy assumptions 1. - 4. above. For any $\varepsilon>0$, $n\in\mathbb{N}$ and any smooth compactly supported test function $F:\mathbb{R}^n\to\mathbb{R}$, there are positive constants C and c that only depend on μ, s, ε, n and F, such that for $i\in[\varepsilon N,(1-\varepsilon)N]$ bulk universality holds*

$$\Big|\mathbb{E}\,F\Big(\big(N\rho(\lambda_i)(\lambda_i-\lambda_{i+j})\big)_{j=1}^{n}\Big) - \mathbb{E}_{\mathrm{GOE}}\,F\Big(\big(N\rho_{\mathrm{sc}}(\lambda_i)(\lambda_i-\lambda_{i+j})\big)_{j=1}^{n}\Big)\Big| \;\le\; CN^{-c}. \tag{2.6}$$

The expectation, $\mathbb{E}_{\mathrm{GOE}}$, means that with respect to the underlying probability measure the random matrix $\mathbf{H}$ is of the form $\mathbf{H}=\frac{1}{\sqrt{2}}(\mathbf{W}+\mathbf{W}^t)$, where the entries of $\mathbf{W}$ are i.i.d. centred Gaussian random variables with variance N^{-1}. The function ρ_{sc} is the density of states in this case, the Wigner semicircle law (cf. (1.1)).

3 Universality in three steps

The three step approach to universality we present here has been developed in a series of papers, [29, 30, 31, 27]. It relies on an idea by Dyson [22] to consider the evolution of the eigenvalues of a matrix as the entries undergo Brownian motion. The local relaxation of this *Dyson Brownian motion (DBM)* is then used to establish spectral universality. The three steps used in this method are:

- Local law and rigidity
- Dyson Brownian motion
- Green's function comparison theorem

In the following we will explain each step.

3.1 Local law and rigidity

Starting point of the proof of universality is to establish a local law in the spirit of Theorem 2.1. The local law provides detailed information about the resolvent of the random matrix under consideration. This, in turn, leads to a good control on the position of the eigenvalues of the random matrix. In fact, a corollary of Theorem 2.1 is the following rigidity result: For all $\varepsilon, D > 0$ we have

$$\max_{i=1}^{N} \mathbb{P}\left[|\lambda_i - \gamma_i| \,\geq\, \frac{N^{\varepsilon}}{N^{2/3} \min\{i, N+1-i\}^{1/3}} \right] \,\leq\, N^{-D} \qquad N \geq N_0(\varepsilon, D)\,. \tag{3.1}$$

Here $N_0(\varepsilon, D)$ is a threshold function depending on the model parameters μ and f. The *classical position*, γ_i, is defined by the identity

$$\int_{-\infty}^{\gamma_i} \rho(\tau)\mathrm{d}\tau \,=\, \frac{i}{N}\,.$$

For a proof of this fact, we refer to Part III of this work.

3.2 Dyson Brownian motion

At this stage of the proof the local law is given and we establish the bulk universality statement, Theorem 2.4, for matrices with a small GOE component. This means that in addition to assumptions 1. - 4. the random matrix $\mathbf{H}$ is assumed to be of the form

$$\mathbf{H} \,=\, \mathbf{H}_0 + \sqrt{T}\,\mathbf{W}, \tag{3.2}$$

with a random matrix $\mathbf{H}_0$, still satisfying assumptions 1. - 4., and an independent GOE-matrix $\mathbf{W}$. The positive parameter, T, encodes the size of the GOE-component. Depending on the strength of the rigidity result from the first step and the strength of the result on the local relaxation of the DBM this parameter is assumed to be of the size $T \in [N^{\varepsilon-1}, N^{-\varepsilon}]$. With (3.1) and Theorem 2.5 from [47], we may choose $T := N^{\varepsilon-1}$. The extra assumption that $\mathbf{H}$ can be written in the form (3.2) is removed in the third step by applying the *Green's function comparison theorem.*

Now we indicate how the small Gaussian component in (3.2) is used to prove bulk universality. Let $(\widetilde{B}_{ij}(t))_{1\leq i\leq j\leq N}$ be a family of independent standard Brownian motions and define the symmetric matrix $\mathbf{B} = (B_{ij})_{i,j=1}^{N}$ by

$$B_{ij}(t) \,:=\, \begin{cases} \widetilde{B}_{ij}(t) & \text{if } i<j\,, \\ \widetilde{B}_{ji}(t) & \text{if } i>j\,, \\ \sqrt{2}\widetilde{B}_{ii}(t) & \text{if } i=j\,. \end{cases}$$

Then the GOE component in (3.2) can be realised by embedding $\mathbf{H}$ into a flow of random matrices satisfying the simple stochastic differential equation

$$\mathrm{d}\mathbf{H}(t) \;=\; \frac{1}{\sqrt{N}}\,\mathrm{d}\mathbf{B}(t)\,, \qquad \mathbf{H}(0) \;=\; \mathbf{H}_0\,,$$

Indeed, $\mathbf{H}$ has the same distribution as the matrix $\mathbf{H}(T) = \mathbf{H}_0 + \mathbf{B}(T)$. The insight by Dyson in [22] was that the Brownian motion flow on the entries of the matrix induces a closed SDE on the eigenvalues. Let $\lambda_1(t) \le \cdots \le \lambda_N(t)$ denote the eigenvalues of $\mathbf{H}(t)$. Then by a straight forward application of Ito-calculus one heuristically derives the SDE,

$$\mathrm{d}\lambda_i(t) \;=\; \frac{1}{\sqrt{N}}\mathrm{d}b_i(t) + \frac{1}{2N}\sum_{j\neq i}\frac{\mathrm{d}t}{\lambda_i-\lambda_j}\,, \qquad i = 1,\dots,N\,. \tag{3.3}$$

The repulsion between the eigenvalues, represented by the second term on the right hand side, is sufficiently strong that the paths of the λ_i do not cross and one can establish (3.3) as a well-defined system of N strongly coupled stochastic differential equations [4]. It has been shown, [47], see also [34], that under this evolution the common distribution of the particles $(\lambda_i(t))_{i=1}^N$ reaches local equilibrium already after short times $t \ge N^{\varepsilon-1}$, provided the rigidity statement (3.1) holds for the initial value, $(\lambda_i(0))_{i=1}^N$. More precisely, (2.6) holds for $\mathbf{H} = \mathbf{H}(t)$ with $t \ge T = N^{\varepsilon-1}$. This proves bulk universality for matrixes of the form (3.2).

3.3 Green's function comparison theorem

The Green's function comparison theorem allows for the removal of the extra assumption that $\mathbf{H}$ can be written in the form (3.2). It is the basis for an approximation argument, asserting that for any given $\mathbf{H}$ one can always find another random matrix, $\widetilde{\mathbf{H}}$, with a small GOE component such that $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ are sufficiently close in an appropriate sense. The closeness of the two matrices then implies that their local eigenvalue statistics coincide. Thus the bulk universality, that was proven for $\widetilde{\mathbf{H}}$ using the DBM, also implies bulk universality for the original matrix, $\mathbf{H}$.

This type of argument first appeared as the *four moment theorem* (Theorem 6) in [58]. There, $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ were Wigner-matrices ($f = 1$ in our setting) and it was required that the first four moments of their entries coincide. In this work we use the Green's function comparison theorem in the form given in [36] as Theorem 2.3. Adjusting the proof and statement of this theorem to our setting shows that if for two random matrices $\mathbf{H}$ and $\widetilde{\mathbf{H}}$, satisfying assumptions 1. - 4., their first four moments match in the sense that

$$|\mathbb{E}\,h_{ij}^s - \mathbb{E}\,\widetilde{h}_{ij}^s| \;\le\; N^{-2-\delta},\;\; s = 3,4\,,$$

then the expectation of sufficiently smooth observables, F, of resolvent elements coincide for $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ in the limit,

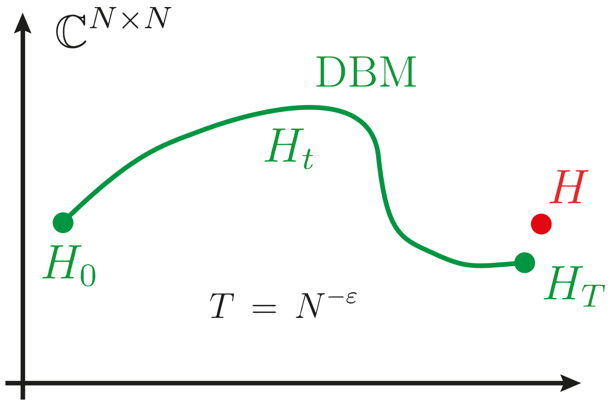

Figure 3.1: To any given matrix $\mathbf{H}$ one can construct a matrix $\mathbf{H}_0$ such that $\mathbf{H}_T$, evolved by the DBM, is close to $\mathbf{H}$ in the four moment sense.

$$\mathbb{E}\,F\Big(N^{1-k_1}\Big[\prod_{j=1}^{k_1}\mathbf{G}\big(z_j^{(1)}\big)\Big]_{i_1 i_2},\dots,N^{1-k_n}\Big[\prod_{j=1}^{k_n}\mathbf{G}\big(z_j^{(n)}\big)\Big]_{i_n i_{n+1}}\Big) \;-\; \mathbb{E}\big(\mathbf{G}\to\widetilde{\mathbf{G}}\big) \;\to\; 0\,, \quad N\to\infty.$$

Here the resolvents are evaluated at spectral parameters $z_j^{(k)}$ with real parts in the bulk of the spectrum and $|\mathrm{Im} z_j^{(k)}| = \eta \in [N^{-1-\varepsilon}, N^{-1}]$. The integers k_l and n are finite, i.e., they do not depend on N. Since the spectral resolution here is just below the typical eigenvalue spacing, individual eigenvalues can be resolved and the we conclude that the local eigenvalue statistics of $\mathbf{H}$ and $\widetilde{\mathbf{H}}$ coincide.

This is useful only for removing the small Gaussian component, assumed in the previous set, only in combination with the *four moment matching* that asserts that for any given $\mathbf{H}$ one can construct and $\widetilde{\mathbf{H}}$ that is close to $\mathbf{H}$ in the four moment sense and has a small Gaussian component [36, 37]. Figure 3.1 illustrates this strategy.

4 Proof of local law

In this section we will explain the strategy of the proof of Theorem 2.1. We will, however, not prove the theorem rigorously here, since it follows directly from Theorem 6.8 in Part II and Theorem 15.6 in Part III. Instead, we will focus on the main ideas and largely ignore technical details. We will only discuss how to show (2.2). The improved averaged local law, (2.3), only requires minor changes, following the size of the error terms more precisely. The proof of the local law starts by showing that the off-diagonal resolvent elements, G_{ij}, converge to zero as the size of the matrix $\mathbf{H}$ grows and that the diagonal entries, G_{ii}, solve a discrete version of the QVE (2.1) with an additional random error term. Analysing the stability of this equation then shows that these entries converge to the solution m of the QVE.

4.1 Resolvent expansion and derivation of QVE

Here we sketch how to derive the QVE for the diagonal elements of $\mathbf{G}$ and how to show that the off-diagonal elements converge to zero as N tends to infinity. More precisely, we will show

$$\max_{i \neq j} \big|G_{ij}(z)\big| \prec \frac{1}{\sqrt{N \,\mathrm{Im}\, z}} \,. \tag{4.1}$$

Here, the relation '$\prec$' means that for any given $\varepsilon, D > 0$ for large enough N the left hand side is smaller than the right hand side up to a factor of N^{ε} and up to a set in the underlying probability space whose probability is at most N^{-D}, i.e., (4.1) is equivalent to (2.2), ignoring the diagonal entries. For a precise definition of '$\prec$' see Definition 15.5 in Part III of this work. In particular, the bounded moment condition on the entries of the random matrix ensures the large deviation estimate

$$\max_{i,j} |h_{ij}| \prec \frac{1}{\sqrt{N}} \,. \tag{4.2}$$

The basis for the derivation of the QVE is the *resolvent expansion method*. Schur's complement formula allows one to compute the inverse of a 2×2-block matrix in terms of the matrices in the individual blocks,

$$\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)^{-1} = \left(\begin{array}{c|c} S^{-1} & -S^{-1}BD^{-1} \\ \hline -D^{-1}CS^{-1} & D^{-1} + D^{-1}CS^{-1}BD^{-1} \end{array}\right), \qquad S := A - BD^{-1}C \,.$$

In the resolvent expansion method one isolates the dependence of the resolvent entries on a particular row and column of the random matrix $\mathbf{H}$ by applying this formula with the following choices. Fix $k \in \{1, \dots, N\}$ and set

$$A := h_{kk} - z \,, \qquad B := (h_{kj})_{j \neq k} \,, \qquad C := (h_{jk})_{j \neq k} \,, \qquad D := \mathbf{H}^{(k)} - z \,,$$

where for any $T \subseteq \{1, \dots, N\}$ we denote by $\mathbf{H}^{(T)}$ the matrix $\mathbf{H}^{(T)} := (h_{ij})_{i,j \notin T}$ of dimension $N - |T|$. Let $\mathbf{G}^{(T)}$ be the resolvent of this matrix, then with these choices Schur's complement formula reads

$$G_{kl} \;=\; -G_{kk} \sum_{i \neq k} h_{ki} G^{(k)}_{il} \;=\; -G_{ll} \sum_{i \neq l} G^{(l)}_{ki} h_{il} \,, \qquad k \neq l \,, \tag{4.3}$$

for the off-diagonal elements. As in (4.3) we often will not write the argument z explicitly. For the diagonal elements we get

$$-\frac{1}{G_{kk}} \;=\; z + \sum_{i,j \neq k} h_{ki} G^{(k)}_{ij} h_{jk} - h_{kk} \,. \tag{4.4}$$

Applying the off-diagonal resolvent formula, (4.3), for G_{kl} and then again for $G^{(k)}_{il}$ results in

$$G_{kl} \;=\; -G_{kk} G^{(k)}_{ll} \sum_{i,j \neq k,l} h_{ki} G^{(kl)}_{ij} h_{jl} - G_{kk} G^{(k)}_{ll} h_{kl} \,. \tag{4.5}$$

The starting point of our analysis are the formulas (4.4) and (4.5). We will now make plausible that for all $k \neq l$ we have

$$|G_{kl}| \;\prec\; \frac{1}{\sqrt{N \operatorname{Im} z}} \quad \text{and} \quad -\frac{1}{G_{kk}} \;=\; z + \sum_{j=1}^{N} (\mathbb{E}\, h_{kj}^2) G_{jj} + d_k \;\; \text{with} \;\; |d_k| \;\prec\; \frac{1}{\sqrt{N \operatorname{Im} z}} \,. \tag{4.6}$$

The equation for the diagonal resolvent entries resembles the QVE (2.1) once we plug in the assumption about the variances of h_{ij}, namely

$$\sum_{j=1}^{N} (\mathbb{E}\, h_{kj}^2) G_{jj} \;=\; \frac{1}{N} \sum_{j=1}^{N} s\Big(\frac{k}{N}, \frac{j}{N}\Big) G_{jj} \,.$$

The right hand side becomes is a Riemann-sum that converges to an integral in the limit as N tends to infinity, provided G_{jj} is sufficiently regular in its index j.

In this simplified presentation we will assume an a priori bound on the diagonal resolvent elements of the form

$$|G_{kk}(z)| + |G^{(l)}_{kk}(z)| + |G^{(lj)}_{kk}(z)| \;\prec\; 1 \,, \qquad \operatorname{Im} z \geq N^{\varepsilon - 1} \,, \tag{4.7}$$

where $k \neq l, j$. This bound holds once the local law is proven, because $G_{kk}(z)$ is then shown to be close to $m(k/N; z)$ and the solution, m, of the QVE is bounded (see Section 4.3). Furthermore, removing the one or two rows and columns from $\mathbf{H}$, does not change the resolvent entries by much, i.e., $G^{(l)}_{kk}$ and $G^{(lj)}_{kk}$ are close to G_{kk}. Of course this argument can only be made once Theorem 2.1 is established. In truth (4.7) and the final result, (2.2), are proven in tandem by a combination of a bootstrap and a continuity argument. The a priori bound is obviously true for $\operatorname{Im} z = 1$ because of the simple fact that $\|\mathbf{G}(z)\| \leq (\operatorname{Im} z)^{-1}$. Then we feed this bound into the derivation of the QVE as we will do below and improve it to the form (2.2) with $\operatorname{Im} z = 1$. Afterwards, the continuity of the resolvent entries in z is used to establish (4.7) for slightly decreased values of $\operatorname{Im} z$ and again fed back into the argument below. This procedure can be continued until $\operatorname{Im} z = N^{\varepsilon - 1}$ and the local law is proven. For more details we refer to Part III of this work. We will also restrict ourselves to bounded values of the spectral parameter z, i.e., $|z| \leq C$, for simplicity. Since the spectrum of $\mathbf{H}$ is expected to lie in a compact interval around the origin, this covers the main difficulty. Proving (2.2) for large $|z|$ can be done by using more qualitative methods, e.g. the moment method.

We will now use (4.5) to get smallness of the off-diagonal resolvent elements. Indeed, plugging the a priori bound, (4.7), and (4.2) into this identity yields

$$|G_{kl}| \;\prec\; \Big|\sum_{i,j\neq k,l} h_{ki}G^{(kl)}_{ij}h_{jl}\Big| + \frac{1}{\sqrt{N}}\,.$$

The reduced resolvent $\mathbf{G}^{(\mathbf{kl})}$ is independent of the k-th and l-th row of $\mathbf{H}$ and both of these rows are independent of each other since $k \neq l$. Thus, the sum on the right hand side of this inequality can be viewed as a quadratic expression of the form $\sum_{ij} b_{ij}X_iY_j$ with independent random variables X_i, Y_j and coefficients b_{ij}. For this expression we apply a large deviation estimate. To get a feeling of its size, we compute its variance,

$$\mathbb{E}\Big|\sum_{ij} b_{ij}X_iY_j\Big|^2 \;=\; \sum_{ij}|b_{ij}|^2\,\mathbb{E}\,|X_i|^2\,\mathbb{E}\,|Y_j|^2\,.$$

The full proof of the large deviation estimate can be found in [25] and confirms the expected result,

$$\Big|\sum_{i,j\neq k,l} h_{ki}G^{(kl)}_{ij}h_{jl}\Big| \;\prec\; \Big(\sum_{i,j\neq k,l}\big|G^{(kl)}_{ij}\big|^2\,\mathbb{E}\,h_{ki}^2\,\mathbb{E}\,h_{jl}^2\Big)^{1/2} \;\leq\; \frac{\|s\|_\infty}{\sqrt{N}}\Big(\frac{1}{N}\sum_{i,j\neq k,l}\big|G^{(kl)}_{ij}\big|^2\Big)^{1/2}. \qquad (4.8)$$

At this stage it is not sufficient to estimate the resolvent entries trivially by $(\operatorname{Im} z)^{-1}$. Not even an a priori estimate of the form $|G^{(kl)}_{ij}| \leq C$ is enough because the sum runs over almost N^2 elements. Instead we use the *Ward-identity*,

$$\sum_{i\neq k,l}\big|G^{(kl)}_{ij}\big|^2 \;=\; \frac{\operatorname{Im} G^{(kl)}_{jj}}{\operatorname{Im} z}\,,$$

which is a simple consequence of the self-adjointness of $\mathbf{H}$. Therefore, we may estimate the expression on the right hand side of (4.8) further and arrive at

$$\Big|\sum_{i,j\neq k,l} h_{ki}G^{(kl)}_{ij}h_{jl}\Big| \;\prec\; \frac{1}{\sqrt{N\operatorname{Im} z}}\Big(\frac{1}{N}\sum_{j\neq k,l}\operatorname{Im} G^{(kl)}_{jj}\Big)^{1/2}.$$

The term in parenthesis on the right hand side is an average over imaginary parts of diagonal resolvent entries, which are estimated using the a priori bound (4.7). This justifies the claim about the off-diagonal resolvent entries in (4.6).

We will now use (4.4) to derive the equation for the diagonal resolvent entries in (4.6). The sum on the right hand side of (4.4) is split into a diagonal and an off-diagonal contribution. The off-diagonal part is handled by a large deviation estimate in the same way as was done for the off-diagonal resolvent elements, G_{kl}. Thus, we find

$$\Big|\sum_{i,j\neq k,\, i\neq j} h_{ki}G^{(k)}_{ij}h_{jk}\Big| \;\prec\; \frac{1}{\sqrt{N\operatorname{Im} z}}\,.$$

The diagonal contribution is of the form $\sum_{i\neq k,} G^{(k)}_{ii}h_{ki}^2$, where the coefficients, given by the diagonal entries of the resolvent $\mathbf{G}^{(k)}$, are independent of the family of non-negative independent random variables $(h_{ki}^2)_i$. By the law of large numbers, this sum is close to its expectation over the k-th row of $\mathbf{H}$. More precisely, we have the large deviation result,

$$\Big|\sum_{i\neq k,} G^{(k)}_{ii}h_{ki}^2 - \sum_{i\neq k,} G^{(k)}_{ii}\,\mathbb{E}\,h_{ki}^2\Big| \;\prec\; \frac{1}{\sqrt{N}}\,.$$

Here, we used the bound on the coefficients, provided by the a priori bound (4.7). Altogether, we have

$$\Big|\frac{1}{G_{kk}} + z + \sum_{i \neq k,} G_{ii}^{(k)} \mathbb{E}\, h_{ki}^2\Big| \prec \frac{1}{\sqrt{N \operatorname{Im} z}}\,.$$

This finishes our heuristic proof of (4.6), except that we still have to get rid of the upper index k of the resolvent elements in the sum. We will not justify this step here, but removing a single row and column from $\mathbf{H}$ changes the resolvent entries only by a small error of size $(N \operatorname{Im} z)^{-1}$.

4.2 Stability of QVE and local law in the bulk

In Section 4.1 we derived the *perturbed discrete QVE* for the diagonal resolvent entries of the form

$$-\frac{1}{G_{kk}(z)} \;=\; z + \frac{1}{N}\sum_{j=1}^{N} s\Big(\frac{k}{N}, \frac{j}{N}\Big) G_{jj}(z) + d_k(z)\,, \tag{4.9}$$

where the random error d_k satisfies the large deviation bound

$$|d_k(z)| \prec \frac{1}{\sqrt{N \operatorname{Im} z}}\,. \tag{4.10}$$

We view (4.9) as a perturbation of the QVE (2.1). The first statement, (2.2), of the local law, Theorem 2.1, is then a consequence of the stability of the QVE under small perturbations. More generally, we study the *perturbed QVE*

$$-\frac{1}{m_d(x;z)} \;=\; z + \int_0^1 s(x,y)\, m_d(y;z)\mathrm{d}y + d(x;z)\,, \tag{4.11}$$

with a general perturbation $d : [0,1] \times \mathbb{H} \to \mathbb{C}$. Here the function d is considered to contain the random error d_k from (4.9), as well as the error made by replacing the Riemann-sum on the right hand side of (4.9) by an integral.

The spectral parameter z is usually considered fixed. We will therefore often not write the dependence of m and related quantities on z explicitly. It is important in our analysis that all bounds are uniform in z. For example, it is easy to read off from the QVE that its solution satisfies the trivial bound $|m(x;z)| \leq (\operatorname{Im} z)^{-1}$. In fact, most of our arguments are very simple if z stays away from the real line, say $\operatorname{Im} z \geq 1$. But the bound becomes useless as z approaches the real line, where the resolvent $\mathbf{G}(z)$ encodes the most detailed information about the spectrum of $\mathbf{H}$. Thus, we will establish bounds that do not decay as $\operatorname{Im} z \downarrow 0$.

In general we cannot expect equation (4.11) to have a unique solution, let alone m_d to depend smoothly on the perturbation d. Nevertheless, for the purpose of this presentation we will assume that $d(\,\cdot\,;z) \mapsto m_d(\,\cdot\,;z)$ is Fréchet-differentiable as a map from the space of continuous functions, $\mathrm{C}[0,1]$, to itself in such a way that $m_0 = m$ is the unique solution of (2.1). In fact, the following heuristic analysis can be made rigorous and shows that for any fixed $z \in \mathbb{H}$ the equation (4.11) has a unique bounded solution $m_d(\,\cdot\,;z)$ that is differentiable in the perturbation contained in a sufficiently small neighbourhood of zero in $\mathrm{C}[0,1]$. The size of this neighbourhood depends on $\int \operatorname{Im} m(x;z)\mathrm{d}x$ and will remain finite even as z approaches the real axis, $z \to \tau$, if the density of states is non-vanishing, i.e., $\rho(\tau) > 0$. This fact is not used in the detailed analysis about stability of the QVE carried out in Part II. There, we simply assume that (4.11) has a solution m_d and estimate the difference between m and m_d directly without using any differential calculus. However, the use of the functional derivative, Dm_d, of the map $d(\,\cdot\,;z) \mapsto m_d(\,\cdot\,;z)$ allows a more transparent presentation and we will use it here for that reason.

Let S be the integral operator on $\mathrm{C}[0,1]$ with kernel $s(x,y)$. Then by differentiating (4.11) with respect to the perturbation d, we find a formula for the linear operator, Dm_d, namely

$$(1-m_d^2 S)Dm_d\, w \,=\, m_d^2 w\,, \qquad w\in \mathrm{C}[0,1]\,. \tag{4.12}$$

Solving this equation for $Dm_d\, w$ requires inverting the linear operator $1-m_d^2 S$. This can be done at $d=0$ by making use of the average imaginary part of m to bound the norm of the inverse of $1-m^2 S$. Here, we will simply state the result of this analysis, summarised in Lemma 8.8 of Part II. For a detailed analysis of this fact we refer to Section 12.5 in Part II. The lemma implies the bound

$$\big\|(1-m(\,\cdot\,;z)^2 S)^{-1}\big\| \,\le\, \frac{C}{\int \mathrm{Im}\, m(x;z)\mathrm{d}x}\,,$$

where C is a generic positive constant. In Section 4.3 we will see that the solution of the QVE is uniformly bounded. Thus, with the bound on $(1-m^2S)^{-1}$ equation (4.12) can be solved for Dm_d and we find the bound

$$\|Dm_d(\,\cdot\,;z)\|\,\big|_{d=0} \,\le\, \frac{C}{\int \mathrm{Im}\, m(x;z)\mathrm{d}x}\,. \tag{4.13}$$

This shows that the QVE is stable with respect to small perturbations wherever the density of states is bounded away from zero (cf. Definition 2.3).

The diagonal resolvent elements satisfy the perturbed QVE (4.9) and for the random error term we have the bound (4.10). Thus, the stability of the QVE implies that in the bulk, i.e., for $\tau\in\mathbb{R}$ with $\rho(\tau)$ bounded away from zero, the diagonal resolvent entries converge to the solution of the QVE,

$$|G_{kk}(\tau+\mathrm{i}\eta)-m(k/N;\tau+\mathrm{i}\eta)| \,\prec\, \frac{1}{\sqrt{N\eta}}\,.$$

4.3 Bounds on solution of QVE

Here, we explain the basic idea behind how to establish a uniform bound on the solution, m, of the QVE. This bound is needed for the stability analysis of the QVE. In fact, we made use of it already in Sections 4.1 and 4.2. Also for proving regularity properties of m this bound is essential.

The proof of the uniform bound on m requires two different arguments in two different regimes. Away from $z=0$ the bound stems from a Perron-Frobenius argument involving the integral operator F defined by

$$(F(z)w)(x) \,:=\, |m(x;z)|\int_0^1 s(x,y)|m(y;z)|w(y)\mathrm{d}y\,, \qquad w\in\mathrm{C}[0,1]\,. \tag{4.14}$$

In a neighbourhood of $z=0$ on the other hand we make use of the positivity of the kernel $s(x,y)$. Altogether, we will see that the solution of the QVE is uniformly bounded on the entire complex upper half plane under our assumptions on the kernel s (cf. assumption 4.),

$$\sup_{z\in\mathbb{H}}\sup_{x\in[0,1]}|m(x;z)| \,<\, \infty\,.$$

Let us first show the mechanism used to for the uniform bound away from $z=0$. Since m has values in $\mathbb{H}$, i.e., the imaginary part of the solution of the QVE is positive, and $s(x,y)>0$, the integral kernel defining the operator F (cf. (4.14)) is positive everywhere. The operator has a unique Perron-Frobenius eigenvalue-eigenvector pair,

$$Ff \,=\, \lambda f \quad \text{with} \quad f(x;z)>0 \quad \text{and} \quad \lambda(z) \,=\, \sup\mathrm{Spec}(F(z))\,.$$

For more details on this fact we refer to Proposition 8.2 in Part II. The operator F appears naturally when we take the imaginary part on both sides of the QVE and multiply them with $|m|$. In this way we get

$$\frac{\operatorname{Im} m(x;z)}{|m(x;z)|} \,=\, |m(x;z)|\operatorname{Im} z + |m(x;z)| \int_0^1 s(x,y)|m(y;z)| \frac{\operatorname{Im} m(y;z)}{|m(y;z)|} \mathrm{d}y \,.$$

On the right hand side we identify the operator F. We estimate this pointwise from below by discarding the imaginary part of z,

$$\frac{\operatorname{Im} m}{|m|} \,=\, |m| \operatorname{Im} z + F\Big(\frac{\operatorname{Im} m}{|m|}\Big) \,\geq\, F\Big(\frac{\operatorname{Im} m}{|m|}\Big) \,.$$

By multiplying both sides of this inequality by the positive Perron-Frobenius eigenfunction, f, of F, integrating over x and using the symmetry of the integral kernel of F we see that

$$\int_0^1 f(x;z) \frac{\operatorname{Im} m(x;z)}{|m(x;z)|} \mathrm{d}x \,\geq\, \lambda(z) \int_0^1 f(x;z) \frac{\operatorname{Im} m(x;z)}{|m(x;z)|} \mathrm{d}x \,.$$

In short, we find that $\lambda(z) \leq 1$. Since F is an integral operator with a positive symmetric integral kernel, this spectral information implies a bound on the operator norm of F on $\mathrm{L}^2[0,1]$ as well. In fact, λ coincides with this norm. Using this information we find a bound on the $\mathrm{L}^2[0,1]$-norm of m as follows. We write the QVE in the form

$$z m(x;z) \,=\, 1 - m(x;z) \int_0^1 s(x,y) m(y;z) \,. \tag{4.15}$$

Then we take the $\mathrm{L}^2[0,1]$-norm over x on both sides and arrive at

$$|z| \Big(\int_0^1 |m(x;z)|^2 \mathrm{d}x \Big)^{1/2} \,\leq\, 1 + \lambda(z) \,\leq\, 2 \,.$$

Therefore, $m(\,\cdot\,, z)$ is bounded in $\mathrm{L}^2[0,1]$ away from $z = 0$.

At this stage we use the regularity of m in the variable x to improve the $\mathrm{L}^2[0,1]$-bound to a uniform bound. Indeed, it is easy to see from the QVE that the 1/2-Hölder-continuity of the integral kernel $s(x,y)$ (cf. assumption 4.) also implies 1/2-Hölder-continuity of m in the x-variable. Using the argument carried out in Lemma 9.4 of Part II shows that boundedness of m in $\mathrm{L}^2[0,1]$ implies a bound also in maximum norm because of this regularity in x.

The uniform bound in the vicinity of $z = 0$ requires a completely different proof. This proof is split into three steps. The first step is to show a bound in $\mathrm{L}^1[0,1]$ on the imaginary axis close to $z = 0$,

$$\sup_{\eta \in (0,1]} \int_0^1 |m(x; \mathrm{i}\eta)| \mathrm{d}x \,<\, \infty \,.$$

In the second step this bound is improved to a uniform bound. Finally, a perturbation argument in the third step extends the uniform bound to a whole neighbourhood of $z \in \mathrm{i}(0,1]$. Here, we will only demonstrate how to do the first two steps.

It is a consequence of the symmetry $\overline{m(x;z)} \,=\, -m(x;-\overline{z})$ of the QVE that its solution is purely imaginary on the imaginary axis. In particular, it suffices to show the $\mathrm{L}^1[0,1]$-bound and the uniform bound for $\operatorname{Im} m$. On the imaginary line the QVE simplifies to

$$\frac{1}{\operatorname{Im} m(x; \mathrm{i}\eta)} \,=\, \eta + \int_0^1 s(x,y) \operatorname{Im} m(y; \mathrm{i}\eta) \mathrm{d}y \,, \qquad \eta \in (0,1] \,,\; x \in [0,1] \,. \tag{4.16}$$

Since the imaginary part of m and $s(x,y)$ are positive this implies that

$$1 \,\geq\, \min_{u,v} s(u,v)\,\mathrm{Im}\, m(x;\mathrm{i}\eta) \int_0^1 \mathrm{Im}\, m(y;\mathrm{i}\eta)\mathrm{d}y\,.$$

Integrating both side over x shows an $\mathrm{L}^1[0,1]$-bound for $\mathrm{Im}\, m$. We remark that in the case that is presented in Part II proving this bound is more involved. There, the assumption that s is positive is replaced by the more general notion of *block fully indecomposability* and (4.16) is reformulated as a minimisation problem. Analysing the properties of the corresponding minimiser shows the $\mathrm{L}^1[0,1]$-bound in the general case.

The $\mathrm{L}^1[0,1]$-bound on $\mathrm{Im}\, m$ is now improved to a uniform bound by a simple argument. Since the integral kernel $s(x,y)$ is bounded, (4.16) and the $\mathrm{L}^1[0,1]$-bound imply a pointwise lower bound on $\mathrm{Im}\, m(x;\mathrm{i}\eta)$ for $\eta \in (0,1]$. The kernel, $s(x,y)$, is positive and continuous and thus bounded away from zero. We infer that the right hand side of (4.16) is bounded away from zero as well. This implies a pointwise upper bound on the imaginary part of m.

In the third step the uniform bound on $\mathrm{Im}\, m(x;\mathrm{i}\eta) = |m(x;\mathrm{i}\eta)|$ for $\eta \in (0,1]$ is extended by a perturbation argument to a uniform bound,

$$\sup_{\eta\in(0,1]}\sup_{|\tau|\leq\varepsilon} |m(x;\tau+\mathrm{i}\eta)| \,<\, \infty\,,$$

for some small $\varepsilon > 0$. We will not carry out this step here.

Altogether, the considerations above show that m is uniformly bounded everywhere. It is an immediate consequence of m satisfying the QVE that a uniform upper bound on m also implies a pointwise lower bound. Indeed, taking into account the large-$|z|$-behaviour that can be read off from the QVE as well, we get

$$\frac{c}{1+|z|} \,\leq\, |m(x;z)| \,\leq\, \frac{C}{1+|z|}\,, \qquad x \in [0,1]\,,\; z \in \mathbb{H}\,,$$

for some positive constants c and C.

4.4 Quadratic equation at the edge

Close to the edges, β and $-\beta$, of the support of the density of states (cf. Theorem 2.2), the stability argument from Section 4.2 breaks down. The density of states approaches zero as z converges to these points and the bound (4.13) on the derivative of m with respect to the perturbation d of the QVE becomes ineffective. The blow-up of the right hand side of this bound has its root in the fact that at the edges the operator $1-m^2S$ is not invertible anymore. Instead, it has a one dimensional kernel

$$\ker(1-m^2S)|_{z=\pm\beta} \,=\, \mathbb{C}e\,, \qquad e \,:=\, \lim_{\eta\downarrow 0} \frac{\mathrm{Im}\, m}{\|\mathrm{Im}\, m\|}\bigg|_{z=\pm\beta+\mathrm{i}\eta}\,. \tag{4.17}$$

That the components of $\mathrm{Im}\, m$ are all mutually comparable in size (cf. Theorem 6.2 of Part II) allows us to take the limit in the definition of e, even though $\mathrm{Im}\, m$ converges to zero as z approaches the edge.

The non-invertibility of $1-m^2S$ in this direction can be seen by looking at the imaginary part of the QVE,

$$\mathrm{Im}\, m \,=\, |m|^2\,\mathrm{Im}\, z + |m|^2 S\,\mathrm{Im}\, m\,.$$

We divide both sides by $\|\mathrm{Im}\, m\|$ and use that as z approaches the edges $\mathrm{Im}\, z/\|\mathrm{Im}\, m\|$ converges to zero. A posteriori this can be verified easily by using the Stieltjes transform representation of m from Theorem 2.2. Since $\operatorname{Im} m = 0$ and $|m|^2 = m^2$ at the edges, we confirm that $\operatorname{Im} m / \|\operatorname{Im} m\|$ is mapped to zero by $1 - m^2 S$ as $z \to \pm\beta$.

In order to carry out a stability analysis of the QVE in the vicinity of $z = \pm\beta$, we split off that bad direction, in which the operator $1 - m^2 S$ is not invertible. Essentially this means that we write the difference between the solution m_d of the perturbed QVE (4.11) and m as

$$m_d(x; \pm\beta) - m(x; \pm\beta) \;=\; \Theta_d \, e(x) + r_d(x) \,.$$

Here, the remainder, r_d, is the spectral projection of $m_d - m$ to the spectral subspace of $1 - m^2 S$ associated to $\operatorname{Spec}(1 - m^2 S) \setminus \{0\}$. Since this operator is not symmetric, e and r_d are not orthogonal to each other. Nevertheless, the operator $1 - m^2 S$ is invertible on the spectral subspace, where the bad direction, e, is taken out. Therefore, the remainder satisfies the bound $\|r_d\| \le C\|d\|$.

The scalar, Θ_d, on the other hand, measures the degree to which $m_d - m$ points in the bad direction. By using the perturbed QVE (4.11) for m_d and (2.1) for m, we derive an equation for this quantity. This is carried out in detail in Proposition 11.2 of Part II. In our situation this equation is of the form

$$\Theta_d^2 \mp \int k_1(x) d(x; \pm\beta) \mathrm{d}x \;=\; \mathcal{O}\Big(|\Theta_d|^3 + |\Theta_d| \Big| \int k_2(x) d(x; \pm\beta) \mathrm{d}x \Big| + \|d\|^2 \Big) \,. \tag{4.18}$$

Here, k_1 and k_2 are continuous functions, depending only of m. They do not depend on the perturbation d. Furthermore, k_1 takes positive values.

The quadratic instability at the edge, reflected by equation (4.18), implies the bound $|\Theta_d| \le \|d\|^{1/2}$ and thus

$$\|m_d(\cdot\,; \pm\beta) - m(\cdot\,; \pm\beta)\| \;\le\; \|d\|^{1/2}. \tag{4.19}$$

This is in contrast to the linear stability of the QVE in the bulk, i.e., where the density of states is bounded away from zero. There the difference between m_d and m was bounded by $\|d\|$ instead of its square root.

In our simplified set-up (4.18) is quadratic in Θ_d in the sense that the coefficient of the second order term in Θ_d does not vanish (cf. Proposition 11.2 and Section 12.5 in Part II). In general, this will not be the case and we need to take the third order term in Θ_d into account. This term is now treated as part of the error on the right hand side. Once the third order term in Θ_d has to be considered, equation (4.18) becomes cubic, i.e., it is of the form

$$\pi_3 \Theta_d^3 + \pi_2 \Theta_d^2 + \pi_1 \Theta \;=\; \mathcal{O}(\|d\|) \,, \tag{4.20}$$

with some coefficients π_k that are explicitly expressed in terms of the solution m of the QVE. Higher order terms in Θ_d are never relevant though, because the coefficients of Θ_d^2 and Θ_d^3 never vanish at the same time, $|\pi_3| + |\pi_2| \ge c > 0$. For more details see Section 11.2 of Part II. The fact that the equation for Θ_d is cubic in the setup of Part II and Part III causes additional technical difficulties there because the roots of the cubic equation have to be followed for the stability analysis.

4.5 Edge shape

At the edges of its support the density of states grows like a square root. This behaviour can be inferred from the quadratic equation (4.18) for the scalar quantity Θ_d that reflects the size the leading order term of the difference, $m_d - m$, between the solutions of (4.11) and (2.1) evaluated at the edges, respectively. Indeed, if we choose the perturbation d to be the constant function (in the x-variable), $d(x;\pm\beta) := \omega$, for some $\omega \in \mathbb{R}$, then $m_d(x;\pm\beta) = m(x;\pm\beta+\omega)$. Thus, following the arguments from Section 4.4, we find

$$m(x;\pm\beta+\omega) - m(x;\pm\beta) \;=\; \Theta_\omega\, e(x) + \mathcal{O}(|\omega|)\,,$$

where Θ_ω satisfies a quadratic equation of the form

$$\Theta_\omega^2 \mp \kappa\omega \;=\; \mathcal{O}(|\omega|^{3/2})\,.$$

Here, the positive constant κ is given by the integral over the function k_1 from (4.18). This simple quadratic equation for Θ_ω is solved explicitly and its solution has the form

$$\Theta_\omega \;=\; \begin{cases} \mathrm{i}\sqrt{\kappa|\omega|}, & \text{if } \omega > 0, z = -\beta \text{ or } \omega < 0, z = \beta\,, \\ \sqrt{\kappa|\omega|}, & \text{if } \omega < 0, z = -\beta \text{ or } \omega > 0, z = \beta\,. \end{cases}$$

In particular, we find the square root behaviour of the imaginary part of m at the edges,

$$\operatorname{Im} m(x;\pm\beta\mp\omega) - \operatorname{Im} m(x;\pm\beta) \;=\; \sqrt{\kappa\omega} + \mathcal{O}(\omega)\,, \qquad \omega \in (0,\varepsilon)\,.$$

In the general case, which is treated in Part II, equation (4.18) will be cubic in Θ_d (cf. (4.20)). Nevertheless, with the choice of the constant perturbation, $d(x;\pm\beta) := \omega$, this equation is still explicitly solvable for Θ_ω. In contrast to our simplified setting, the solution can have qualitatively different shapes, depending on the coefficients of the cubic equation. These shapes reflect the ways in which the density of states approaches zero and their analysis requires considerable effort in Part II of this work. This approach to zero may happen at the extreme edges of the support of ρ as in our simplified setting. But in general the support may consist of intervals with gaps between them and ρ may also approach zero inside the interior of one of these intervals, leading to a cusp singularity with a cubic root growth on both sides.

4.6 Local law at the edge

The diagonal resolvent elements satisfy the perturbed QVE (4.9) which has a quadratic stability behaviour at the edges (cf. (4.19)). The proof of the local law at the edge of the spectrum, however, requires a more careful use of the quadratic equation for Θ_d than the one leading to (4.19). First, the quadratic equation has to be derived not just at the edges but also inside a small neighbourhood of them, because $G_{kk}(z)$ is compared to $m(x;z)$ and not to $m(x;\pm\beta)$. Then another important tool, the *fluctuation averaging*, is used to improve the bounds on integrals over the error function which already appeared in the quadratic equation at the edges, (4.18). Basically, the fluctuation averaging improves the naive error bound $\|d\| \prec (N\operatorname{Im} z)^{-1/2}$ from (4.10) on an integral against any bounded deterministic function k to

$$\Big|\int k(x)\, d(x;z)\,\mathrm{d}x\Big| \;\prec\; \frac{1}{N\operatorname{Im} z}\,.$$

By solving the quadratic equation for Θ_d we then infer the bound

$$|\Theta_d| \;\prec\; \frac{1}{\sqrt{N\operatorname{Im} z}}\,.$$

Since $G_{kk}(z) = m_d(k/N;z)$ is viewed as the solution of (4.11) this bound implies the local law close to the edge,

$$|G_{kk}(z) - m(k/N;z)| \;\prec\; \frac{1}{\sqrt{N\operatorname{Im} z}}\,.$$

5 Open problems and future projects

The project initiated by [1] and [2] can be extended and continued in several directions. First we mention a few open question within the model introduced in Part III that are not answered by this work. Then we discuss a few possible extension of the model that can be considered a natural continuation of recent developments in random matrix theory.

The formulation of the local law in the general setting, Theorem 15.6, suggests that the speed of convergence of G_{kk} to its limit deteriorates close to *cusp singularities* inside the interior of the support of the density of states. However, this is expected to be an artefact of the existing proof. It is therefore still an open problem to prove the optimal local law at the cusp.

Closely related to the local law at the singularities of the density of states is the question of universality at these points. It is expected that the eigenvalues at the extreme edges as well as the ones at the edges of internal gaps in the support of the density of states follow the universal Tracy-Widom distributions. At the cusp singularities, however, the distribution of the eigenvalues is conjectured to be given by corresponding Pearcey functions [62]. For random matrices with a large GUE component this has been shown by Brézin and Hikami in [17, 18]. However, their method relies on the Itzykson-Zuber integral formula and is thus not easily extendable to the real symmetric and symplectic symmetry class.

In the presentation of our results so far, we used several times that the values of the kernel, s, encoding the variances of the entries of the random matrix, is bounded away from zero. In the models considered in Part II and III this condition is replaced by the more general assumption of *uniform primitivity* (cf. Section 6.1 in Part II). It asserts that the kernel of a finite power of the operator S with kernel s be bounded away from zero. A simple extension of the setting from Part II and III can be obtained by relaxing the assumption of uniform primitivity even further. Of particular importance are random matrices with 2×2-block structure

$$\mathbf{H} = \left(\begin{array}{c|c} \mathbf{0} & \mathbf{A}^* \\ \hline \mathbf{A} & \mathbf{0} \end{array}\right),$$

where the $N \times M$-matrix $\mathbf{A}$ has independent entries. By squaring the eigenvalues of this matrix one obtains the spectrum of the covariance matrices $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$. For the purpose of obtaining local laws, matrices $\mathbf{H}$ of this form can be analysed by following the methods presented in Part II and III of this work. Their stability analysis at the origin $z = 0$, however, requires some extra steps. If we consider the case $N = M$ and subtract from the matrix $\mathbf{A}$ with independent entries a complex multiple of the identity, i.e., $\mathbf{A} \to \mathbf{A} - w\mathbf{1}$, then the matrix $\mathbf{H}$ can be used to study the local law of the non-Hermitian matrix $\mathbf{A}$ itself (see e.g. [60]) by using *Girko's Hermitisation trick*.

We will now briefly discuss a few models that go well beyond the scope of what is presented in this work. Motivated by recent results on random band matrices one may consider self-adjoint band matrices with independent entries (up to symmetry constraints) whose density of states is not the semicircle law. At the moment establishing local laws for band matrices still relies on a special block structure of the matrix and explicit formulas from supersymmetry [55, 56, 54, 7]. Without this structure and the semicircle law in the limit more robust tools may be required. Even showing bulk universality for the matrices considered in Part III but without assuming a non-vanishing variance profile, e.g. with a macroscopic band of size $\sim N$, is an open problem.

Finally, the results on Gaussian random matrices with translation invariant correlation, that are discussed in Section 21 of Part III, have natural extensions in two directions. The first is to eliminate the assumption that the entries are Gaussian. The second is to consider more general correlations without translation invariance. The correlated case is particularly interesting because it provides a random matrix model in which the entries of the resolvent naturally admit a non-trivial profile in which the off-diagonal entries do not vanish in the limit. The same feature is expected for random Schrödinger operators.

Part II

6 Introduction to Part II

In this part we analyse in detail the solution of the system of non-linear equations

$$-\frac{1}{m_i} = z + \sum_{j=1}^{N} S_{ij} m_j \,, \qquad i = 1, \dots, N \,. \tag{6.1}$$

For any fixed z in the complex upper half plane $\mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ the solution is a set of N complex numbers $m_1, \dots, m_N \in \mathbb{H}$. If the matrix $\mathbf{S}$ is stochastic, then the solution is a constant vector, $m_i = m_{\mathrm{sc}}$, with the Stieltjes transform of the semicircle law in each entry. For a general variance matrix S_{ij} the situation is more complicated. The solution m_i to (6.1) genuinely depends on i. The average density, $\frac{1}{\pi N} \sum_i \operatorname{Im} m_i$, is not the semicircle law any more. For an analysis of well-posedness and a few qualitative results about (6.1) and its solution see the work of Helton, Far, and Speicher [43], and Anderson and Zeitouni [5]. We will comment on these previous results in the end of Subsection 6.1. Prior to this work it was not known what kind of density shapes may emerge instead of the familiar semicircle law. Furthermore, in order to use equation (6.1) for the application to random matrices with independent entries and to prove that the diagonal resolvent entries G_{ii} are close to m_i a thorough stability analysis is needed. It is well known that even in the simple scalar case, when $\mathbf{S}$ is stochastic, the stability deteriorates near the endpoints of the semicircle law, i.e., if z is close to ± 2, the points where $m(z)$ has a square-root singularity. For generalised Wigner matrices the proof of the local semicircle law down to the optimal scale [38, 26] heavily relied on the precise understanding of how the deterioration of the stability can be offset by better estimates on the error terms. A similarly accurate stability analysis is necessary to establish local laws for random matrices with general variance matrices $\mathbf{S}$.

In the current part of this work we present general results on the singularities and stability of (6.1). This analysis is of interest in its own right since (6.1) appears in other contexts as well (see the end of the next section). The analysis of the corresponding random matrices will be presented in Part III. The current part is self-contained and random matrices will not appear here. They are mentioned only to motivate the problem.

6.1 Set-up and the main results

We formulate a general continuum version of (6.1), which allows us to treat all dimensions N in a unified manner. We may also think of this continuum limit as the resulting equation when we take the large N-limit of (6.1) and assume that the entries of $\mathbf{S}$ scale like N^{-1}.

We start with an abstract set $\mathfrak{X}$ labelling the matrix elements. We introduce the Banach space

$$\mathscr{B} := \Big\{ w : \mathfrak{X} \to \mathbb{C} : \sup_{x \in \mathfrak{X}} |w_x| < \infty \Big\} \,, \tag{6.2}$$

of bounded complex valued functions on $\mathfrak{X}$, equipped with the norm

$$\|w\|_{\mathscr{B}} := \sup_{x \in \mathfrak{X}} |w_x| \,. \tag{6.3}$$

We will also define the subset,

$$\mathscr{B}_+ := \big\{ w \in \mathscr{B} : \operatorname{Im} w_x > 0 \;\text{ for all } x \in \mathfrak{X} \big\} \,, \tag{6.4}$$

of functions with values in the upper half-plane $\mathbb{H}$.

Let $S : \mathscr{B} \to \mathscr{B}$ be a bounded linear operator. The main object of study of this part of the work is the continuum limitöf (6.1),

$$-\frac{1}{m(z)} = z + Sm(z), \qquad \forall\, z \in \mathbb{H}, \tag{6.5}$$

and its solution $m : \mathbb{H} \to \mathscr{B}_+$. Here we view $m : \mathfrak{X} \times \mathbb{H} \to \mathbb{H}$, $(x, z) \mapsto m_x(z)$ as a function of the two variables x and z, but we will often suppress the x and/or z dependence of m and other related functions. The symbol m will always refer to a solution of (6.5). We will refer to (6.5) as the **Quadratic Vector Equation** (**QVE**).

We assume that $\mathfrak{X}$ is equipped with a probability measure π and a σ-algebra $\mathcal{S}$ such that $(\mathfrak{X}, \mathcal{S}, \pi)$ constitutes a probability space. We will denote the space of measurable functions $u : \mathfrak{X} \to \mathbb{C}$, satisfying $\|u\|_p := (\int_{\mathfrak{X}} |u_x|^p \pi(\mathrm{d}x))^{1/p} < \infty$, as $\mathrm{L}^p = \mathrm{L}^p(\mathfrak{X}; \mathbb{C})$, $p \geq 1$. The usual L^2-inner product, and the averaging are denoted by

$$\langle u, w\rangle := \int_{\mathfrak{X}} \overline{u_x} w_x \pi(\mathrm{d}x) \qquad \text{and} \qquad \langle w\rangle := \langle 1, w\rangle, \qquad u, w \in \mathrm{L}^2. \tag{6.6}$$

For a linear operator A mapping a Banach space X to another Banach space Y we denote the corresponding operator norm by $\|A\|_{X\to Y}$. Finally, if w is a function on $\mathfrak{X}$ and T is a linear operator acting on such functions then $w + T$ denotes the linear operator $u \mapsto wu + Tu$, i.e., we interpret w as a multiplication operator when appropriate.

In the entire Part II we assume that the operator $S : \mathscr{B} \to \mathscr{B}$ in (6.5) satisfies:

A1. *Symmetry*: If $u, w \in \mathscr{B}$ then $\langle u, Sw\rangle = \langle Su, w\rangle$;

A2. *Non-negativity*: If $u \in \mathscr{B}$, with $\inf_x u_x \geq 0$, then $\inf_x (Su)_x \geq 0$;

A3. *Normalisation*: $\|S\|_{\mathscr{B}\to\mathscr{B}} = 1$.

By a simple rescaling, the normalisation condition **A3.** is only for convenience, it merely reflects the boundedness of S. Indeed, if we replace S with λS, for some constant $\lambda > 0$, then the modified QVE is solved by $m^\lambda : \mathbb{H} \to \mathscr{B}$, where $m^\lambda_x(z) := \lambda^{1/2} m_x(\lambda^{1/2} z)$.

Theorem 6.1 (Existence and uniqueness). *Assume S satisfies* **A1-3.** *Then for any $z \in \mathbb{H}$, the equation*

$$-\frac{1}{m} = z + Sm, \tag{6.7}$$

has a unique solution $m = m(z) \in \mathscr{B}_+$. These solutions form an analytic function $z \mapsto m(z)$ from $\mathbb{H}$ to $\mathscr{B}_+$. Moreover, for each $x \in \mathfrak{X}$ there exists a measure v_x on $\mathbb{R}$ with its support contained in $[-2, 2]$ and $v_x(\mathbb{R}) = \pi$, such that

$$m_x(z) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{v_x(\mathrm{d}\tau)}{\tau - z}, \qquad \forall\, z \in \mathbb{H}. \tag{6.8}$$

The measures v_x constitute a measurable function $v := (x \mapsto v_x) : \mathfrak{X} \to \mathcal{M}(\mathbb{R})$, where $\mathcal{M}(\mathbb{R})$ denotes the space of finite Borel measures on $\mathbb{R}$ equipped with the weak topology.

Furthermore, the solution $m(z)$ satisfies the L^2-bound

$$\|m(z)\|_2 \leq \frac{2}{|z|}, \qquad \forall\, z \in \mathbb{H}. \tag{6.9}$$

The existence and uniqueness part follows from the main result of [43]. We remark that if the solution space $\mathscr{B}_+$ is replaced by $\mathscr{B}$, then the equation (6.7) in general may have multiple, even infinitely many, solutions. Since v generates the solution m through (6.8) we call the x-dependent family of measures $v = (v_x)$ the **generating measure**. In order to obtain results beyond the existence and uniqueness we will generally assume that S has the following additional properties:

A4. *Smoothing*: S can be extended to a bounded operator from L^2 to $\mathscr{B}$, i.e., $\|S\|_{\mathrm{L}^2\to\mathscr{B}} < \infty$, and is represented by a symmetric non-negative measurable kernel $S : \mathfrak{X}^2 \to [0,\infty)$.

A5. *Uniform primitivity*: There exist a power $L \in \mathbb{N}$, and a constant $\rho > 0$, such that

$$u \in \mathscr{B}\,,\ u \geq 0 \quad \Longrightarrow \quad (S^L u)_x \,\geq\, \rho \int_{\mathfrak{X}} u_y \,\pi(\mathrm{d}y)\,, \qquad \forall\, x \in \mathfrak{X}\,. \tag{6.10}$$

The finiteness of the norm $\|S\|_{\mathrm{L}^2\to\mathscr{B}}$ in condition **A4.** means that the integral kernel S_{xy} representing the operator S satisfies

$$\|S\|_{\mathrm{L}^2\to\mathscr{B}} \;=\; \sup_{x\in\mathfrak{X}}\biggl(\,\int_{\mathfrak{X}} (S_{xy})^2 \pi(\mathrm{d}y)\biggr)^{1/2} \;<\; \infty\,. \tag{6.11}$$

In particular, S is a Hilbert-Schmidt operator on L^2. The condition **A5.** is an effective lower bound on the coupling between the components m_x in the QVE. In the context of symmetric matrices this property is known as *primitivity* - hence our terminology.

For brevity we introduce the concept of **comparison relations**: If $\varphi = \varphi(u)$ and $\psi = \psi(u)$ are non-negative functions on some set U, then the notation $\varphi \lesssim \psi$, or equivalently, $\psi \gtrsim \varphi$, means that there exists a constant $0 < C < \infty$ such that $\varphi(u) \leq C\psi(u)$ for all $u \in U$. If $\psi \lesssim \varphi \lesssim \psi$ then we write $\varphi \sim \psi$, and say that φ and ψ are **comparable**. Furthermore, we use $\psi = \phi + \mathcal{O}(\xi)$ as a shorthand for $|\psi - \phi| \lesssim |\xi|$, where $\xi = \xi(u) \in \mathbb{C}$. When the implicit constants C in the comparison relations depend on some parameters Λ we say that the **comparison relations depend on** Λ. Typically, Λ consists of constants appearing in the hypotheses and we refer to them as **model parameters**.

For any $I \subseteq \mathbb{R}$ we introduce the seminorm on functions $w : \mathbb{H} \to \mathscr{B}$:

$$|\!|\!| w |\!|\!|_I \;:=\; \sup\bigl\{\, \|w(z)\|_{\mathscr{B}} : \mathrm{Re}\, z \in I\,,\ \mathrm{Im}\, z \in (0,\infty)\bigr\}\,. \tag{6.12}$$

THEOREM 6.2 (Regularity of generating density). *Suppose S satisfies the assumptions* **A1-A5.**, *and the solution m of (6.5) is uniformly bounded everywhere, i.e.,*

$$|\!|\!| m |\!|\!|_{\mathbb{R}} \,\leq\, \Phi\,,$$

for some constant $\Phi < \infty$. Then the following hold true:

(i) *The generating measure has a Lebesgue density (also denoted by v), i.e., $v_x(\mathrm{d}\tau) = v_x(\tau)\mathrm{d}\tau$. The components of the* **generating density** *are symmetric functions of τ and comparable, i.e.,*

$$v_x(-\tau) = v_x(\tau) \qquad \textit{and} \qquad v_x(\tau) \,\sim\, v_y(\tau)\,, \quad \forall\, \tau \in \mathbb{R}\,,\, \forall\, x, y \in \mathfrak{X}\,.$$

In particular, the support of v_x is independent of x, and hence we write $\mathrm{supp}\, v$ *for this common support.*

(ii) *$v(\tau)$ is real analytic in τ, everywhere except at points $\tau \in \operatorname{supp} v$ where $v(\tau) = 0$. More precisely, there exists $C_0 \sim 1$, such that the derivatives are bounded by*

$$\left\|\partial_\tau^k v(\tau)\right\|_{\mathscr{B}} \le k!\left(\frac{C_0}{\langle v(\tau)\rangle^3}\right)^k, \qquad \forall\, k \in \mathbb{N},$$

whenever $\langle v(\tau)\rangle > 0$.

(iii) *The density is uniformly 1/3-Hölder-continuous everywhere, i.e.,*

$$\|v(\tau_2) - v(\tau_1)\|_{\mathscr{B}} \lesssim |\tau_2 - \tau_1|^{1/3}, \qquad \forall\, \tau_1, \tau_2 \in \mathbb{R}.$$

The comparison relations in these statements depend on the model parameters ρ, L, $\|S\|_{\mathrm{L}^2\to\mathscr{B}}$ and Φ.

Here we assumed an a priori uniform bound on $|||m|||_{\mathbb{R}}$. In Theorem 6.8 below we will present sufficient conditions on S to guarantee $|||m|||_{\mathbb{R}} < \infty$. We remark that without such a bound a regularity result weaker than Theorem 6.2 can still be proven (cf. Corollary 10.3).

The next theorem describes the behaviour of the generating density in the regime where the average generating density, $\langle v\rangle$, is small. We start with defining two universal shape functions.

Definition 6.3 (Shape functions). *Define $\Psi_{\mathrm{edge}} : [0,\infty) \to [0,\infty)$, and $\Psi_{\min} : \mathbb{R} \to [0,\infty)$, by*

$$\Psi_{\mathrm{edge}}(\lambda) := \frac{\sqrt{(1+\lambda)\,\lambda}}{\left(1+2\lambda + 2\sqrt{(1+\lambda)\,\lambda}\,\right)^{2/3} + \left(1+2\lambda - 2\sqrt{(1+\lambda)\,\lambda}\,\right)^{2/3} + 1}, \tag{6.13a}$$

$$\Psi_{\min}(\lambda) := \frac{\sqrt{1+\lambda^2}}{(\sqrt{1+\lambda^2}+\lambda)^{2/3} + (\sqrt{1+\lambda^2}-\lambda)^{2/3} - 1} - 1. \tag{6.13b}$$

As the names suggest, the appropriately rescaled versions of Ψ_{edge} and $\Psi_{\min}$ will describe how $v_x(\tau_0+\omega)$ behaves when τ_0 is an edge of $\operatorname{supp} v$ and when τ_0 is a local minimum of $\langle v\rangle$ with $\langle v(\tau_0)\rangle > 0$ sufficiently small, respectively.

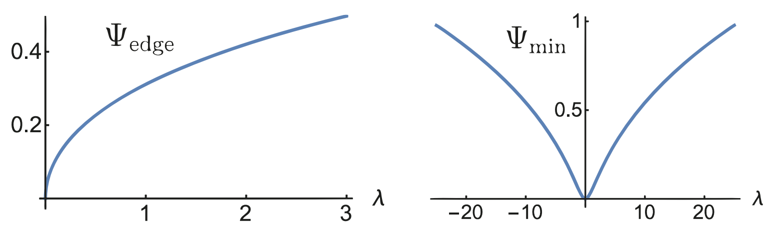

Figure 6.1: The two shape functions Ψ_{edge} and $\Psi_{\min}$.

The next theorem is our main result. Together with Theorem 6.2 it classifies the behaviour of the generating density of a general bounded solution of the QVE.

Theorem 6.4 (Shape of generating density near its small values). *Assume* **A1-A5.**, *and*

$$|||m|||_{\mathbb{R}} \le \Phi,$$

for some $\Phi < \infty$. Then the support of the generating measure consists of $K' \sim 1$ disjoint intervals, i.e.,

$$\operatorname{supp} v = \bigcup_{i=1}^{K'} [\alpha_i, \beta_i], \quad \textit{where} \quad \beta_i - \alpha_i \sim 1, \quad \textit{and} \quad \alpha_i < \beta_i < \alpha_{i+1}. \tag{6.14}$$

Moreover, for all $\varepsilon > 0$ there exist $K'' = K''(\varepsilon) \sim 1$ points $\gamma_1, \dots, \gamma_{K''} \in \operatorname{supp} v$ such that $\tau \mapsto \langle v(\tau)\rangle$ has a local minimum at $\tau = \gamma_k$ with $\langle v(\gamma_k)\rangle \le \varepsilon$, $1 \le k \le K''$. These minima are well separated from each other and from the edges, i.e.,

$$|\gamma_i - \gamma_j| \sim 1, \quad \forall\, i \ne j, \quad \textit{and} \quad |\gamma_i - \alpha_j| \sim 1, \quad |\gamma_i - \beta_j| \sim 1, \qquad \forall\, i, j. \tag{6.15}$$

Let $\mathbb{M}$ denote the set of edges and these internal local minima:

$$\mathbb{M} := \{\alpha_i\} \cup \{\beta_j\} \cup \{\gamma_k\}, \tag{6.16}$$

then small neighbourhoods of $\mathbb{M}$ cover the entire domain where $0 < \langle v \rangle \le \varepsilon$, i.e., there exists $C \sim 1$ such that

$$\begin{aligned} &\big\{\tau \in \operatorname{supp} v : \langle v(\tau)\rangle \le \varepsilon\big\} \\ &\quad\subseteq \bigcup_i [\alpha_i, \alpha_i + C\varepsilon^2] \,\cup\, \bigcup_j [\beta_j - C\varepsilon^2, \beta_j] \,\cup\, \bigcup_k [\gamma_k - C\varepsilon^3, \gamma_k + C\varepsilon^3]\,. \end{aligned} \tag{6.17}$$

The generating density is described by expansions around the points of $\mathbb{M}$, i.e. for any $\tau_0 \in \mathbb{M}$ we have

$$v_x(\tau_0 + \omega) \;=\; v_x(\tau_0) \,+\, h_x \Psi(\omega) \,+\, \mathcal{O}\Big(v_x(\tau_0)^2 + \Psi(\omega)^2\Big), \qquad \omega \in I, \tag{6.18}$$

where $h_x \sim 1$ depends on τ_0. The interval $I = I(\tau_0)$ and the function $\Psi : I \to [0,\infty)$ depend only on the type of τ_0 according to the following list:

- *Left edge: If $\tau_0 = \alpha_i$ then (6.18) holds with $v_x(\tau_0) = 0$, $I = [0,\infty)$, and*

$$\Psi(\omega) \;=\; (\alpha_i - \beta_{i-1})^{1/3}\, \Psi_{\mathrm{edge}}\bigg(\frac{\omega}{\alpha_i - \beta_{i-1}}\bigg), \tag{6.19a}$$

  with the convention $\beta_0 - \alpha_1 = 1$.

- *Right edge: If $\tau_0 = \beta_j$ then (6.18) holds with $v_x(\tau_0) = 0$, $I = (-\infty, 0]$, and*

$$\Psi(\omega) \;=\; (\alpha_{j+1} - \beta_j)^{1/3}\, \Psi_{\mathrm{edge}}\bigg(\frac{-\,\omega}{\alpha_{j+1} - \beta_j}\bigg), \tag{6.19b}$$

  with the convention $\alpha_{K'+1} - \beta_{K'} = 1$.

- *Minimum: If $\tau_0 = \gamma_k$ then (6.18) holds with $I = \mathbb{R}$, and*

$$\Psi(\omega) \;=\; \rho_k\, \Psi_{\min}\bigg(\frac{\omega}{\rho_k^3}\bigg), \quad \text{where} \quad \rho_k \;\sim\; \langle v(\gamma_k)\rangle\,. \tag{6.19c}$$

The comparison relations depend on the model parameters ρ, L, $\|S\|_{\mathrm{L}^2\to\mathscr{B}}$ and Φ.

Figure 6.2 shows an average generating measure which exhibits each of the possible singularities described by (6.18) and (6.19).

Note that the expansions (6.18) become useful for the non-zero minima γ_k only when $\varepsilon > 0$ is chosen to be so small that the term $h_x\Psi(\omega)$ dominates $v_x(\tau_0)^2$, which itself is smaller than ε^2.

Remark 6.5 (Qualitative behaviour near minima). *The edge shape function $\Delta^{1/3}\Psi_{\mathrm{edge}}(\omega/\Delta)$ interpolates between a square root and a cubic root growth with the switch in the growth rate taking place when its argument becomes of the size Δ. Similarly, the function $\rho\,\Psi_{\min}(\omega/\rho^3)$ can be seen as a cubic root cusp $\omega \mapsto |\omega|^{1/3}$ regularised at scale ρ^3.*

We point out that the set of points $\{\gamma_1, \dots, \gamma_{K''}\}$ in Theorem 6.4 do not necessary include every local minimum of $\langle v \rangle$ below the threshold ε. However $\mathbb{M}$ contains a nearby representative of all such local minima in the following sense:

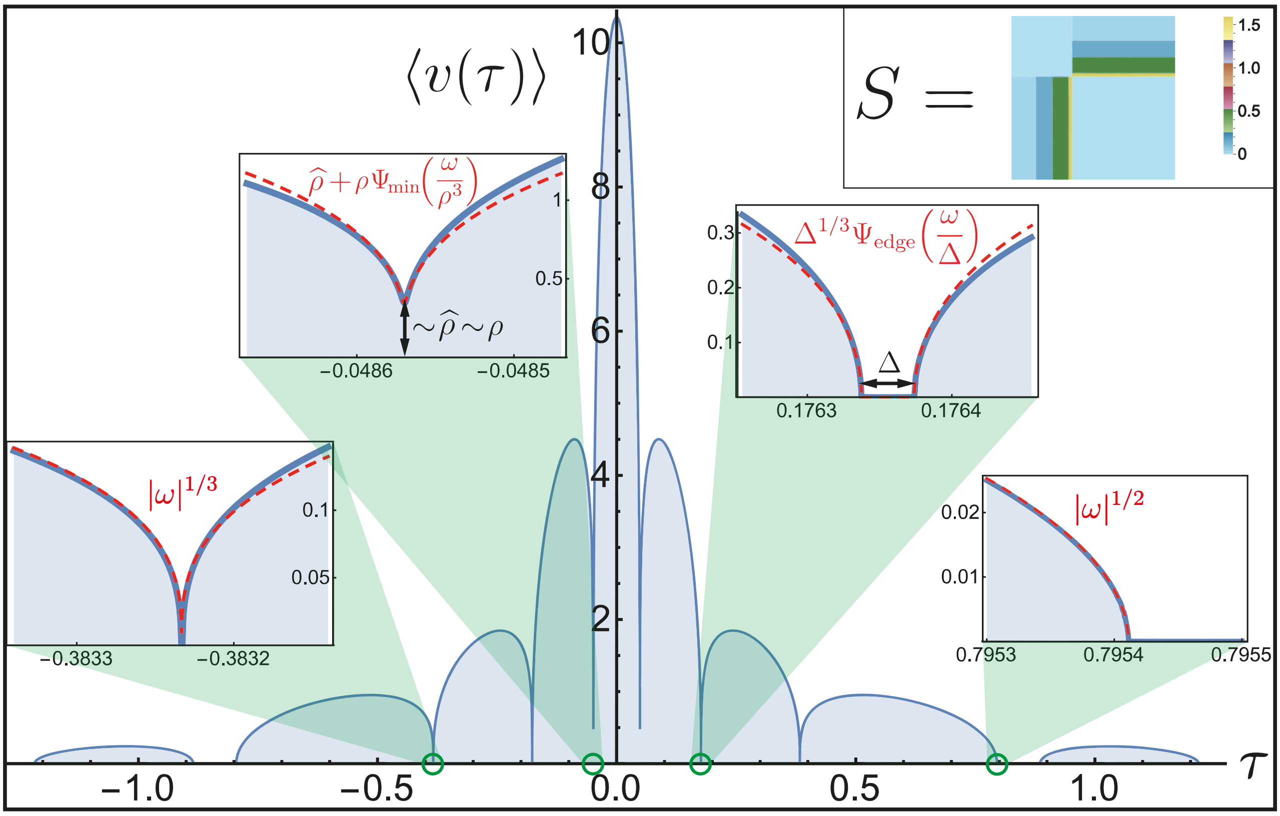

Figure 6.2: Average generating measure $\langle v\rangle$ for the given kernel S. All the possible shapes appear in this example.

Remark 6.6 (Choice of non-zero minima). *The local minima of $\langle v\rangle$ which are not edges of $\operatorname{supp} v$ are either tightly clustered or well separated from each other in the following sense: If $\gamma,\gamma'\in\operatorname{supp} v\backslash\partial\operatorname{supp} v$ are two local minima of $\langle v\rangle$ then either*

$$|\gamma-\gamma'|\lesssim\min\{\langle v(\gamma)\rangle,\langle v(\gamma')\rangle\}^4\,,\qquad or\qquad |\gamma-\gamma'|\sim 1\,.$$

Picking a single representative γ_k from each cluster of minima satisfying $\langle v\rangle\le\varepsilon$ yields a set $\mathbb{M}$ (cf. (6.16)*) for which Theorem 6.4 holds. In particular, the choice of $\mathbb{M}$ is not unique.*

We will now discuss two sufficient and checkable conditions that together with **A1-5.** imply $\|\!|m|\!\|_{\mathbb{R}}<\infty$, a key input of Theorems 6.2 and 6.4. The first one involves a regularity assumption on the family of row functions, or simply **rows**, of S,

$$S_x:\mathfrak{X}\to[0,\infty),\; y\mapsto S_{xy}\,,\qquad x\in\mathfrak{X}\,, \tag{6.20}$$

as elements of L^2. It expresses that no row should be too different from all the other rows (cf. (6.22) below). This condition will imply the boundedness of $\|m(z)\|_{\mathscr{B}}$ away from $z=0$. The second condition will ensure the boundedness around $z=0$. To state it we need the following definition.

Definition 6.7 (Full indecomposability). *A $K\times K$ matrix $\mathbf{T}$ with nonnegative elements $T_{ij}\ge 0$, is called* **fully indecomposable** *(FID) provided that for any subsets $I,J\subset\{1,\dots,K\}$, with $|I|+|J|\ge K$, the submatrix $(T_{ij})_{i\in I,j\in J}$ contains a non-zero entry.*

The integral operator $S:\mathscr{B}\to\mathscr{B}$ is **block fully indecomposable** *if there exist an integer K, a fully indecomposable matrix $\mathbf{T}=(T_{ij})_{i,j=1}^K$ and a measurable partition $\mathcal{I}:=\{I_j\}_{j=1}^K$ of $\mathfrak{X}$, such that*

$$\pi(I_i)=\frac{1}{K}\,,\qquad and\qquad S_{xy}\ge T_{ij}\,,\quad whenever\quad (x,y)\in I_i\times I_j\,, \tag{6.21}$$

for every $1 \le i, j \le K$.

Our main result concerning the uniform boundedness is the following.

THEOREM 6.8 (Uniform bounds). *Suppose that in addition to* **A1-4.**, *S satisfies*

$$\liminf_{\varepsilon \downarrow 0} \inf_{x \in \mathfrak{X}} \int_{\mathfrak{X}} \frac{\pi(\mathrm{d}y)}{\varepsilon + \|S_x - S_y\|_2^2} \;=\; \infty\,, \tag{6.22}$$

and it is block fully indecomposable. Then the solution of the QVE is uniformly bounded, $|||m|||_{\mathbb{R}} < \infty$, and S has the property **A5.** *Hence the conclusions of both Theorem 6.2 and Theorem 6.4 hold.*

In Section 9 we present a more quantitative version of this result (Theorem 9.1) where the upper bound $\Phi < \infty$, in $|||m|||_{\mathbb{R}} \le \Phi$, depends on S only through a few parameters appearing in the quantitative versions (cf. assumptions **B1.** and **B2.**) of the estimate (6.22), and of the block fully indecomposability condition.

In the prominent example $(\mathfrak{X}, \pi(\mathrm{d}x)) = ([0,1], \mathrm{d}x)$ the condition (6.22) is satisfied if the rows $x \mapsto S_x \in \mathrm{L}^2$, are piecewise 1/2-Hölder continuous, in the sense that for some finite partition $\{I_k\}$ of $[0,1]$ into non-trivial intervals, the bound

$$\|S_x - S_y\|_2 \le C_1 |x - y|^{1/2}\,, \qquad \forall\, x, y \in I_k\,, \tag{6.23}$$

holds for every k. Furthermore, it is easy to see that S is block fully indecomposable provided it has a positive diagonal in the sense that for some $\varepsilon, \delta > 0$:

$$S_{xy} \ge \varepsilon \cdot \mathbb{1}\{|x - y| \le \delta\}\,. \tag{6.24}$$

Next we discuss the special situation in which the generating measure is supported on a single interval. A sufficient condition for this to hold is that the rows (6.20) of S cannot be split into two well separated subsets of L^1, as expressed by the quantity ξ_S below.

THEOREM 6.9 (Generating density supported on single interval). *Assume S satisfies* **A1-5.** *and $|||m|||_{\mathbb{R}} \le \Phi$ for some $\Phi < \infty$. Then there exists a threshold $\xi_* \sim 1$ such that under the assumption*

$$\sup_{A \subset \mathfrak{X}} \inf_{\substack{x \in A \\ y \notin A}} \|S_x - S_y\|_1 \;\le\; \xi_* \tag{6.25}$$

the generating density is supported on a single interval, i.e. $\operatorname{supp} v = [-\beta, \beta]$ *with $\beta \in [c, 2]$ for some $c \sim 1$. Moreover, for every $\delta \in (0, \beta)$ we have*

$$v_x(\tau) \;\gtrsim\; \delta^{1/2}\,, \qquad\qquad \tau \in [-\beta + \delta, \beta - \delta] \tag{6.26a}$$

$$v_x(-\beta + \omega) \;=\; v_x(\beta - \omega) \;=\; h_x \omega^{1/2} + \mathcal{O}(\omega)\,, \qquad \omega \in [0, \delta], \tag{6.26b}$$

where $h \in \mathscr{B}$ with $h_x \sim 1$. Moreover, $v(\tau)$ is uniformly 1/2-Hölder continuous in τ. Here ρ ,L, $\|S\|_{\mathrm{L}^2 \to \mathscr{B}}$ and Φ are considered the model parameters.

Combining the last two theorems we remark that if $\mathfrak{X} = [0,1]$ and S satisfies **A1-4.**, it is block fully indecomposable, and the row functions S_x are 1/2-Hölder continuous on the whole set $[0,1]$, then the conclusions (6.26a) and (6.26b) of Theorem 6.9 hold. Figure 6.3 shows an average generating density corresponding to an integral operator S with a smooth kernel.

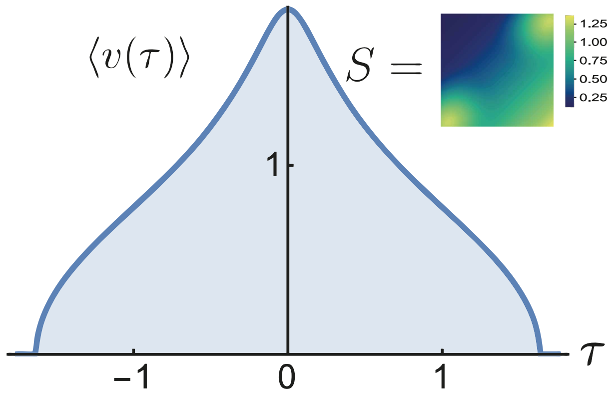

Figure 6.3: The smooth profile of S leads to a generating density that is supported on a single interval.

Finally we discuss the stability properties of the QVE (6.5). This result will be a cornerstone of the local law for Wigner-type random matrices proven in Part III of this work. Fix $z \in \overline{\mathbb{H}}$, and suppose $g \in \mathscr{B}$ satisfies

$$-\frac{1}{g} \;=\; z + Sg + d\,. \tag{6.27}$$

This equation is viewed as a perturbation of the QVE (6.5) by a ßmallffunction $d \in \mathscr{B}$. Our final result provides a bound on the difference between g and the unperturbed solution $m(z)$. The difference will be measured both in strong sense (in $\mathscr{B}$-norm) and in weak sense (integrated against a fixed bounded function).

Theorem 6.10 (Stability). *Assume S satisfies* **A1-A5.** *and $|||m|||_{\mathbb{R}} \leq \Phi$, for some $\Phi < \infty$. Suppose $g, d \in \mathscr{B}$ satisfy the perturbed QVE (6.27) for some fixed $z \in \overline{\mathbb{H}}$. There exists a small $\lambda \sim 1$, such that the following hold:*

(i) **Rough stability:** *Suppose that for some $\varepsilon \in (0,1)$,*

$$\langle v(\operatorname{Re} z)\rangle \;\geq\; \varepsilon\,, \qquad \textit{or} \qquad \operatorname{dist}(z, \operatorname{supp} v) \;\geq\; \varepsilon\,, \tag{6.28}$$

and g is sufficiently close to $m(z)$,

$$\|g - m(z)\|_{\mathscr{B}} \;\leq\; \lambda\varepsilon\,. \tag{6.29}$$

Then their distance is bounded in terms of d as

$$\|g - m(z)\|_{\mathscr{B}} \;\lesssim\; \varepsilon^{-2}\|d\|_{\mathscr{B}} \tag{6.30a}$$
$$|\langle w, g - m(z)\rangle| \;\lesssim\; \varepsilon^{-6}\|w\|_{\mathscr{B}}\|d\|_{\mathscr{B}}^2 \,+\, \varepsilon^{-2}|\langle J(z)w, d\rangle|\,, \qquad \forall\, w \in \mathscr{B}\,, \tag{6.30b}$$

for some z-dependent family of linear operators $J(z) : \mathscr{B} \to \mathscr{B}$, that depends only on S, and satisfies $\|J(z)\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$.

(ii) **Refined stability:** *There exist z-dependent families $t^{(k)}(z) \in \mathscr{B}$, $k = 1, 2$, depending only on S, and satisfying $\|t^{(k)}(z)\|_{\mathscr{B}} \lesssim 1$, such that the following holds. Defining*

$$\varpi(z) \;:=\; \operatorname{dist}(z, \operatorname{supp} v|_{\mathbb{R}}) \tag{6.31a}$$
$$\rho(z) \;:=\; \langle v(\operatorname{Re} z)\rangle \tag{6.31b}$$
$$\delta(z, d) \;:=\; \|d\|_{\mathscr{B}}^2 \,+\, |\langle t^{(1)}(z), d\rangle| \,+\, |\langle t^{(2)}(z), d\rangle|\,, \tag{6.31c}$$

assume g is close to $m(z)$, in the sense that

$$\|g - m(z)\|_{\mathscr{B}} \;\leq\; \lambda\varpi(z)^{2/3} + \lambda\rho(z)\,. \tag{6.32}$$

Then their distance is bounded in terms of the perturbation as

$$\|g - m(z)\|_{\mathscr{B}} \;\lesssim\; \Upsilon(z, d) \,+\, \|d\|_{\mathscr{B}} \tag{6.33a}$$
$$|\langle w, g - m(z)\rangle| \;\lesssim\; \Upsilon(z, d)\,\|w\|_{\mathscr{B}} \,+\, |\langle T(z)w, d\rangle|\,, \qquad \forall\, w \in \mathscr{B}\,, \tag{6.33b}$$

for some z-dependent family of linear operators $T(z) : \mathscr{B} \to \mathscr{B}$, that depends only on S, and satisfies $\|T(z)\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$. Here the key control parameter is

$$\Upsilon(z, d) \;:=\; \min\biggl\{\frac{\delta(z, d)}{\rho(z)^2}\,, \frac{\delta(z, d)}{\varpi(z)^{2/3}}, \delta(z, d)^{1/3}\biggr\}\,. \tag{6.34}$$

Note that the existence of g solving (6.27) for a given d is part of the assumptions of Theorem 6.10. An important aspect of the estimates (6.30) and (6.33) is that the upper bounds depend only on the unperturbed problem, i.e., on z and S, possibly through $m(z)$, in addition to the perturbation d itself.

The condition (6.32) of (ii) in the preceding theorem becomes ineffective when z approaches the critical points (6.16). A stronger but less transparent perturbation estimate is given as Proposition 13.1 below.

The guiding principle behind these estimates is that the bounds (6.30a) and (6.33a) are linear in $\|d\|_{\mathscr{B}}$, while the weaker bounds (6.30b) and (6.33b) are quadratic in $\|d\|_{\mathscr{B}}^2$ and linear in a specific average in d. The motivation behind the refined bound is that in the application in Part III of this work the perturbation d will have no definite sign and its (weighted) average will typically be comparable to $\|d\|_{\mathscr{B}}^2$. In (ii) of the theorem we see how the stability estimates deteriorate as z approaches the part of the real line where $\langle v\rangle$ becomes small, in particular near the edges of $\operatorname{supp} v$.

Another trivial application of our general stability result is to show that the QVE (6.5) is stable under perturbations of S.

REMARK 6.11 (Perturbations on S). *Suppose S and T are two integral operators satisfying* **A1-5.** *Let m and g be the unique solutions of the two QVE's*

$$-\frac{1}{m} = z + Sm \quad \textit{and} \quad -\frac{1}{g} = z + Tg\,.$$

Then g can be considered as a solution of the perturbed QVE (6.27), with

$$d \,:=\, (T-S)g\,.$$

In particular, d satisfies $\|d\|_{\mathscr{B}} \leq \|T-S\|_{\mathrm{L}^2\to\mathscr{B}}\|g\|_2 \leq 2\|T-S\|_{\mathrm{L}^2\to\mathscr{B}}|z|^{-1}$ by Theorem 6.1. Thus if $|\!|\!| m |\!|\!|_{\mathbb{R}} < \infty$ and the difference $\|T-S\|_{\mathrm{L}^2\to\mathscr{B}}$ is sufficiently small, then Theorem 6.10 can be used to control $\|g-m\|_{\mathscr{B}}$.

Besides the random matrix application in Part III of this work, the QVE (6.5) has previously appeared in at least two different contexts.

In [5] the authors show that the empirical distributions of the eigenvalues of a class random matrices with dependent entries converge to a probability measure μ on $\mathbb{R}$ as the dimension of the matrices becomes large. The measure μ is determined through the so-called *color equations* (cf. equation (3.9) on p. 1135):

$$\int_C \frac{s(c,c')P(\mathrm{d}c')}{\lambda - \Psi(c',\lambda)} \;=\; \Psi(c,\lambda)\,, \qquad \int_{\mathbb{R}} \frac{\mu(\mathrm{d}\tau)}{\lambda-\tau} = \int_C \frac{P(\mathrm{d}c)}{\lambda - \Psi(c,\lambda)}\,,$$

where $\lambda \in \mathbb{C}$ and the colour space is $C = [0,1]\times\mathbb{S}^1$, with $\mathbb{S}^1$ denoting the unit circle on the complex plane. Identifying $(\mathfrak{X},\pi) = (C,P)$, $x = c$, $\pi(\mathrm{d}x) := P(\mathrm{d}c)$, $z = \lambda$, we see that the colour equation is equivalent to the QVE (6.5). Indeed, we have the correspondence

$$S_{xy} = s(c,c')\,, \qquad m_x(z) = \frac{-1}{\lambda - \Psi(c,\lambda)}\,,$$

so that from (6.8) we see that $\langle v(\mathrm{d}\tau)\rangle = \pi\,\mu(\mathrm{d}\tau)$. Hence our results cover the asymptotic spectral statistics of a large class of random matrices with correlations.

We also remark that the QVE (6.5) naturally appears in the theory of Laplace-like operators

$$(Hf)(x) \;=\; \sum_{y\sim x} t_{xy}\,(f(x)-f(y))\,, \qquad f: V\to\mathbb{C}\,,$$

on rooted tree graphs Γ with vertex set V (see [45] for a review article and references therein). Set $m_x = (H_x - z)^{-1}(x,x)$, where H_x is the operator H restricted to the forward subtree with root x. A simple resolvent formula then shows that the QVE holds with $S_{xy} = |t_{xy}|^2 \mathbb{1}\{x < y\}$, where $x < y$ indicates that x is closer to the root of Γ than y. In this example S is not symmetric, but in a related model it may be chosen symmetric (rooted trees of finite cone types associated with a substitution matrix S, see [52]). In particular, real analyticity of the density of states (away from the spectral edges) in this model follows from Theorem 1.2 [We thank C. Sadel for pointing out this connection].

Convention 6.12 (Generic constants). *We denote by $C, C', C_1, C_2, \ldots$ and $c, c', c_1, c_2, \ldots$, etc., generic constants that depend only on the model parameters. The constants C, C', c, c' may change their values from one line to another, while the enumerated constants, such as c_1, C_2, have constant values within an argument or a proof.*

6.2 Outline of proof of Theorem 6.4

In this section we will explain and motivate the basic steps leading to our main results.

Stieltjes transform representation, L^2- and uniform bounds: It is a structural property of the QVE that its solution admits a representation as the Stieltjes transform of some generating measure on the real line (cf. (6.8)). This representation implies that m can be fully reconstructed from its own imaginary part near the real line.

From the Stieltjes transform representation of m a trivial bound, $|m_x(z)| \le (\operatorname{Im} z)^{-1}$, directly follows. A detailed analysis of the QVE near the real axis, however, requires $\operatorname{Im} z$-independent, bounds as its starting point. Away from $z = 0$ the L^2-bound, (6.9), meets this criterion. The estimate (6.9) is a structural property of the QVE in the sense that it follows from positivity and symmetry of S alone, and therefore quantitative assumptions such as **A4.** and **A5.** are not needed. This bound is derived from spectral information about a specific operator $F = F(z)$, constructed from the solution $m = m(z)$, that appears naturally when taking the imaginary part on both sides of the QVE. Indeed, (6.5) implies

$$\frac{\operatorname{Im} m}{|m|} = |m| \operatorname{Im} z + F \frac{\operatorname{Im} m}{|m|}, \qquad Fu := |m| S(|m|u). \tag{6.35}$$

As $\operatorname{Im} z$ approaches zero we may view this as an eigenvalue equation for the positive symmetric linear operator F. In the limit this eigenvalue equals 1 and $f = \operatorname{Im} m/|m|$ is the corresponding eigenfunction, provided $\operatorname{Im} m$ does not vanish. The Perron-Frobenius theorem, or more precisely, its generalisation to compact operators, the Krein-Rutman theorem, implies that this eigenvalue coincides with the spectral radius of F. This, in turn, implies the L^2-bound on m. These steps are carried out in detail at the end of Section 7. In fact, the norm of $F(z)$, as an operator on L^2, approaches 1 if and only if z approaches the support of the generating measure. Otherwise it stays below 1.

Requiring additional regularity conditions on S, such as **A4-5.** and the conditions of Theorem 6.8 from Section 9 enables us to improve the L^2-bound on m to a uniform bound. Already at this stage, we see that $z = 0$ needs a special treatment, because the L^2-bound is not effective here. The block fully indecomposability condition is designed to ensure uniform boundedness of m in a vicinity of $z = 0$. The uniform bounds are a prerequisite for most of our results concerning regularity and stability of the solution of the QVE.

Stability in the region where $\operatorname{Im} m$ is large: Stability properties of the QVE under small perturbations are essential, not just for applications in random matrix theory, but also as tools to analyse the regularity of the solution $m(z) \in \mathscr{B}_+$ as a function of z. Indeed, the stability of the QVE translates directly to regularity properties of the generating measure as described by Theorem 6.2. The stability of the solution deteriorates as $\operatorname{Im} m$ becomes small, this happens around the expansion points in $\mathbb{M}$ from Theorem 6.4

In order to see this deterioration of the stability, let us suppose that for a small perturbation $d \in \mathscr{B}$, the perturbed QVE has a solution $g(d)$ which depends smoothly on d,

$$-\frac{1}{g(d)} = z + Sg(d) + d\,. \tag{6.36}$$

For $d = 0$ we get back our original solution $g(0) = m$, with $m = m(z)$. We take the functional derivative with respect to d on both sides of the equation. In this way we derive a formula for the (Fréchet-)derivative $Dg(0)$, evaluated on some $w \in \mathscr{B}$:

$$(1 - m^2 S)Dg(0)w = m^2 w\,. \tag{6.37}$$

This equation shows that the invertibility of the linear operator $1 - m^2 S$ is relevant to the stability of the QVE. Assuming uniform lower and upper bounds on $|m|$, the invertibility of $1 - m^2 S$ is equivalent to the invertibility of the following related operator:

$$B := U - F = \frac{|m|}{m^2}(1 - m^2 S)|m|\,, \qquad Uw := \frac{|m|^2}{m^2}w\,.$$

Here, $|m|$ on the right of S is interpreted as a multiplication operator by $|m|$. Similarly, U is a unitary multiplication operator and F was introduced in (6.35). Away from the support of the generating measure the spectral radius of F stays below 1 and the invertibility of B is immediate. On the support of the generating measure the spectral radius of F equals 1. Here, the fundamental bound on the inverse of B is

$$\|B^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim \langle \operatorname{Im} m\rangle^{-1}\,, \tag{6.38}$$

apart from some special situations (cf. Lemma 8.8).

Let us understand the mechanism that leads to this bound in the simplest case, namely when $x \mapsto m_x(z)$ is a constant function. In this situation, the operator U is simply multiplication by a complex phase, $U = \mathrm{e}^{\mathrm{i}\varphi}$ with $\varphi \in (-\pi, \pi]$. The uniform bounds on m ensure that the operator F inherits certain properties from S. Among these are the conditions **A4.** and **A5.**. From these two properties we infer a spectral gap $\varepsilon > 0$,

$$\operatorname{Spec}(F) \subseteq [-1+\varepsilon, 1-\varepsilon] \cup \{1\}\,,$$

on the support of the generating measure. We readily verify the following bound on the norm of the inverse of B:

$$\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \le \begin{cases} |\mathrm{e}^{\mathrm{i}\varphi} - 1|^{-1} \sim \langle \operatorname{Im} m(z)\rangle^{-1} & \text{if } \varphi \in [-\varphi_*, \varphi_*]\,; \\ \varepsilon^{-1} & \text{otherwise.} \end{cases}$$

Here, $\varphi_* \in [0, \pi/2]$ is the threshold defined through $\cos\varphi_* = 1 - \varepsilon/2$, where the spectral radius $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2}$ becomes more relevant for the bound than the spectral gap. Similar bounds for the special case, $m_x = m_{\mathrm{sc}}$ (cf. (1.4)), first appeared in [36].

The bound (6.38) on the inverse of B implies a bound on the derivative $Dg(0)$ from (6.37). For a general perturbation d this means that the QVE is stable wherever the average generating measure is not too small. If d is chosen to be a constant function $d_x = z' - z$ then this argument yields the bound for the difference $m(z') - m(z)$, as $g(z' - z) = m(z')$. This can be used to estimate the derivative of $m(z)$ with respect to z and to prove existence and Hölder-regularity of the Lebesgue-density of the generating measure. This analysis is carried out in Sections 8 and 10.

Stability in the regime where $\operatorname{Im} m$ is small: The bound (6.38) becomes ineffective as $\langle \operatorname{Im} m \rangle$ becomes small. In fact, the norm of B^{-1} diverges owing to a single isolated eigenvalue, $\beta \in \mathbb{C}$, close to zero. This point is associated to the spectral radius of F, and the corresponding eigenvector, $Bb = \beta b$, is close to the Perron-Frobenius eigenvector of F, i.e., $b = f + \mathcal{O}(\langle \operatorname{Im} m \rangle)$, with $Ff = f$. The special direction b, in which B^{-1} becomes unbounded, is treated separately in Section 11. It is split off from the derivative $Dg(0)$ in the stability analysis. For the coefficient Θ, indicating how much of this derivative points in the bad direction b, a cubic equation is derived (cf. Subsection 11.2). Section 13 is concerned with deriving this equation and expanding its coefficients in terms of $\langle \operatorname{Im} m \rangle \ll 1$ at the edge.

Universal shape of v near its small values: In this regime understanding the dependence of the solution $g(d)$ of (6.36), is essentially reduced to understanding the scalar quantity Θ. This quantity satisfies a cubic equation (cf. Proposition 11.2), in which the coefficients depend only on the unperturbed solution m. In particular, we can follow the dependence of $m_x(z)$ on z by analysing the solution of this equation by choosing $d_x = z' - z$, a constant function. The special structure of the coefficients of the cubic equation, in combination with specific selection principles, based on the properties of the solution of the QVE, allows only for a few possible shapes that the solution Θ of the cubic equation may have. This is reflected in the universal shapes that describe the growth behaviour of the generating density at the boundary of its support. In Section 12 we will analyse the three branches of solutions for the cubic equation in detail and select the one that coincides with Θ. This will complete the proof of Theorem 6.4.

7 Existence, uniqueness and L^2-bound

This section contains the proof of Theorem 6.1. Namely assuming,

- *S satisfies* **A1-3.**,

we show that the QVE (6.5) has a unique solution, whose components m_x are Stieltjes transforms (6.8) of $x-$dependent probability measures, supported on the interval $[-2, 2]$. We also show (cf. (6.9)) that $m(z) \in \mathrm{L}^2$ whenever $z \neq 0$. Existence and uniqueness for a more general class of equations have already been proven in [43] using different methods. Nevertheless, we have included an elementary proof of existence and uniqueness for completeness.

The strategy of the proof is to interpret (6.5) as a fixed point equation for a contraction in an appropriate metric space. For this purpose we use the standard hyperbolic metric $d_{\mathbb{H}}$ on the complex upper half plane $\mathbb{H}$. This metric has the additional benefit of being invariant under $z \mapsto -z^{-1}$, which enables us to exchange the numerator and denominator on the left hand side of the QVE. The hyperbolic metric has been used in this fashion already in [39].

We start by summarising a few basic properties of $d_{\mathbb{H}}$. These will be expressed through the function

$$D(\zeta, \omega) := \frac{|\zeta - \omega|^2}{(\operatorname{Im} \zeta)(\operatorname{Im} \omega)}, \qquad \forall\, \zeta,\, \omega \in \mathbb{H}, \tag{7.1}$$

which is related to the hyperbolic metric through the formula

$$D(\zeta, \omega) = 2\,(\cosh d_{\mathbb{H}}(\zeta, \omega) - 1)\,. \tag{7.2}$$

Lemma 7.1 (Properties of hyperbolic metric). *The following three properties hold for D:*

1. *Isometries: If $\psi : \mathbb{H} \to \mathbb{H}$, is a linear fractional transformation, of the form*

$$\psi(\zeta) \;=\; \frac{\alpha\zeta + \beta}{\gamma\zeta + \mu}\,, \qquad \begin{bmatrix} \alpha & \beta \\ \gamma & \mu \end{bmatrix} \in \mathrm{SL}_2(\mathbb{R})\,,$$

then

$$D\big(\psi(\zeta), \psi(\omega)\big) \;=\; D(\zeta, \omega)\,.$$

2. *Contraction: If ζ, $\omega \in \mathbb{H}$ are shifted in the positive imaginary direction by $\lambda > 0$ then*

$$D(\mathrm{i}\lambda + \zeta, \mathrm{i}\lambda + \omega) \;=\; \Big(1 + \frac{\lambda}{\operatorname{Im}\zeta}\Big)^{-1}\Big(1 + \frac{\lambda}{\operatorname{Im}\omega}\Big)^{-1} D(\zeta, \omega)\,. \tag{7.3}$$

3. *Extremal values: Let K be a compact convex subset of $\mathbb{H}^2$ and let $\operatorname{ex} K$ denote the set of extreme points of K, i.e. all points in K, which cannot be written as a non-trivial convex combination of other points in K. Then*

$$\sup\big\{D(\zeta_1, \zeta_2) : (\zeta_1, \zeta_2) \in K\big\} = \sup\big\{D(\zeta_1, \zeta_2) : (\zeta_1, \zeta_2) \in \operatorname{ex} K\big\}\,. \tag{7.4}$$

Properties 1 and 2 follow immediately from (7.2) and (7.1). Property 3 is proven in Appendix B.2.

Corollary 7.2. *Suppose $0 \neq \phi \in \mathscr{B}^*$ is a bounded linear functional on $\mathscr{B}$ which is non-negative, i.e., $\phi(u) \geq 0$ for all $u \in \mathscr{B}$ with $u \geq 0$. Let $u, w \in \mathscr{B}_+$ with imaginary parts bounded away from zero, $\inf_x \operatorname{Im} u_x$, $\inf_x \operatorname{Im} w_x > 0$. Then*

$$D\big(\phi(u), \phi(w)\big) \;\leq\; \sup_{x \in \mathfrak{X}} D(u_x, w_x)\,. \tag{7.5}$$

Proof. By Property 1 in Lemma 7.1 we may assume the normalisation $\phi(1) = 1$, because for any $\lambda > 0$ the map $\mathbb{H} \ni \zeta \mapsto \lambda\zeta \in \mathbb{H}$ is an isometry of $\mathbb{H}$. Now we use Property 3 with K the closed convex hull of the points $(u_x, w_x) \in \mathbb{H}^2$ for $x \in \mathfrak{X}$. Since $(\phi(u), \phi(w)) \in K$, and

$$\operatorname{ex} K \;\subseteq\; \big\{(u_x, w_x) :\; x \in \mathfrak{X}\big\}\,, \tag{7.6}$$

the statement of the corollary follows from Property 3 of Lemma 7.1. □

In order to show existence and uniqueness of the solution of the QVE for a given S, we see that for any fixed $z \in \mathbb{H}$, a solution $m = m(z) \in \mathscr{B}_+$ of (6.7) is a fixed point of the map

$$\Phi(\,\bullet\,; z) : \mathscr{B}_+ \to \mathscr{B}_+\,, \qquad \Phi(u; z) \;:=\; -\frac{1}{z + Su}\,. \tag{7.7}$$

We can fix a constant $\eta_0 \in (0, 1)$ such that z lies in the domain

$$\mathbb{D}_{\eta_0} \;:=\; \big\{z \in \mathbb{H} :\; |z| < \eta_0^{-1}\,,\; \operatorname{Im} z > \eta_0\big\}\,. \tag{7.8}$$

We will now see that $\Phi(\,\bullet\,; z)$ is a contraction on the subset

$$\mathscr{B}_{\eta_0} \;:=\; \bigg\{u \in \mathscr{B}_+ :\; \|u\|_{\mathscr{B}} \leq \frac{1}{\eta_0}\,,\; \inf_{x \in \mathfrak{X}} \operatorname{Im} u_x \geq \frac{\eta_0^3}{4}\bigg\}\,, \tag{7.9}$$

equipped with the metric

$$d(u, w) \;:= \sup_{x \in \mathfrak{X}} d_{\mathbb{H}}(u_x, w_x)\,, \qquad u, w \in \mathscr{B}_{\eta_0}\,. \tag{7.10}$$

On $\mathscr{B}_{\eta_0}$ the metric d is equivalent to the metric induced by the uniform norm (6.3) of $\mathscr{B}$. Since $\mathscr{B}_{\eta_0}$ is closed in the uniform norm metric it is a complete metric space w.r.t. d.

Lemma 7.3. *For any $z \in \mathbb{D}_{\eta_0}$, the function $\Phi(\,\bullet\,; z)$ maps $\mathscr{B}_{\eta_0}$ into itself and satisfies*

$$\sup_{x\in\mathfrak{X}} D\Big(\big(\Phi(u;z)\big)_x, \big(\Phi(w;z)\big)_x\Big) \,\le\, \frac{1}{(1+\eta_0^2)^2} \sup_{x\in\mathfrak{X}} D(u_x, w_x)\,, \qquad u,w\in\mathscr{B}_{\eta_0}\,, \tag{7.11}$$

with D from (7.1).

Proof. First we show that $\mathscr{B}_{\eta_0}$ is mapped to itself. For this let $u \in \mathscr{B}_{\eta_0}$ be arbitrary. We start with the upper bound

$$|\Phi(u;z)| \,\le\, \frac{1}{\mathrm{Im}(z+Su)} \,\le\, \frac{1}{\mathrm{Im}\, z} \,\le\, \frac{1}{\eta_0}\,, \tag{7.12}$$

where in the second inequality we employed the non-negativity property of S and that $\mathrm{Im} u \ge 0$. Since $\|S\|_{\mathscr{B}\to\mathscr{B}} = 1$, we also find a lower bound,

$$|\Phi(u;z)| \,\ge\, \frac{1}{|z|+|Su|} \,\ge\, \frac{\eta_0}{2}\,. \tag{7.13}$$

Now we use this as an input to establish the lower bound on the imaginary part,

$$\mathrm{Im}\,\Phi(u;z) \,=\, \frac{\mathrm{Im}(z+Su)}{|z+Su|^2} \,\ge\, |\Phi(u;z)|^2\,\mathrm{Im}\, z \,\ge\, \frac{\eta_0^3}{4}\,. \tag{7.14}$$

We are left with showing the inequality in (7.11). For that we use the three properties of D in Lemma 7.1. By Property 1, the function D is invariant under the isometries $\zeta \mapsto -1/\zeta$ and $\zeta \mapsto \zeta - \mathrm{Re} z$ of $\mathbb{H}$ and therefore for any $x \in \mathfrak{X}$ we have

$$\begin{aligned} D\big(\big(\Phi(u;z)\big)_x, \big(\Phi(w;z)\big)_x\big) \,&=\, D\big(z+(Su)_x,\, z+(Sw)_x\big) \\ &=\, D\big(\mathrm{i}\,\mathrm{Im}\, z+(Su)_x,\, \mathrm{i}\,\mathrm{Im}\, z+(Sw)_x\big)\,, \end{aligned} \tag{7.15}$$

for all $u, w \in \mathscr{B}_{\eta_0}$. In case the non-negative functional $S_x \in \mathscr{B}^*$, defined through $S_x(u) := (Su)_x$, vanishes identically, the expression in (7.15) vanishes as well. Thus we may assume that $S_x \neq 0$. In view of Property 2 we can estimate

$$D\big(\mathrm{i}\,\mathrm{Im}\, z+(Su)_x,\, \mathrm{i}\,\mathrm{Im}\, z+(Sw)_x\big) \,\le\, \Big(1+\frac{\mathrm{Im}\, z}{\mathrm{Im}(Su)_x}\Big)^{-1}\Big(1+\frac{\mathrm{Im}\, z}{\mathrm{Im}(Sw)_x}\Big)^{-1} D\big((Su)_x, (Sw)_x\big)\,.$$

Since we have the normalisation $\|S\|_{\mathscr{B}\to\mathscr{B}} = 1$ and $\|u\|_{\mathscr{B}} \le 1/\eta_0$ for $u \in \mathscr{B}_{\eta_0}$, as well as the lower bound on the imaginary part, $\mathrm{Im} z \ge \eta_0$ for $z \in \mathbb{D}_{\eta_0}$, we conclude

$$D\Big(\big(\Phi(u;z)\big)_x, \big(\Phi(w;z)\big)_x\Big) \,\le\, (1+\eta_0^2)^{-2}\, D\big((Su)_x, (Sw)_x\big)\,. \tag{7.16}$$

Using Corollary 7.2 we find

$$D\big((Su)_x, (Sw)_x\big) \,\le \sup_{x\in\mathfrak{X}} D(u_x, w_x)\,. \tag{7.17}$$

This finishes the proof of the lemma. □

Lemma 7.3 shows that the sequence of iterates $(u^{(n)})_{n=0}^\infty$, with $u^{(n+1)} := \Phi(u^{(n)}; z)$, is Cauchy for any initial function $u^{(0)} \in \mathscr{B}_{\eta_0}$ and any $z \in \mathbb{D}_{\eta_0}$. Therefore, $(u^{(n)})_{n\in\mathbb{N}}$ converges to the unique fixed point $m = m(z) \in \mathscr{B}_{\eta_0}$ of $\Phi(\,\bullet\,; z)$. We have therefore shown existence and uniqueness of (6.7) for any given $z \in \mathbb{D}_{\eta_0}$ and thus, since η_0 was arbitrary, even for all $z \in \mathbb{H}$.

7.1 Stieltjes transform representation

In order to show that m_x can be represented as a Stieltjes transform (cf. (6.8)), we will first prove that m_x is a holomorphic function on $\mathbb{H}$. We can use the same argument as above on a space of function which are also z dependent. Namely, we consider the complete metric space, obtained by equipping the set

$$\mathfrak{B}_{\eta_0} \;:=\; \big\{\mathfrak{u} : \mathbb{D}_{\eta_0} \to \mathscr{B}_{\eta_0} \,:\, \mathfrak{u} \text{ is holomorphic}\big\}\,. \tag{7.18}$$

of $\mathscr{B}_{\eta_0}$-valued functions $\mathfrak{u}$ on $\mathbb{D}_{\eta_0}$ with the metric

$$d_{\eta_0}(\mathfrak{u}, \mathfrak{w}) \;:=\; \sup_{z\in\mathbb{D}_{\eta_0}} d(\mathfrak{u}(z), \mathfrak{w}(z))\,, \qquad \mathfrak{u}, \mathfrak{w} \in \mathfrak{B}_{\eta_0}\,. \tag{7.19}$$

Here the holomorphicity of $\mathfrak{u}$ means that the map $z \mapsto \langle \phi, \mathfrak{u}(z)\rangle$ is holomorphic on $\mathbb{D}_{\eta_0}$ for any element ϕ in the dual space of $\mathscr{B}$. Since the constant $(1+\eta_0^2)^{-2}$ in (7.11) only depends on η_0, but not on z, we see that the function $\mathfrak{u} \mapsto \Phi(\mathfrak{u})$, defined by

$$(\Phi(\mathfrak{u}))(z) \;:=\; \Phi(\mathfrak{u}(z); z)\,, \qquad \forall\, \mathfrak{u} \in \mathfrak{B}_{\eta_0}, \tag{7.20}$$

inherits the contraction property from $\Phi(\,\bullet\,; z)$. Thus the iterates $\mathfrak{u}^{(n)} := \Phi^n(\mathfrak{u}^{(0)})$ for any initial function $\mathfrak{u}^{(0)} \in \mathfrak{B}_{\eta_0}$ converge to the unique holomorphic function $m : \mathbb{D}_{\eta_0} \to \mathscr{B}_{\eta_0}$, which satisfies $m(z) = (\Phi(m))(z)$ for all $z \in \mathbb{D}_{\eta_0}$. Since $\eta_0 > 0$ was arbitrary and by the uniqueness of the solution on $\mathbb{D}_{\eta_0}$, we see that there is a holomorphic function $m : \mathbb{H} \to \mathscr{B}_+$ which satisfies $m(z) = (\Phi(m))(z) = \Phi(m(z); z)$, for all $z \in \mathbb{H}$. This function $z \mapsto m(z)$ is the unique holomorphic solution of the QVE.

Now we show the representation (6.8) for $m(z)$. We use that a holomorphic function $\phi : \mathbb{H} \to \mathbb{H}$ is a Stieltjes transform of a probability measure on the real line if and only if $|\mathrm{i}\eta\,\phi(\mathrm{i}\eta)+1| \to 0$ as $\eta \to \infty$ (cf. Theorem 3.5 in [40], for example). In order to see that

$$\lim_{\eta\to\infty} \sup_x \big|\mathrm{i}\eta\, m_x(\mathrm{i}\eta) + 1\big| \;=\; 0\,, \tag{7.21}$$

we write the QVE in the form:

$$z m_x(z) + 1 \;=\; m_x(z)\,(Sm(z))_x\,.$$

Using the normalisation $\|S\|_{\mathscr{B}\to\mathscr{B}} = 1$, we obtain

$$|z m_x(z) + 1| \;\le\; \|m(z)\, Sm(z)\|_{\mathscr{B}} \;\le\; \|m(z)\|_{\mathscr{B}}^2\,.$$

The right hand side is bounded by using $\operatorname{Im} m(z) \ge 0$ and the fact that S preserves positivity:

$$|m(z)| \;=\; \frac{1}{|z + Sm(z)|} \;\le\; \frac{1}{\operatorname{Im}(z + Sm(z))} \;\le\; \frac{1}{\operatorname{Im} z}\,, \qquad \forall\, z \in \mathbb{H}\,. \tag{7.22}$$

Choosing $z = \mathrm{i}\eta$, we get $|\mathrm{i}\eta\, m(\mathrm{i}\eta) + 1| \le \eta^{-2}$, and hence (7.21) holds true. This completes the proof of the Stieltjes transform representation (6.8).

As the next step we show that the measures v_x, $x \in \mathfrak{X}$, in (6.8) are supported on $[-2, 2]$. We start by extending these measures to functions on the upper half plane.

Definition 7.4 (Extended generating density). *Let m be the solution of the QVE. Then we define*

$$v_x(z) \;:=\; \operatorname{Im} m_x(z)\,, \qquad \forall\, x \in \mathfrak{X},\; z \in \mathbb{H}\,. \tag{7.23}$$

However, denote by $\operatorname{supp} v := \cup_x \operatorname{supp} v_x|_{\mathbb{R}}$ *the union of the supports of the generating measures* (6.8) *on the real line.*

This extension is consistent with the generating measure v_x appearing in (6.8) since $v_x(z)$, $z \in \mathbb{H}$, is obtained by regularising the generating measure with the Cauchy-density at the scale $\eta > 0$. Indeed, (7.23) is equivalent to

$$v_x(\tau + \mathrm{i}\eta) \;=\; \int_{-\infty}^{\infty} \frac{1}{\eta}\Pi\Big(\frac{\tau-\omega}{\eta}\Big)\, v_x(\mathrm{d}\omega)\,, \qquad \Pi(\lambda) := \frac{1}{\pi}\frac{1}{1+\lambda^2}\,, \qquad \tau \in \mathbb{R}\,, \eta > 0\,. \tag{7.24}$$

To show that the supports of the measures $v_x(\mathrm{d}\omega)$, for $x \in \mathfrak{X}$, lie in $[-2, 2]$, it suffices to show that $v_x(\tau + \mathrm{i}\eta)$ converges locally uniformly to zero on the set of $\tau \in \mathbb{R}$ that satisfy $|\tau| > 2$ as $\eta \to 0$. For this purpose let us fix $\tau \in \mathbb{R}$ with $|\tau| > 2$. As the first step we show

$$\|m(\tau + \mathrm{i}\eta)\|_{\mathscr{B}} \;\le\; \frac{2}{|\tau|}\,, \qquad |\tau| > 2\,,\ \eta > 0\,. \tag{7.25}$$

Using the trivial bound from (7.22) on m and the normalisation $\|S\|_{\mathscr{B}\to\mathscr{B}} = 1$, we see that for $\eta > |\tau|/2$ the inequality (7.25) is certainly fulfilled. Furthermore, the function $z \mapsto \|m(z)\|_{\mathscr{B}}$ is continuous on $\mathbb{H}$, as can be seen from the representation (6.8) of m as a Stieltjes transform. It is therefore enough to show that, provided (7.25) is satisfied for some fixed $\eta_0 > 0$, then there is an $\varepsilon > 0$ such that the inequality still holds for all η in the ε-ball around η_0, i.e. for $|\eta - \eta_0| < \varepsilon$. In fact, since $|\tau| > 2$ and by continuity we find for any η_0 for which (7.25) holds an $\varepsilon > 0$, such that for all $\eta > 0$ with $|\eta - \eta_0| \le \varepsilon$ we still have

$$\|m(\tau + \mathrm{i}\eta)\|_{\mathscr{B}} \;<\; \frac{|\tau|}{2}\,. \tag{7.26}$$

The QVE and $\|S\|_{\mathscr{B}\to\mathscr{B}} = 1$ then imply that for these $\eta > 0$ we also get

$$\|m(\tau + \mathrm{i}\eta)\|_{\mathscr{B}} \;\le\; \frac{1}{|\tau| - \|m(\tau+\mathrm{i}\eta)\|_{\mathscr{B}}} \;<\; \frac{1}{|\tau| - |\tau|/2} \;=\; \frac{2}{|\tau|}\,. \tag{7.27}$$

This shows the upper bound (7.25) for all $\eta > 0$ with $|\eta - \eta_0| \le \varepsilon$ and by continuity we can extend it to all $\eta > 0$.

Now we use the imaginary part of the QVE,

$$\frac{v(z)}{|m(z)|^2} \;=\; -\operatorname{Im}\frac{1}{m(z)} \;=\; \operatorname{Im} z + Sv(z)\,. \tag{7.28}$$

With (7.25) and the normalisation of S as an operator on $\mathscr{B}$, this leads to

$$\|v(\tau + \mathrm{i}\eta)\|_{\mathscr{B}} \;\le\; \frac{4}{\tau^2}\big(\eta + \|v(\tau+\mathrm{i}\eta)\|_{\mathscr{B}}\big)\,. \tag{7.29}$$

From this, using again that $|\tau| > 2$, we see that

$$\|v(\tau + \mathrm{i}\eta)\|_{\mathscr{B}} \;\le\; \frac{4\,\eta}{\tau^2 - 4}\,, \qquad \forall\, |\tau| \,>\, 2\,,\ \eta \,>\, 0\,. \tag{7.30}$$

We conclude that as $\eta \to 0$, the function $\tau \mapsto v_x(\tau + \mathrm{i}\eta)$ converges locally uniformly to zero on the set $|\tau| > 2$.

7.2 L^2-bound and operator F

In this subsection we prove the upper bound (6.9) on the L^2-norm of the solution. The proof of this L^2-bound relies on the analysis of the following symmetric non-negative operator $F(z)$, generated by the solution $m(z)$. The operator $F(z)$ will play a central role in the upcoming analysis.

Definition 7.5 (Operator F). *The operator $F(z) : \mathscr{B} \to \mathscr{B}$ for $z \in \mathbb{H}$, is defined by*

$$F(z)w := |m(z)|\, S(|m(z)|\, w)\,, \qquad w \in \mathscr{B}\,, \tag{7.31}$$

where $m(z)$ is the solution of the QVE at z.

We start the analysis by writing the QVE in the form

$$z\, m(z) \;=\; -1 - m(z) S m(z)\,. \tag{7.32}$$

Taking the L^2-norm on both sides yields

$$\|m(z)\|_2 \;\le\; \frac{1}{|z|}\big(1 + \|m(z)\, Sm(z)\|_2\big)\,. \tag{7.33}$$

The function mSm in (7.33) is bounded pointwise from above by the image of $F(z)$ on the constant function $x \mapsto 1$. In the rest of this subsection we will show that $F(z)$ is a bounded operator, not just on $\mathscr{B}$, but also on L^2 with the uniform operator norm bound,

$$\|F(z)\|_{\mathrm{L}^2\to\mathrm{L}^2} \;<\; 1\,, \qquad \forall\, z \in \mathbb{H}\,. \tag{7.34}$$

The bound (6.9), i.e., $\|m\|_2 \le 2\,|z|^{-1}$, then follows from $\|mSm\|_2 \le \||m|S|m|\|_2 \le \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}$ and (7.33).

To prove the remaining estimate (7.34) we first realise that S is a bounded operator on L^q, $1 \le q \le \infty$. Indeed, for a fixed $x \in \mathfrak{X}$ the functional $S_x : \mathscr{B} \to \mathbb{R}$, $w \mapsto (Sw)_x$ is bounded and non-negative in the sense that

$$S_x(w) \;\ge\; 0\,, \qquad \forall\, w \ge 0\,.$$

We define a normalised version, P_x, of the functional, S_x, through

$$(S1)_x\, P_x(w) \;:=\; (Sw)_x\,, \qquad \forall\, w \in \mathscr{B}\,. \tag{7.35}$$

In case $(S1)_x = 0$ we simply chose P_x to be an arbitrary bounded linear functional on $\mathscr{B}$ which is non-negative and normalised, i.e., $P_x 1 = 1$. For example, we may choose the evaluation functional at x. Using **A3.** we bound the normalising constants from above,

$$(S1)_x \;\le\; \sup_y (S1)_y \;=\; \|S1\|_{\mathscr{B}} \;\le\; \|S\|_{\mathscr{B}\to\mathscr{B}} \;=\; 1\,.$$

Let us show that S is bounded on L^q, with $\|S\|_{\mathrm{L}^q\to\mathrm{L}^q} \le 1$. By a version of Jensen's inequality for bounded linear, non-negative and normalised functionals on $\mathscr{B}$ we find

$$|(Sw)_x|^q \;=\; (S1)_x^q\, |P_x(w)|^q \;\le\; (S1)_x\, P_x(|w|^q) \;\le\; (S(|w|^q))_x\,.$$

By the symmetry **A1.** of S we obtain the operator norm bounds,

$$\|Sw\|_p^p \;=\; \langle 1, |Sw|^p\rangle \;\le\; \langle 1, S(|w|^p)\rangle \;=\; \langle S1, |w|^p\rangle \;\le\; \|S1\|_{\mathscr{B}} \|w\|_p^p \;\le\; \|w\|_p^p\,.$$

For each $z \in \mathbb{H}$ the operator $F(z)$ is also bounded on L^q, as $|m(z)|$ is trivially bounded by $(\operatorname{Im} z)^{-1}$. Furthermore, from the Stieltjes transform representation (6.8) it follows that $m_x(z)$ is also bounded away from zero uniformly in x,

$$\operatorname{Im} m_x(z) \;\ge\; \frac{\operatorname{Im} z}{(2+|z|)^2}\,, \qquad \forall\, x \in \mathfrak{X}\,. \tag{7.36}$$

The uniform bound (7.34) follows by considering the imaginary part (7.28) of the QVE. Rewriting this equation in terms of $F = F(z)$ we get

$$\frac{v}{|m|} = |m| \operatorname{Im} z + F\frac{v}{|m|}\,. \tag{7.37}$$

In order to avoid excess clutter we will suppressed the dependence on z in the equations, since z can be considered a fixed parameter here. The trivial lower bound (7.36) on $\operatorname{Im} m(z) = v(z)$ and the trivial upper bound $|m(z)| \le (\operatorname{Im} z)^{-1}$ imply that there is a scalar function $\varepsilon : \mathbb{H} \to (0,1)$, such that

$$F\frac{v}{|m|} \le (1-\varepsilon)\frac{v}{|m|}\,, \qquad \varepsilon := (\operatorname{Im} z) \inf_x \frac{|m_x|^2}{v_x} \in (0,1)\,. \tag{7.38}$$

We apply Lemma 7.6 below with the choices,

$$T := \frac{F}{1-\varepsilon} \qquad \text{and} \qquad h := \frac{v}{|m|} \gtrsim \frac{(\operatorname{Im} z)^2}{1+|z|^2}\,,$$

to conclude $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2} < 1$. This finishes the proof of Theorem 6.1.

Lemma 7.6 (Subcontraction). *Let T be a bounded symmetric operator on L^2 that preserves non-negative functions, i.e., if $u \ge 0$ almost everywhere, then also $Tu \ge 0$ almost everywhere. If there exists an almost everywhere positive function $h \in \mathrm{L}^2$, such that almost everywhere $Th \le h$, then $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2} \le 1$.*

The proof of Lemma 7.6 is postponed to Appendix B.2

8 Properties of solution

In this section we prove various technical estimates for the solution m of the QVE and the associated operator F (cf. (7.5)). For the stability analysis, we introduce the concept of the (spectral) gap of an operator.

Definition 8.1 (Spectral gap). *Let $T : \mathrm{L}^2 \to \mathrm{L}^2$ be a compact self-adjoint operator. The spectral gap* $\operatorname{Gap}(T)$ *is the difference between the two largest eigenvalues of $|T|$. If $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2}$ is a degenerate eigenvalue of $|T|$ then* $\operatorname{Gap}(T) = 0$.

Note that with this definition $\operatorname{Gap}(T) = \operatorname{Gap}(|T|)$. The following proposition collects the most important estimates in the special case when the solution is uniformly bounded.

Proposition 8.2 (Estimates when solution is bounded). *Suppose S satisfies* **A1-5.**. *Additionally, assume that for some $I \subseteq \mathbb{R}$, and $\Phi < \infty$ the uniform bound*

$$|||m|||_I \le \Phi\,,$$

applies. Then, considering $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}}, \Phi)$ as model parameters, the following estimates apply for every $z \in \mathbb{H}$, with $\operatorname{Re} z \in I$:

(i) *The solution m of the QVE satisfies the bounds*

$$|m_x(z)| \sim \frac{1}{1+|z|}\,, \qquad \forall\, x \in \mathfrak{X}\,. \tag{8.1}$$

(ii) *The imaginary part is comparable with its average, i.e.*

$$v_x(z) \sim \langle v(z)\rangle\,, \qquad \forall\, x \in \mathfrak{X}\,. \tag{8.2}$$

(iii) *The largest eigenvalue* $\lambda(z)$ *of* $F(z)$ *is single, and satisfies*

$$\lambda(z) = \|F(z)\|_{\mathrm{L}^2\to\mathrm{L}^2} \sim \frac{1}{1+|z|^2}\,. \tag{8.3}$$

(iv) *The operator* $F(z)$ *has a uniform spectral gap, i.e.,*

$$\mathrm{Gap}(F(z)) \sim \|F(z)\|_{\mathrm{L}^2\to\mathrm{L}^2} \tag{8.4}$$

(v) *The unique eigenvector* $f(z) \in \mathscr{B}$*, satisfying*

$$F(z)f(z) = \lambda(z)f(z)\,, \qquad f_x(z) \geq 0\,, \quad \text{and} \quad \|f(z)\|_2 = 1\,, \tag{8.5}$$

is comparable to 1*, i.e.*

$$f_x(z) \sim 1\,, \qquad \forall\, x \in \mathfrak{X}\,. \tag{8.6}$$

The proof of this proposition follows by combining the various auxiliary results presented in the next subsection (cf. p. 44).

8.1 Relations between components of m and F

In order to avoid excess clutter we will often suppress the symbol z from expressions whenever z can be considered as a fixed parameter. Moreover, the standing assumption in this section, unless explicitly stated otherwise, is that:

- *S satisfies* **A1-5.**

The comparison relations in this subsection hence depend on the model parameters ρ, L and $\|S\|_{\mathrm{L}^2\to\mathscr{B}}$. In particular, we do not assume that the solution is uniformly bounded here.

The smoothing condition **A4.** implies that for every $x \in \mathfrak{X}$ the linear functional $S_x : \mathrm{L}^2 \to \mathbb{R}$, $w \mapsto (Sw)_x$ is bounded. Hence, the row-function $y \mapsto S_{xy}$ is in L^2. The family of functions satisfies $\sup_x \|S_x\|_2 = \|S\|_{\mathrm{L}^2\to\mathscr{B}}$. The bound (6.11) implies that S is a Hilbert-Schmidt operator. The uniform primitivity **A5.** in turn implies that the integrated rows of S are bounded from below. To see this, fix x and consider a constant function $u = 1$ in (6.10):

$$\rho \,\leq\, \int (S^L)_{xy}\pi(\mathrm{d}y) \,\leq\, \biggl(\int S_{xu}\,\pi(\mathrm{d}u)\biggr)\biggl(\sup_t \int (S^{L-1})_{ty}\,\pi(\mathrm{d}y)\biggr) \,\leq\, \|S^{L-1}\|_{\mathscr{B}\to\mathscr{B}}\,\|S_x\|_1\,.$$

Using $\|S^{L-1}\|_{\mathscr{B}\to\mathscr{B}} \leq \|S\|_{\mathscr{B}\to\mathscr{B}}^{L-1} = 1$, we thus get

$$\inf_x \|S_x\|_1 \,\geq\, \rho \sim 1\,. \tag{8.7}$$

The following lemma shows that $\|m(z)\|_{\mathscr{B}}$ can diverge only when either $|z|$ or $\langle v(z)\rangle$ become zero. Furthermore, if a component $|m_x(z)|$ with $x \in \mathfrak{X}$, approaches zero while z stays bounded, some another component $|m_y(z)|$ will always diverge at the same time.

LEMMA 8.3 (Constraints on solution). *If S satisfies* **A1-5.** *then:*

(i) *The solution m of the QVE satisfies for every $x \in \mathfrak{X}$ and $z \in \mathbb{H}$:*

$$\min\biggl\{\frac{1}{|z|}, |z| + \frac{1}{\|m(z)\|_{\mathscr{B}}}\biggr\} \lesssim |m_x(z)| \lesssim \min\biggl\{\frac{1}{\inf_y |m_y(z)|^{2L-2}\langle v(z)\rangle}, \frac{1}{\mathrm{dist}(z,[-2,2])}\biggr\}. \tag{8.8}$$

(ii) *The imaginary part, $v_x(z)$ is comparable to its average, such that for every $x \in \mathfrak{X}$ and $z \in \mathbb{H}$ with $|z| \le 4$:*

$$\inf_y \left|m_y(z)\right|^{2L} \lesssim \frac{v_x(z)}{\langle v(z)\rangle} \lesssim \left(1+\frac{1}{\inf_y |m_y(z)|}\right)^2 \|m(z)\|_{\mathscr{B}}^4 . \tag{8.9}$$

For $|z| \ge 4$ the function v satisfies $v_x(z) \sim \langle v(z)\rangle$.

These bounds simplify considerably when $\|m\|_{\mathscr{B}}$ stays bounded (cf. Proposition 8.2 below).

Proof. We start by proving the lower bound on $|m|$. This is done by establishing an upper bound on $1/|m|$. Using the QVE we find

$$\frac{1}{|m|} = |z+Sm| \le |z| + \|S\|_{\mathrm{L}^2\to\mathscr{B}}\|m\|_2 \lesssim |z| + \min\left\{\frac{1}{|z|},\, \|m\|_{\mathscr{B}}\right\}.$$

For the last inequality, we used the fact that $\|m\|_2$ is less than or equal to both $\|m\|_{\mathscr{B}}$ and $2|z|^{-1}$. Taking the reciprocal on both sides shows the lower bound.

Now we will prove the upper bound on $|m|$. To this end, recall that

$$m_x(z) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{v_x(\mathrm{d}\tau)}{\tau - z},$$

where v_x/π is a probability measure. Since $\operatorname{supp} v \subseteq [-2,2]$, we may bound the absolute value of the denominator of the integrand from below by $\operatorname{dist}(z,[-2,2])$, and find

$$|m_x(z)| \le \frac{1}{\operatorname{dist}(z,[-2,2])}.$$

For the derivation of the second upper bound we rely on the positivity of the imaginary part of m:

$$|m| = \frac{1}{|\mathrm{Im}\,(z+Sm)|} \le \frac{1}{Sv}. \tag{8.10}$$

In order to continue we will now bound Sv from below. This is achieved by estimating v from below by $\langle v\rangle$. Indeed, writing the imaginary part of the QVE, as

$$\frac{v}{|m|^2} = -\mathrm{Im}\,\frac{1}{m} = \mathrm{Im}\, z + Sv,$$

and ignoring $\mathrm{Im}\, z > 0$, yields

$$v \ge |m|^2 Sv \ge \phi^2 Sv, \tag{8.11}$$

where we introduced the abbreviation

$$\phi := \inf_x |m_x|.$$

Now we make use of the uniform primitivity **A5.** of S and of (8.11). In this way we get the lower bound on Sv,

$$Sv \ge \phi^2 S^2 v \ge \ldots \ge \phi^{2L-2} S^L v \ge \phi^{2L-2}\rho\,\langle v\rangle,$$

Plugging this back into (8.10) finishes the proof of the upper bound on $|m|$.

We continue by showing the claim concerning $v/\langle v\rangle$. We start with the lower bound. We use (8.11) in an iterative fashion and employ assumption **A5.**,

$$v \;\geq\; \phi^2 Sv \;\geq\; \dots \;\geq\; \phi^{2L}\, S^L v \;\geq\; \phi^{2L}\, \rho\, \langle v\rangle\,. \tag{8.12}$$

This proves the lower bound $v/\langle v\rangle \gtrsim \phi^{2L}$.

In order to derive upper bounds for the ratio $v/\langle v\rangle$, we first write

$$v \;=\; |m|^2(\operatorname{Im} z + Sv) \;\leq\; \|m\|_{\mathscr{B}}^2\,(\operatorname{Im} z + Sv)\,. \tag{8.13}$$

We will now bound $\operatorname{Im} z$ and Sv in terms of $\langle v\rangle$. We start with $\operatorname{Im} z$. By dropping the term Sv from (8.13), and estimating $|m| \geq \phi$, we get $v \geq \phi^2 \operatorname{Im} z$, or by averaging:

$$\operatorname{Im} z \leq \frac{\langle v\rangle}{\phi^2}\,. \tag{8.14}$$

In order to bound Sv, we apply S on both sides of (8.13), and use the bound on $\operatorname{Im} z$, to get

$$Sv \;\leq\; \left(\frac{\langle v\rangle}{\phi^2} + S^2 v\right)\|m\|_{\mathscr{B}}^2\,. \tag{8.15}$$

The expression involving S^2 is useful, as we may now estimate the kernel $(S^2)_{xy}$ uniformly:

$$(S^2)_{xy} \;\leq\; \langle S_x, S_y\rangle \;\leq\; \|S_x\|_2\|S_y\|_2 \;\leq\; \sup_x \|S_x\|_2^2 \;=\; \|S\|_{\mathrm{L}^2\to\mathscr{B}}^2 \;\sim\; 1\,. \tag{8.16}$$

In particular, $S^2 v \leq \|S\|_{\mathrm{L}^2\to\mathscr{B}}^2\langle v\rangle \sim \langle v\rangle$, and thus

$$Sv \;\lesssim\; \left(1 + \frac{1}{\phi^2}\right)\|m\|_{\mathscr{B}}^2\,\langle v\rangle\,.$$

With this and (8.14) plugged back into (8.13) we find

$$v \;\lesssim\; \left(1 + \frac{1}{\phi}\right)^2\|m\|_{\mathscr{B}}^4\,\langle v\rangle\,.$$

Here we have also used that $\|m\|_{\mathscr{B}} \gtrsim 1$ for $|z| \leq 4$. This fact follows directly from the QVE,

$$1 \;=\; |m|\,|z + Sm| \;\leq\; \|m\|_{\mathscr{B}}\,(|z| + \|S\|_{\mathscr{B}\to\mathscr{B}}\|m\|_{\mathscr{B}}) \;=\; \|m\|_{\mathscr{B}}\,(|z| + \|m\|_{\mathscr{B}})\,.$$

The claim about v for $|z| \geq 4$ is easily seen from the fact that $v_x(z)$ is the harmonic extension of the measure $v_x(\mathrm{d}\tau)$ which is supported on $[-2, 2]$. □

Since the solution, $m(z)$ for $z \in \mathbb{H}$, of the QVE is bounded by the trivial bound (cf. (7.22)), also the operator $F(z)$ introduced in Definition 7.5 is seen to be a Hilbert-Schmidt operator. Consistent with the notation for S we write $F_{xy}(z)$ for the symmetric non-negative measurable kernel representing this operator. The largest eigenvalue and the corresponding eigenvector of $F(z)$ will play a key role when we analyse the sensitivity of $m(z)$ to z, or more generally, to any perturbations of the QVE.

LEMMA 8.4 (Operator F). *Assume that S satisfies* **A1-5.**. *Then for every $z \in \mathbb{H}$ the operator $F(z)$, defined in* (7.31), *is a Hilbert-Schmidt integral operator on* L^2, *with the integral kernel*

$$F_{xy}(z) \;=\; |m_x(z)|\, S_{xy}\, |m_y(z)|\,. \tag{8.17}$$

The norm $\lambda(z) := \|F(z)\|_{\mathrm{L}^2\to\mathrm{L}^2}$ is a single eigenvalue of $F(z)$, and it satisfies:

$$\|F(z)\|_{\mathrm{L}^2\to\mathrm{L}^2} \;=\; 1 - \frac{\operatorname{Im} z}{\alpha(z)}\,\langle f(z)\,|m(z)|\rangle \;\leq\; 1\,, \qquad z\in\mathbb{H}\,. \tag{8.18}$$

Here the positive eigenvector $f:\mathbb{H}\to\mathscr{B}$ is defined by (8.5), while $\alpha:\mathbb{H}\to(0,\infty)$ is the size of the projection of $v/|m|$ onto the direction f:

$$\alpha(z) \;:=\; \Big\langle f(z), \frac{v(z)}{|m(z)|}\Big\rangle\,. \tag{8.19}$$

Proof. The existence and uniqueness of $\|F(z)\|_{\mathrm{L}^2\to\mathrm{L}^2}$ as a non-degenerate eigenvalue and $f(z)$ as the corresponding eigenvector satisfying (8.5) follow from Lemma 8.5 below by choosing $r := |m(z)|$, using the trivial bound $\|m(z)\|_{\mathscr{B}} \lesssim \eta^{-1}$ to argue (using (8.8)) that also $r_- := \inf_x|m_x| > 0$.

In order to obtain (8.18) we take the inner product of (7.37) with $f = f(z)$. Since $F(z)$ is symmetric, we find

$$\Big\langle \frac{f\,v}{|m|}\Big\rangle \;=\; \langle f\,|m|\rangle\operatorname{Im} z + \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\Big\langle\frac{f\,v}{|m|}\Big\rangle\,. \tag{8.20}$$

Reorganising the terms yields the identity (8.18). □

The following lemma demonstrates how the spectral gap, $\operatorname{Gap}(F(z))$, the norm and the associated eigenvector of $F(z)$ depend on the component wise estimates of $|m_x(z)|$. Since we will later need this result for a general positive function $r:\mathfrak{X}\to(0,\infty)$ in the role of $|m(z)|$ we introduce the following bounds.

Lemma 8.5 (Maximal eigenvalue of scaled S). *Assume S satisfies* **A1-5.** *Consider an integral operator $\widehat{F}(r):\mathrm{L}^2\to\mathrm{L}^2$, parametrised by $r\in\mathscr{B}$, and defined through the integral kernel*

$$\widehat{F}_{xy}(r) \;:=\; r_r S_{xy} r_y\,, \qquad r\in\mathscr{B}\,. \tag{8.21}$$

If there exist upper and lower bounds, $0<r_-\leq r_+<\infty$, such that

$$r_-\leq r_x\leq r_+\,,\qquad \forall x\in\mathfrak{X}\,,$$

then $\widehat{F}(r)$ is Hilbert-Schmidt, and $\widehat{\lambda}(r) := \|\widehat{F}(r)\|_{\mathrm{L}^2\to\mathrm{L}^2}$ is a single eigenvalue with upper and lower bounds,

$$r_-^2 \;\lesssim\; \widehat{\lambda}(r) \;\lesssim\; r_+^2\,. \tag{8.22}$$

Furthermore, $\widehat{F}(r)$ has is a spectral gap,

$$\operatorname{Gap}(\widehat{F}(r)) \;\gtrsim\; r_-^{2L}\,r_+^{-8}\,\widehat{\lambda}(r)^{-L+5}\,, \tag{8.23}$$

and the unique eigenvector, $\widehat{f}(r)\in\mathrm{L}^2$, satisfying

$$\widehat{F}(r)\widehat{f}(r) = \widehat{\lambda}(r)\widehat{f}(r)\,,\qquad \widehat{f}_x(r)\geq 0\,,\quad \text{and}\quad \|\widehat{f}(r)\|_2 = 1\,, \tag{8.24}$$

is comparable to its average in the following sense,

$$\bigg(\frac{r_-^2}{\widehat{\lambda}(r)}\bigg)^{L} \;\lesssim\; \frac{\widehat{f}_x(r)}{\langle\widehat{f}(r)\rangle} \;\lesssim\; \frac{r_+^4}{\widehat{\lambda}(r)^2}\,. \tag{8.25}$$

If $\widehat{F}$ is interpreted as a bounded operator on $\mathscr{B}$, then the following relationship between the norm of the L^2-resolvent and the $\mathscr{B}$-resolvent holds:

$$\|(\widehat{F}(r)-\zeta)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \;\lesssim\; \frac{1}{|\zeta|}\Big(1 + r_+^2\|(\widehat{F}(r)-\zeta)^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2}\Big)\,,\qquad \zeta\notin\operatorname{Spec}(\widehat{F}(r))\cup\{0\}\,. \tag{8.26}$$

Feeding (8.22) into (8.25) yields $\Phi^{-2L}\langle\widehat{f}(r)\rangle \lesssim \widehat{f}(r) \lesssim \Phi^4\langle\widehat{f}(r)\rangle$, where $\Phi := r_+/r_-$. If $L \geq 4$ then similarly $\mathrm{Gap}(\widehat{F}(r)) \gtrsim \Phi^{-2L}$. For the proof of Lemma 8.5 we need the following simple result that is proven in the appendix. It can be found in this form or another in many text books. In the context of graph theory it says that the adjacency matrix of a connected graph has a spectral gap.

LEMMA 8.6 (Spectral gap for positive bounded operators). *Let T be a symmetric Hilbert-Schmidt integral operator on $\mathrm{L}^2(\mathfrak{X})$ with kernel $T_{xy} = T_{yx}$. Assume that $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2} = 1$ is a non-degenerate eigenvalue, and that h is an eigenvector satisfying $Th = h$, $\|h\|_2 = 1$, and $h_x \geq 0$.*

If there exist $\varepsilon > 0$ and $\Phi < \infty$, such that

$$T_{xy} \geq \varepsilon\,, \qquad \textit{and} \qquad h_x \leq \Phi\,, \tag{8.27}$$

for almost every $x, y \in \mathfrak{X}$, then

$$\mathrm{Gap}(T) \geq \frac{\varepsilon}{\Phi^2}\,.$$

PROOF OF LEMMA 8.5. Since S is compact, and $r \leq r_+$ also $\widehat{F} = \widehat{F}(r)$ is compact. The operator $\widehat{F}$ preserves the cone of non-negative functions $u \geq 0$. Hence by the Krein-Rutman theorem $\widehat{\lambda} = \|\widehat{F}\|_{\mathrm{L}^2\to\mathrm{L}^2}$ is an eigenvalue, and there exists a non-negative normalised eigenfunction $\widehat{f} \in \mathrm{L}^2(\mathfrak{X})$ corresponding to $\widehat{\lambda}$. The uniform primitivity assumption **A4.** and the smoothing property **A5.** combine to

$$\inf_{x,y\in\mathfrak{X}} (S^L)_{xy} \geq \rho\,.$$

Since $r_- > 0$, it follows that the integral kernel of $\widehat{F}^L$ is also strictly positive everywhere. In particular, $\widehat{F}$ is irreducible, and thus the eigenfunction $\widehat{f}$ is unique.

Now we derive the upper bound for $\widehat{\lambda}$. Since $\|w\|_p \leq \|w\|_q$ for $p \leq q$, we obtain

$$\widehat{\lambda}^2 = \|\widehat{F}\|^2_{\mathrm{L}^2\to\mathrm{L}^2} = \|\widehat{F}^2\|_{\mathrm{L}^2\to\mathrm{L}^2} \leq \|\widehat{F}^2\|_{\mathrm{L}^1\to\mathscr{B}} = \max_{x,y}\,(\widehat{F}^2)_{xy} \leq r_+^4\|S\|^2_{\mathrm{L}^2\to\mathscr{B}}\,, \tag{8.28}$$

which implies $\widehat{\lambda} \lesssim r_+^2$. Here we have used $(S^2)_{xy} = \langle S_x, S_y\rangle \leq \|S_x\|_2\|S_y\|_2$, and $\sup_x\|S_x\|_2 = \|S\|_{\mathrm{L}^2\to\mathscr{B}}$ to estimate:

$$(\widehat{F}^2)_{xy} \leq r_+^4(S^2)_{xy} \leq r_+^4\|S\|^2_{\mathrm{L}^2\to\mathscr{B}}\,. \tag{8.29}$$

For the lower bound on $\widehat{\lambda}$, we use first (8.7) to get $\iint \pi(\mathrm{d}x)\pi(\mathrm{d}y)S_{xy} \sim 1$. Therefore

$$\widehat{\lambda} = \|\widehat{F}\|_{\mathrm{L}^2\to\mathrm{L}^2} \geq \langle 1, \widehat{F}1\rangle \geq r_-^2 \iint \pi(\mathrm{d}x)\pi(\mathrm{d}y)S_{xy} \sim r_-^2. \tag{8.30}$$

Now we show the upper bound for the eigenvector. Applying (8.29), and $\langle\widehat{f}\rangle = \|\widehat{f}\|_1 \leq \|\widehat{f}\|_2 = 1$, yields

$$\widehat{\lambda}^2\widehat{f}_x = (\widehat{F}^2\widehat{f})_x \lesssim r_+^4\langle\widehat{f}\rangle \leq r_+^4.$$

This shows the upper bound on $\widehat{f}_x/\langle\widehat{f}\rangle$ and, in addition, $\widehat{f}_x \lesssim r_+^4/\widehat{\lambda}^2$.

In order estimate the ratios, $\widehat{f}_x/\langle\widehat{f}\rangle$ for $x \in \mathfrak{X}$, from below, we consider the operator

$$T := \Big(\frac{\widehat{F}}{\widehat{\lambda}}\Big)^L. \tag{8.31}$$

Using $\inf_{x,y}(S^L)_{xy} \geq \rho$, we get

$$\inf_{x,y} T_{xy} \;\geq\; \frac{r_-^{2L}}{\widehat{\lambda}^L}\,(S^L)_{xy} \;\gtrsim\; \biggl(\frac{r_-^2}{\widehat{\lambda}}\biggr)^{L}.$$

Hence, we find a lower bound on f through

$$\widehat{f}_x = (T\widehat{f})_x \;\gtrsim\; \biggl(\frac{r_-^2}{\widehat{\lambda}}\biggr)^{L}\langle \widehat{f}\,\rangle\,. \tag{8.32}$$

In order to prove (8.23), we apply Lemma 8.6 to the operator T, to get

$$\mathrm{Gap}(T) \geq \frac{\inf_{x,y} T_{xy}}{\|\widehat{f}\|_{\mathscr{B}}^2} \;\gtrsim\; \frac{(r_-^2/\widehat{\lambda})^L}{(r_+^4/\widehat{\lambda}^2)^2} \;=\; r_-^{2L} r_+^{-8}\widehat{\lambda}^{-(L-4)}\,.$$

This implies, using $(1-\sigma)^\tau \leq 1-\tau\sigma$ for $\sigma,\tau \in (0,1)$, and $L \sim 1$,

$$\frac{\mathrm{Gap}(\widehat{F})}{\widehat{\lambda}} \;=\; 1-\bigl(1-\mathrm{Gap}(T)\bigr)^{1/L} \;\geq\; \frac{\mathrm{Gap}(T)}{L} \;\sim\; r_-^{2L} r_+^{-8}\widehat{\lambda}^{-(L-4)}\,.$$

Finally, we show the bound (8.26). Here the smoothing condition **A4.** on S is crucial. Let $d,w \in \mathscr{B}$ satisfy $(\widehat{F}-\zeta)^{-1}w = d$. For $\zeta \notin \mathrm{Spec}(\widehat{F}) \cup \{0\}$, we have

$$\|d\|_2 \;\leq\; \|(\widehat{F}-\zeta)^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2}\,\|w\|_2 \;\leq\; \|(\widehat{F}-\zeta)^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2}\,\|w\|_{\mathscr{B}}\,. \tag{8.33}$$

Now, using $\|S\|_{\mathrm{L}^2\to\mathscr{B}} \lesssim 1$, we bound the uniform norm of d from above by the corresponding L^2-norm:

$$|\zeta|\,\|d\|_{\mathscr{B}} \;=\; \|\widehat{F}d-w\| \;\leq \|\widehat{F}\|_{\mathrm{L}^2\to\mathscr{B}}\|d\|_2 + \|w\|_{\mathscr{B}} \;\leq\; r_+^2\|S\|_{\mathrm{L}^2\to\mathscr{B}}\|d\|_2 + \|w\|_{\mathscr{B}}\,.$$

The estimate (8.26) now follows by using the operator norm on L^2 for the resolvent, i.e., the inequality (8.33), to estimate $\|d\|_2$ by $\|w\|_{\mathscr{B}}$. □

Proof of Proposition 8.2. All the claims follow by combining Lemma 8.3, Lemma 8.4 and Lemma 8.5. Indeed, let $z \in I + \mathrm{i}(0,\infty)$, so that $\|m(z)\|_{\mathscr{B}} \lesssim 1$. Then from (8.8) we see that

$$\min\bigl\{|z|^{-1},1\bigr\} \;\lesssim\; |m_x(z)| \;\lesssim\; \min\bigl\{\,1,\mathrm{dist}(z,[-2,2])^{-1}\bigr\}\,,$$

which is equivalent to the bound (i). Similarly, (8.9) yields the claim (ii).

For the claims concerning the operator $F(z)$ we use the formula (8.17) to identify $F(z) = \widehat{F}(|m(z)|)$, where $\widehat{F}(r)$ for $0 \leq r \in \mathscr{B}$, is the operator from Lemma 8.5. Now the parts (iii-v) follow from Lemma 8.5 by setting $r_+ := \|m(z)\|_{\mathscr{B}} \lesssim 1$ and $r_- := \inf_x|m(z)| \gtrsim 1$. □

8.2 Stability and operator B

The next lemma introduces the operator B that plays a central role in the stability analysis of the QVE. At the end of this subsection (Lemma 8.10) we present the first stability result for the QVE which is effective when both m is uniformly bounded and B^{-1} is bounded as operator on $\mathscr{B}$.

Lemma 8.7 (Perturbations). *Suppose $g,d \in \mathscr{B}$ satisfy the perturbed QVE,*

$$-\frac{1}{g} \;=\; z + Sg + d\,, \tag{8.34}$$

at some fixed $z \in \mathbb{H}$, and suppose $m = m(z)$ solves the unperturbed QVE. Then the scaled difference

$$u := \frac{g - m(z)}{|m(z)|}, \tag{8.35}$$

satisfies the equation

$$Bu = \mathrm{e}^{-\mathrm{i}a} u F u + |m| d + |m| \mathrm{e}^{-\mathrm{i}a} u d, \tag{8.36}$$

where the operator $B = B(z)$, and the angle-function $a = a(z) \in \mathscr{B}$, with $a_x(z) \in [0, 2\pi)$, are defined by:

$$B := \mathrm{e}^{-\mathrm{i}2a} - F, \qquad \textit{and} \qquad \mathrm{e}^{\mathrm{i}a} := \frac{m}{|m|}. \tag{8.37}$$

Proof. Multiplying on both sides of

$$\frac{g - m}{mg} = \frac{-1}{g} - \frac{-1}{m} = S(g - m) + d, \tag{8.38}$$

with $m g$, and writing everything in terms of u, F, $\mathrm{e}^{\mathrm{i}a}$, and $|m|$, yields

$$\begin{aligned} |m| u &= m(m + u|m|) S(|m|u) + m(m + |m|u) d \\ &= |m|^2 \mathrm{e}^{\mathrm{i}a} F u + |m| \mathrm{e}^{\mathrm{i}a} u F u + |m|^2 \mathrm{e}^{\mathrm{i}2a} d + |m|^2 \mathrm{e}^{\mathrm{i}a} d. \end{aligned}$$

Moving the second term in the last line to the other side, dividing by $\mathrm{e}^{\mathrm{i}2a}|m|$ on both sides, and using the definition of B to write $\mathrm{e}^{-\mathrm{i}2a} u - F u = Bu$, yields (8.36). □

Lemma 8.7 show that the inverse of the non-selfadjoint operator $B(z)$ plays an important role in the stability of the QVE against perturbations. In the next lemma we estimate the size of this operator in terms of the solution of the QVE at z.

Lemma 8.8 (Bounds on B^{-1}). *Let S satisfy* **A1-5.** *and suppose $z \in \mathbb{H}$ satisfies $\varepsilon \le |z| \le 4$ for some $\varepsilon > 0$. Then*

$$\|B(z)^{-1}\|_{\mathrm{L}^2 \to \mathrm{L}^2} \lesssim \langle v(z) \rangle^{-12}, \quad \textit{and} \quad \|B(z)^{-1}\|_{\mathscr{B} \to \mathscr{B}} \lesssim \langle v(z) \rangle^{-14}, \tag{8.39}$$

where the comparison relations depend on ε in addition to the usual model parameters $\|S\|_{\mathrm{L}^2 \to \mathscr{B}}$, ρ and L.

If there exist $\Phi < \infty$ and $\tau \in [-4, 4]$ such that $|||m|||_{\{\tau\}} = \sup_{\eta > 0} \|m(\tau + \mathrm{i}\eta)\|_{\mathscr{B}} < \Phi$, then

$$\begin{aligned} \|B(z)^{-1}\|_{\mathscr{B} \to \mathscr{B}} &\lesssim 1 + \|B(z)^{-1}\|_{\mathrm{L}^2 \to \mathrm{L}^2} && (8.40) \\ &\lesssim \big(|\sigma(z)| + \langle v(z) \rangle\big)^{-1} \langle v(z) \rangle^{-1} \qquad \forall\, z \in \mathbb{H},\ \textit{s.t.}\ \mathrm{Re}\, z = \tau. && (8.41) \end{aligned}$$

Here the function $\sigma : \mathbb{H} \to \mathbb{R}$, is defined by

$$\sigma(z) := \big\langle f(z)^3 \,\mathrm{sign}\, \mathrm{Re}\, m(z) \big\rangle. \tag{8.42}$$

The bound Φ is considered a model parameter of the comparison relations in (8.40) *and* (8.41)*.*

We will see below that (8.41) is sharp in terms of powers of $\langle v \rangle$. On the other hand, the exponents in (8.39) may be improved. For the proof of Lemma 8.8 we need the following auxiliary result which is proven in the appendix.

Lemma 8.9 (Norm of B^{-1}-type operators on L^2). *Let T be a compact self-adjoint and U a unitary operator on $\mathrm{L}^2(\mathfrak{X})$. Suppose that $\operatorname{Gap}(T) > 0$, $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2} \le 1$ and that for the normalized eigenvector h, corresponding to the non-degenerate eigenvalue $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2}$ of T, is not left invariant by U. Then*

$$\|(U-T)^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \le \frac{50}{\|T\|_{\mathrm{L}^2\to\mathrm{L}^2}\operatorname{Gap}(T)\,|1-\|T\|_{\mathrm{L}^2\to\mathrm{L}^2}\langle h, Uh\rangle|}. \tag{8.43}$$

Proof of Lemma 8.8. We will prove both the general case, when $\|m(z)\|_{\mathscr{B}}$ is not known to be bounded, and the case where $\|m(z)\|_{\mathscr{B}} \le \Phi$, partly in parallel. Depending on the case, z is always assumed to lie inside the appropriate domain, i.e., $|z| \in [\varepsilon, 4]$, in the former case, and $\operatorname{Re} z \in I$ in the latter. Besides this, we consider z to be fixed. Correspondingly, the comparison relations in this proof depend on either $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}}, \varepsilon)$ or $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}}, \Phi)$. We will also drop the explicit z-arguments in order to make the following formulas more transparent. In both cases the lower bound $|m_x(z)| \gtrsim 1$ follows from (8.8).

We start the analysis by noting that it suffices to consider only the norm of B^{-1} on L^2, since

$$\|B^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1 + \|m\|_{\mathscr{B}}^2\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2}. \tag{8.44}$$

In order to see this, we use the smoothing property **A4.** of S as in the proof of (8.26) before. In fact, besides replacing the complex number ζ with the function $\mathrm{e}^{2\mathrm{i}a}$, the proof of (8.33) carries over without further changes.

By the general property (8.18) of F we know that $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \le 1$. Furthermore, it is immanent from the definition of F and (8.7) that $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \gtrsim \inf_x |m_x|^2 \gtrsim 1$ in both of the considered cases. This shows that the hypotheses of Lemma 8.9 are met, and hence

$$\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \lesssim \operatorname{Gap}(F)^{-1}\big|1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\big\langle \mathrm{e}^{\mathrm{i}2a}f^2\big\rangle\big|^{-1}, \tag{8.45}$$

where we have also used $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \sim 1$. Now, by basic trigonometry,

$$\big\langle \mathrm{e}^{\mathrm{i}2a}f^2\big\rangle = \big\langle (1-2\sin^2 a)f^2\big\rangle + \mathrm{i}2\big\langle f^2 \sin a\, \cos a\big\rangle.$$

and therefore, using $|\alpha + \mathrm{i}\beta| \gtrsim |\alpha| + |\beta|$, $\alpha, \beta \in \mathbb{R}$, we get

$$\begin{aligned} &\big|1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\big\langle \mathrm{e}^{\mathrm{i}2a}f^2\big\rangle\big| \\ &\gtrsim \big|1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} + 2\|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\big\langle f^2\sin^2 a\big\rangle\big| + \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\big|\big\langle f^2 \sin a\, \cos a\big\rangle\big| \\ &\gtrsim 1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} + \|f\sin a\|_2^2 + \big|\big\langle f^2\sin a\, \cos a\big\rangle\big|. \end{aligned} \tag{8.46}$$

Here, we have again used $1 \lesssim \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \le 1$. Substituting this back into (8.45) yields

$$\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \lesssim \operatorname{Gap}(F)^{-1}\Big(1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} + \|f\sin a\|_2^2 + |\langle f^2 \sin a\, \cos a\rangle|\Big)^{-1}. \tag{8.47}$$

Case 1 (m with L^2-bound): In this case we drop the $\langle f^2 \sin a\, \cos a\rangle$ term and estimate

$$\|f\sin a\|_2 \ge \|f\|_2 \inf_x \sin a_x = \inf_x \frac{v_x}{|m_x|} \gtrsim \langle v\rangle^2, \tag{8.48}$$

where the bounds $\|m\|_{\mathscr{B}} \lesssim \varepsilon^{-C}\langle v\rangle^{-1} \sim \langle v\rangle^{-1}$ and $v \gtrsim \varepsilon^C\langle v\rangle \sim \langle v\rangle$ from Lemma 8.3 were used in the last inequality. Plugging (8.48) back into (8.47), and using (8.23) to estimate $\operatorname{Gap}(F) = \operatorname{Gap}(\widehat{F}(|m|)) \gtrsim \varepsilon^C\|m\|_{\mathscr{B}}^{-8} \gtrsim \langle v\rangle^8$ yields the desired bound:

$$\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \lesssim \operatorname{Gap}(F)^{-1}\|f\sin a\|_2^{-2} \lesssim \langle v\rangle^{-8}\langle v\rangle^{-4} \sim \langle v\rangle^{-12}. \tag{8.49}$$

The operator norm bound on $\mathscr{B}$ follows by combining this estimate with (8.44), and then using (8.8) to estimate $\|m\|_{\mathscr{B}} \lesssim \varepsilon^{-2L+2}\langle v\rangle^{-1} \sim \langle v\rangle^{-1}$.

Case 2 (m uniformly bounded): Now we assume $\|m\|_{\mathscr{B}} \leq \Phi \sim 1$, and thus all the bounds of Proposition 8.2 are at our disposal. This will allow us to extract useful information from the term $|\langle f^2 \sin a\, \cos a\rangle|$ in (8.47) that was neglected in the derivation of (8.49). Clearly, $|\langle f^2 \sin a\, \cos a\rangle|$ can have an important effect to (8.47) only when the term $\|f \sin a\|_2$ becomes small. Moreover, using $|m_x| \sim 1$ we see that this is equivalent to $\sin a_x = v_x/|m_x| \sim \langle v\rangle$ being small. Since also $\langle v\rangle \gtrsim \operatorname{Im} z$, for $|z| \lesssim 1$, the imaginary part of z will also be small in the relevant regime.

Writing the imaginary part of the QVE in terms $\sin a = v/|m|$, we get

$$\sin a = |m| \operatorname{Im} z \,+\, F \sin a\,. \tag{8.50}$$

Since we are interested in a regime where $\operatorname{Im} z$ is small, this implies, recalling $Ff = f$, that $\sin a$ will then almost lie in the span of f. To make this explicit, we decompose

$$\sin a = \alpha\, f + (\operatorname{Im} z)\, t\,, \qquad \text{with} \quad \alpha = \langle f, \sin a\rangle\,, \tag{8.51}$$

for some $t \in \mathscr{B}$ satisfying $\langle f, t\rangle = 0$. Let $Q^{(0)}$ denote the orthogonal projection $Q^{(0)}w := w - \langle f, w\rangle f$. Solving for t in (8.50) yields:

$$t \,=\, (\operatorname{Im} z)^{-1} Q^{(0)} \sin a \,=\, (1-F)^{-1} Q^{(0)} |m|\,. \tag{8.52}$$

Proposition 8.2 implies $\operatorname{Gap}(F) \sim 1$. Therefore we have

$$\|Q^{(0)}(1-F)^{-1}Q^{(0)}\|_{\mathrm{L}^2\to\mathrm{L}^2} \,\lesssim\, \operatorname{Gap}(F)^{-1} \,\sim\, 1\,.$$

In fact, since $f_x \sim 1$, a formula analogous to (8.44) applies, and thus we find

$$\|Q^{(0)}(1-F)^{-1}Q^{(0)}\|_{\mathscr{B}\to\mathscr{B}} \,\lesssim\, 1\,.$$

Applying this in (8.52) yields $\|t\|_{\mathscr{B}} \lesssim 1$, and therefore

$$\sin a \,=\, \alpha f \,+\, \mathcal{O}_{\mathscr{B}}(\operatorname{Im} z)\,. \tag{8.53}$$

Here, the notation $\mathcal{O}_{\mathscr{B}}(\varphi)$ for some positive function φ means that the expression is bounded by φ, up to a constant $C \sim 1$, after taking the supremum norm. Moreover, since we will later use the smallness of $\langle v\rangle \sim \sin a_x \sim \alpha$, we may expand

$$\cos a \,=\, (\operatorname{sign} \cos a)\sqrt{1-\sin^2 a} \,\,=\, \operatorname{sign} \operatorname{Re} m \,+\, \mathcal{O}_{\mathscr{B}}(\alpha^2)\,. \tag{8.54}$$

Combining this with (8.53) yields

$$\begin{aligned} \big\langle f^2 \sin a \cos a\big\rangle \,&=\, \big\langle f^2\big(\alpha f + \mathcal{O}_{\mathscr{B}}(\operatorname{Im} z)\big)\big(\operatorname{sign} \operatorname{Re} m \,+\, \mathcal{O}_{\mathscr{B}}(\alpha^2)\big)\big\rangle \\ &=\, \sigma\, \alpha \,+\, \mathcal{O}\big(\langle v\rangle^3 + \operatorname{Im} z\big)\,, \end{aligned} \tag{8.55}$$

where we have again used $\alpha \sim \langle v\rangle$, and used the definition, $\sigma = \langle f^3 \operatorname{sign}(\operatorname{Re} m)\rangle$, from the statement of the lemma.

For the term $1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}$ in the denominator of the r.h.s. of the main estimate (8.47) we make use of the explicit formula (8.18) for the spectral radius of F,

$$1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \,=\, \frac{\operatorname{Im} z}{\alpha}\, \langle f\, |m|\rangle\,. \tag{8.56}$$

By Proposition 8.2 we have $f_x \sim 1$, $|m_x| \sim 1$ and $\mathrm{Gap}(F) \sim 1$. Using this knowledge in combination with (8.55), (8.56) and $\alpha \sim \langle v \rangle$ we estimate the r.h.s. of (8.47) further:

$$\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \,\lesssim\, \frac{\langle v \rangle}{\langle v \rangle^3 + \langle f|m| \rangle \operatorname{Im} z \,+\, \bigl|\sigma\langle v \rangle^2 \,+\, \mathcal{O}\bigl(\langle v \rangle^4 + \langle v \rangle \operatorname{Im} z\bigr)\bigr|}\,. \tag{8.57}$$

Let us now see how from this and (8.44) the claim (8.41) follows: We have nothing to prove on the domain of z where $\langle v \rangle \gtrsim |\sigma|$ is satisfied, because then the $\langle v \rangle$-term is enough for the final result. We may therefore assume that $\langle v \rangle \le c|\sigma|$ for an arbitrarily small constant c. We are also done if $\operatorname{Im} z \gtrsim |\sigma|\langle v \rangle^2$ since then we may use the second summand on the r.h.s. of (8.57) to get the $|\sigma|\langle v \rangle$-term we need for (8.41). In particular, we can assume that the error term in (8.57) is $\mathcal{O}\bigl(|\sigma|\langle v \rangle^3\bigr)$. But by choosing c small enough this term is smaller than the leading term $|\sigma|\langle v \rangle^2$ and we get (8.41). □

We can now show that the perturbed QVE (8.34) is stable as long as v and $\|m\|_{\mathscr{B}}$ are bounded away from zero and infinity.

Lemma 8.10 (Stability when m and B^{-1} bounded). *Suppose $g, d \in \mathscr{B}$ satisfy the perturbed QVE (8.34), with $\inf_x |g_x| > 0$, at some point $z \in \mathbb{H}$. Assume*

$$\|m(z)\|_{\mathscr{B}} \,\le\, \Phi\,, \qquad \textit{and} \qquad \|B(z)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \,\le\, \Psi\,, \tag{8.58}$$

for some constants $\Phi, \Psi \ge 1$. Then there exists a linear operator $J(z)$ acting on $\mathscr{B}$, and depending only on S, with $\|J(z)\|_{\mathrm{L}^2\to\mathrm{L}^2} + \|J(z)\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$, such that if

$$\|g - m(z)\|_{\mathscr{B}} \,\le\, \frac{1}{2\,\|S\|_{\mathrm{L}^2\to\mathscr{B}}\Phi\,\Psi}\,, \tag{8.59}$$

then the correction $g - m(z)$ satisfies

$$\|g - m(z)\|_{\mathscr{B}} \,\le\, 2\,\Psi\,\Phi^2\|d(z)\|_{\mathscr{B}} \tag{8.60a}$$

$$|\langle w, g - m(z) \rangle| \,\lesssim\, \Psi^3\Phi^5\|w\|_2\|d\|^2_{\mathscr{B}} \,+\, \Psi\,\Phi^2|\langle J(z)w, d \rangle|\,, \tag{8.60b}$$

for any $w \in \mathscr{B}$. Here the comparison relations depends on the model parameters L, ρ and $\|S\|_{\mathrm{L}^2\to\mathscr{B}}$.

Proof. Expressing (8.38) in terms of $h = |m|u$, and re-arranging we obtain

$$h \,=\, |m|\,B^{-1}\bigl[\mathrm{e}^{-\mathrm{i}a}h\,Sh \,+\, (|m| + \mathrm{e}^{-\mathrm{i}a}h)d\,\bigr]\,. \tag{8.61}$$

Taking the $\mathscr{B}$-norm of (8.61), and using, $\|h\,Sh\|_{\mathscr{B}} \le \|S\|_{\mathscr{B}\to\mathscr{B}}\|h\|^2_{\mathscr{B}} = \|h\|^2_{\mathscr{B}}$, we obtain

$$\|h\|_{\mathscr{B}} \,\le\, \Phi\,\Psi\bigl(\|h\|_{\mathscr{B}} + \|d\|_{\mathscr{B}}\bigr)\,\|h\|_{\mathscr{B}} \,+\, \Psi\,\Phi^2\|d\|_{\mathscr{B}}\,.$$

Under the hypothesis (8.59) the first term is less than $(1/2)\|h\|_{\mathscr{B}}$, and hence absorbing it into the left hand side, we get $\|h\|_{\mathscr{B}} \le 2\Psi\Phi^2\|d\|_{\mathscr{B}}$. Here we used $1 = \|S\|_{\mathscr{B}\to\mathscr{B}} \le \|S\|_{\mathrm{L}^2\to\mathscr{B}}$. This proves (8.60a).

In order to prove (8.60b) we apply the linear functional $u \mapsto \langle w, u \rangle$ on (8.61), and obtain

$$\langle w, h \rangle \,=\, \langle \widetilde{w}, h\,Sh \rangle \,+\, \langle \widetilde{w}, hd \rangle \,+\, \Psi\,\Phi^2\langle Jw, d \rangle\,, \tag{8.62}$$

where we have identified the operator J in the statement of the lemma along with an auxiliary function,

$$\widetilde{w} \,:=\, \mathrm{e}^{\mathrm{i}a}\,(B^{-1})^*(|m|w) \qquad \text{and} \qquad J \,:=\, (\Psi\Phi^2)^{-1}|m|\,(B^{-1})^*(|m|\,\bullet\,)\,.$$

Clearly, B^* is like B except the angle function a is replaced by $-a$ in the definition (8.37). Hence, it follows that $\|\widetilde{w}\|_2 \le \|m\|_{\mathscr{B}}\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2}\|w\|_2 \le \Psi\Phi\,\|w\|_2$. Using (8.41) we see similarly that $\|J\|_{\mathrm{L}^2\to\mathrm{L}^2} \lesssim 1$. Moreover, (8.60a) implies, $\|hSh\|_2 \le \|S\|_{\mathrm{L}^2\to\mathscr{B}}\|h\|_{\mathscr{B}}^2 \lesssim \Psi^2\Phi^4\|d\|_{\mathscr{B}}^2$, and $\|hd\|_2 \lesssim \Phi^2\Psi\|d\|_{\mathscr{B}}^2$. Thus using Cauchy-Schwarz on the first two terms on the right hand side of (8.62) yields:

$$
\begin{aligned}
|\langle \widetilde{w}, hSh\rangle| \;&\le\; \|\widetilde{w}\|_2\|hSh\|_2 \;\lesssim\; (\Psi\Phi\|w\|_2)(\Psi\Phi^2\|d\|_{\mathscr{B}})^2 \;\sim\; \Psi^3\Phi^5\|w\|_2\|d\|_{\mathscr{B}}^2 \\
|\langle \widetilde{w}, hd\rangle| \;&\le\; \|\widetilde{w}\|_2\|hd\|_2 \;\lesssim\; (\Psi\Phi\|w\|_2)(\Psi\Phi^2\|d\|_{\mathscr{B}}^2) \;\sim\; \Psi^2\Phi^4\|w\|_2\|d\|_{\mathscr{B}}^2 \,.
\end{aligned}
$$

Plugging these back into (8.62) yields (8.60b) with some extra term $\Psi^2\Phi^4\|w\|_2\|d\|_{\mathscr{B}}^2$, which can be ignored as Φ and Ψ are constants larger than 1. $\square$

9 Uniform bounds

Under the assumptions **A1-5.** the solution m is in L^2. Our main results, such as Theorem 6.4, however, rely on the assumption that the solution m of the QVE is uniformly bounded in x, i.e., that there exists $\Phi < \infty$, depending only some model parameters, such that

$$
|||m|||_{\mathbb{R}} \le \Phi \,. \tag{9.1}
$$

In this section we show that the following extra conditions on S imply (9.1).

B1. *No outlier rows:* There exist constants $0 < \gamma < \vartheta < \infty$, such that

$$
\inf_{x\in\mathfrak{X}} \int_{\mathfrak{X}} \frac{\pi(\mathrm{d}y)}{(\gamma + \|S_x - S_y\|_2)^2} \;\ge\; \frac{1}{\vartheta^2} \,. \tag{9.2}
$$

B2. *Quantitative block fully indecomposability:* There exist two constants $\varphi > 0$, $K \in \mathbb{N}$, a fully indecomposable matrix $\mathbf{Z} = (Z_{ij})_{i,j=1}^K$, with $Z_{ij} \in \{0,1\}$, and a measurable partition $\mathcal{I} := \{I_j\}_{j=1}^K$ of $\mathfrak{X}$, such that for every $1 \le i,j \le K$ the following holds:

$$
\pi(I_j) \;=\; \frac{1}{K}\,, \qquad \text{and} \qquad S_{xy} \;\ge\; \varphi Z_{ij}\,, \quad \text{whenever} \quad (x,y) \in I_i \times I_j \,. \tag{9.3}
$$

The condition **B1.** is a quantitative version of (6.22). Similarly, the condition **B2.** amounts to a quantitative way of requiring S to be a block fully indecomposable operator (cf. Definition 6.7). The main result of this section is the following quantitative version of Theorem 6.8.

THEOREM 9.1 (Quantitative uniform bounds). *The solution m of the QVE* (6.5) *is uniformly bounded under the following set of assumptions:*

(i) **Away from zero:** *If S satisfies* **A1-5.** *and* **B1.** *then*

$$
\|m(z)\|_{\mathscr{B}} \;\le\; \frac{\sqrt{\vartheta}}{\gamma}\,, \qquad \forall\, z \in \mathbb{H}\,,\; |z| \ge 2\sqrt{\vartheta}\,. \tag{9.4}
$$

(ii) **Neighbourhoods of zero:** *Assume that S satisfies* **A1-4.** *and* **B2.** *Then there exist constants $\delta > 0$ and $\Phi < \infty$, both depending only on the parameters φ, K from* **B2.**, *s.t.*

$$
|||m|||_{[-\delta,\,\delta]} \;\le\; \Phi \,. \tag{9.5}
$$

Remark 9.2 (Positive diagonal). *Theorem 9.1 implies that for any S with a positive diagonal, i.e., with*

$$S_{xy} \geq \varepsilon \cdot \mathbb{1}\{|x-y| \leq \delta\}\,, \tag{9.6}$$

for some $\varepsilon, \delta > 0$ the solution of the corresponding QVE is bounded in a neighbourhood of $z = 0$, because such an S satisfies **B2**.

In Subsections 14.2 and 14.4 we have collected simple examples that demonstrates how the solution can become unbounded without the conditions **B1.** and **B2.**, respectively. Note that in (ii) we do not need to assume **A5.** This follows from (iii) of the following proposition and the estimate (9.17) below.

Proposition 9.3. *If $\mathbf{T} = (T_{ij})_{i,j=1}^{K}$ is a symmetric FID matrix then*

(i) *If $\mathbf{P}$ is a permutation matrix then $\mathbf{PT}$ and $\mathbf{TP}$ are FID.*

(ii) *There exists a permutation matrix $\mathbf{P}$ such that $(\mathbf{TP})_{ii} > 0$ for every $i = 1, \dots, K$.*

(iii) $(\mathbf{T}^{K-1})_{ij} > 0$ *for every $1 \leq i, j \leq K$.*

These properties are well known [8].

9.1 Uniform bound away from zero

In order to quantify the non-conformity of rows of S we introduce the family of strictly increasing auxiliary functions $\Gamma_x : [0,\infty] \to [0,\infty]$, parametrized by $x \in \mathfrak{X}$, by setting:

$$\Gamma_x(\tau) := \int_{\mathfrak{X}} \frac{\pi(\mathrm{d}y)}{(\tau^{-1} + \|S_y - S_x\|_2)^2}\,. \tag{9.7}$$

Recall that a generalised inverse of a non-decreasing function $\mu : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ on $\overline{\mathbb{R}} := [-\infty, \infty]$, is a non-decreasing function $\mu^{-1} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, defined by

$$\mu^{-1}(\lambda) := \inf\bigl\{\tau \in \mathbb{R} : \mu(\tau) \geq \lambda\bigr\}\,, \tag{9.8}$$

where the infimum may be infinite. The next lemma shows that a component m_x of m may diverge only if the x-th row of S is sufficiently far away from the other rows in L^2-sense.

Lemma 9.4 (Similarity in rows implies boundedness). *Let $z \in \mathbb{H}$, and assume $\|m(z)\|_2 \leq \Phi$ for some constant $\Phi < \infty$. Then*

$$|m_x(z)| \;\leq\; \frac{\Gamma_x^{-1}(\Phi^4)}{\Phi}\,, \qquad x \in \mathfrak{X}\,. \tag{9.9}$$

Proof. Since $\|m\|_2 \leq \Phi$, we have

$$\Phi^2 \;\geq\; \int_{\mathfrak{X}} \biggl|\frac{1}{m_y}\biggr|^{-2} \pi(\mathrm{d}y) \;\geq\; \int_{\mathfrak{X}} \biggl(\frac{1}{|m_x|} + \Phi\,\|S_y - S_x\|_2\biggr)^{-2} \;=\; \frac{\Gamma_x\bigl(\Phi|m_x|\bigr)}{\Phi^2}\,, \tag{9.10}$$

which is equivalent to (9.9). In order to get the second inequality in (9.10) we used the fact that m solves the QVE, and the Cauchy-Schwarz inequality:

$$\biggl|\frac{1}{m_y}\biggr| \;=\; \biggl|\frac{1}{m_x} - \frac{1}{m_x} + \frac{1}{m_y}\biggr| \;=\; \biggl|\frac{1}{m_x} + \bigl\langle S_x - S_y, m\bigr\rangle\biggr| \;\leq\; \biggl|\frac{1}{m_x}\biggr| + \|S_y - S_x\|_2\|m\|_2\,.$$

Since $\|m\|_2 \leq \Phi$ the second inequality of (9.10) follows. $\square$

In order to prove (i) of Theorem 6.8 we first express the hypothesis **B1**, i.e., (9.2), in terms of the generalised inverse of Γ_x:

$$\sup_{x\in\mathfrak{X}} \Gamma_x^{-1}\Big(\frac{1}{\vartheta^2}\Big) \le \frac{1}{\gamma}\,. \tag{9.11}$$

On the other hand, expressing the general L^2-bound (6.9) in terms of the parameter ϑ, we obtain

$$\|m(z)\|_2 \le \frac{2}{|z|} \le \frac{1}{\vartheta^{1/2}}, \quad \text{when} \quad |z| \ge 2\vartheta^{1/2}\,.$$

Thus using $\Phi := \vartheta^{-1/2}$ as the L^2-bound in (9.9), and then combining with (9.11) yields

$$|m_x(z)| \,\le\, \frac{1}{\vartheta^{-1/2}}\,\Gamma_x^{-1}\Big(\frac{1}{\vartheta^2}\Big) \,\le\, \frac{1}{\gamma} \,\le\, \frac{\vartheta^{1/2}}{\gamma}\,, \quad \text{when} \quad |z| \ge 2\vartheta^{1/2}\,.$$

Since this estimate holds for every x, the uniform bound (9.4) follows.

If the sets $[-\delta,\delta]$ and $\mathbb{R}\backslash[-2\vartheta^{-1/2}, 2\vartheta^{-1/2}]$, with $\delta,\vartheta>0$ from Theorem 9.1, overlap, then we have a uniform bound everywhere. This is the content of the following corollary. In particular, it proves Theorem 6.8, the qualitative version of Theorem 9.1.

Corollary 9.5 (Quantitative uniform bound everywhere). *Assume* **A1-5.** *and that there exist constants* $\delta,\Phi\sim 1$ *such that* (9.5) *holds. If additionally,*

$$\lim_{\tau\to\infty}\Gamma_x(\tau) > \delta^{-4} \qquad \forall\, x\in\mathfrak{X}\,,$$

then the solution of the QVE is everywhere uniformly bounded:

$$|\!|\!|m|\!|\!|_{\mathbb{R}} \,\le\, \max\Big\{\Phi,\, \delta\sup_x \Gamma_x^{-1}(\delta^{-4})\Big\}\,.$$

Consider now the special case $(\mathfrak{X},\pi(\mathrm{d}x)) = ([0,1],\mathrm{d}x)$. Since $|y-x|^{-1}$ is not integrable over $y\in[0,1]$ for any $x\in[0,1]$ we see that 1/2-Hölder regular rows (for definition, see (6.23)) yield uniformly bounded solutions $m(z)$ away from $z=0$. An easy computation yields the following quantitative estimate.

Remark 9.6 (Piecewise 1/2-Hölder continuous rows). *Suppose the rows of* $x\mapsto S_x\in\mathrm{L}^2$ *are 1/2-Hölder continuous in the sense that* (6.23) *applies for some partition* $\{I_k\}$*. Then for any* $\delta>0$,

$$\|m\|_{\mathbb{R}\backslash[-\delta,\delta]} \,\le\, \frac{\delta\exp(2C_1^2\delta^{-4})}{C_1\sqrt{\min_k |I_k|}}\,,$$

where the constant C_1 *is from* (6.23).

9.2 Uniform bound around zero

It is clear from Lemma 8.3 and (9.4) that $\mathrm{Re}\, z=0$ is a special point for the QVE. Note that the real and imaginary parts of the solution m of QVE are odd and even functions of $\mathrm{Re}\, z$ with fixed $\mathrm{Im}\, z$, respectively, i.e.,

$$m(-\overline{z}) \,=\, -\overline{m(z)}\,, \qquad \forall\, z\in\mathbb{H}\,. \tag{9.12}$$

In particular, $\mathrm{Re}\, m(\mathrm{i}\eta) = 0$ for $\eta>0$, and therefore the QVE becomes an equation for $v=\mathrm{Im}\, m$ alone,

$$\frac{1}{v(\mathrm{i}\eta)} \,=\, \eta \,+\, Sv(\mathrm{i}\eta)\,, \qquad \forall\, \eta \,>\, 0\,. \tag{9.13}$$

It is therefore not surprising that there is a connection between the regularity of S at $E = 0$ and the question of whether S is **scaleable**. By this we mean that there exists a positive measurable function h on $\mathfrak{X}$, such that

$$h_x (Sh)_x \,=\, 1\,, \qquad \forall\, x \in \mathfrak{X}\,. \tag{9.14}$$

It has been shown in Corollary 3.10 (p.276) of [13] that (9.14) has a unique bounded solution if S is block fully indecomposable. Here we will show additionally that $\|h\|_{\mathscr{B}} \sim 1$ where the comparison relation is defined w.r.t. the model parameters $(\|S\|_{\mathrm{L}^2\to\mathscr{B}}, \varphi, K)$.

In order to prove (ii) of Theorem 9.1 we use the fact that the solution of the QVE at $\operatorname{Re} z = 0$ is a minimiser of a functional on positive integrable functions L^1_+, where

$$\mathrm{L}^p_+ := \bigl\{ w \in \mathrm{L}^p :\; w_x > 0, \text{ for } \pi\text{-a.e. } x \in \mathfrak{X} \bigr\}\,, \qquad p \in [1,\infty]\,. \tag{9.15}$$

LEMMA 9.7 (Characterisation as minimiser). *Suppose S satisfies* **A1-3.** *and $\eta > 0$. Then the imaginary part $v(\mathrm{i}\eta) = \operatorname{Im} m(\mathrm{i}\eta)$ of the solution of the QVE is π-almost everywhere on $\mathfrak{X}$ equal to the unique minimiser of the functional $J_\eta : \mathrm{L}^1_+ \to \mathbb{R}$,*

$$J_\eta(w) \,:=\, \langle w, Sw\rangle \,-\, 2\,\langle \log w\rangle \,+\, 2\eta\,\langle w\rangle\,, \tag{9.16}$$

i.e.,

$$J_\eta(v(\mathrm{i}\eta)) \,=\, \inf_{w\in \mathrm{L}^1_+} J_\eta(w)\,.$$

The proof of the lemma is given in to Appendix B.4.

PROOF OF (II) OF THEOREM 9.1. Since $\mathbf{Z}$ is a K-dimensional fully indecomposable matrix it follows that $\min_{i,j}(\mathbf{Z}^{K-1})_{ij} \ge 1$. This implies that S is uniformly primitive,

$$(S^{K-1})_{xy} \,\ge\, \varphi^{K-1} \sum_{i,j=1}^{K} (\mathbf{Z}^{K-1})_{ij}\, \mathbb{1}\{x \in I_i,\, y \in I_j\}\,. \tag{9.17}$$

Showing the uniform bound (9.5) on m is somewhat involved and hence we split the proof into two parts. First we consider the case $\operatorname{Re} z = 0$ and show that the solution of QVE, $m(\mathrm{i}\eta) = \mathrm{i}v(\mathrm{i}\eta)$, is uniformly bounded. Afterwards we use a perturbative argument, which allows us to extend the uniform bound on m to a neighbourhood of the imaginary axis.

Because of the trivial bound $v(\mathrm{i}\eta) \le \|m(\mathrm{i}\eta)\|_{\mathscr{B}} \le \eta^{-1}$, we restrict ourselves to the case $\eta \le 1$.

STEP 1 (UNIFORM BOUND AT $\operatorname{Re} z = 0$): Here we will show the uniform bound

$$\sup_{\eta>0} \|v(\mathrm{i}\eta)\|_{\mathscr{B}} \lesssim 1\,. \tag{9.18}$$

As a first step we show that it suffices to establish a bound on the average of v only,

$$\|v(\mathrm{i}\eta)\|_{\mathscr{B}} \lesssim \langle v(\mathrm{i}\eta)\rangle\,, \qquad \forall \eta \in (0,1)\,. \tag{9.19}$$

To see (9.18) we recall (8.7) and use Jensen's inequality (cf. representation (7.35)) to get

$$\frac{1}{\int_{\mathfrak{X}} \pi(\mathrm{d}y)\, S_{xy}\, v_y} \,\lesssim\, \int_{\mathfrak{X}} \pi(\mathrm{d}y)\, \frac{S_{xy}}{v_y}\,.$$

This is used for $v = v(\mathrm{i}\eta)$ together with the QVE on the imaginary axis (cf. (9.13)) in the chain of inequalities,

$$v \,=\, \frac{1}{\eta + Sv} \,\le\, \frac{1}{Sv} \,\lesssim\, S\Bigl(\frac{1}{v}\Bigr) \,=\, S(\eta + Sv) \,\le\, \eta + S^2 v \,\lesssim\, \eta + \langle v\rangle\,. \tag{9.20}$$

In the last inequality we used the uniform upper bound (8.16) on the integral kernel of S^2. This establishes (9.19).

In order to bound $\langle v\rangle$ we argue as follows: First we note that

$$\langle v\rangle \,\le\, \max_{i=1}^{K} \langle v\rangle_i \,, \tag{9.21}$$

where $\langle v\rangle_i$ is the average of v over the i-th part of the partition from **B2.**, i.e.,

$$\langle w\rangle_i \,:=\, K \int_{I_i} \pi(\mathrm{d}x)\, w_x \,, \qquad \forall\, i = 1, \dots, K \,, \tag{9.22}$$

where we have used $\pi(I_i) = K^{-1}$. Let us also introduce a discretised version $\widetilde{J} : (0,\infty)^K \to \mathbb{R}$ of the functional J_η by

$$\widetilde{J}(\mathbf{w}) \,:=\, \frac{\varphi}{K} \sum_{i,j=1}^{K} w_i Z_{ij} w_j - 2\sum_{i=1}^{K} \log w_i \,, \qquad \mathbf{w} \,=\, (w_i)_{i=1}^K \in (0,\infty)^K \,, \tag{9.23}$$

where the matrix $\mathbf{Z}$ and the model parameter $\varphi > 0$ are from **B2.** The discretised functional is smaller than J_η, in the following sense:

$$\widetilde{J}\big(\langle w\rangle_1, \dots, \langle w\rangle_K\big) \,\lesssim\, J_\eta(w) \,, \qquad \forall\, w \in \mathscr{B} \,,\; w > 0 \,. \tag{9.24}$$

To see this we use **B2.** to estimate $S_{xy} \ge \varphi Z_{ij}$, $(x,y) \in I_i \times I_j$, for the quadratic term in the definition (9.16) of J_η. Moreover, we use Jensen's inequality to move the local average inside the logarithm. In other words, (9.24) follows, since

$$\begin{aligned} J_\eta(w) \,&\ge\, \varphi \sum_{i,j=1}^{K} \pi(I_i)\langle w\rangle_i Z_{ij}\, \pi(I_j)\langle w\rangle_j - 2\sum_{i=1}^{K} \pi(I_i)\, \langle \log w\rangle_i \\ &\ge\, \frac{1}{K}\biggl\{ \frac{\varphi}{K} \sum_{i,j=1}^{K} \langle w\rangle_i Z_{ij} \langle w\rangle_j - 2\sum_{i=1}^{K} \log \langle w\rangle_i \biggr\} \\ &=\, \frac{1}{K}\, \widetilde{J}\big(\langle w\rangle_1, \dots, \langle w\rangle_K\big) \,, \end{aligned} \tag{9.25}$$

for an arbitrary $w \in \mathrm{L}^1_+$. Since $K \in \mathbb{N}$ is considered as a model parameter in the statement (ii) of Theorem 9.1. the estimate (9.24) follows.

Now, by Lemma 9.7 the solution $v = v(\mathrm{i}\eta)$ of the QVE at $z = \mathrm{i}\eta$ is the (unique) minimiser of the functional $J_\eta : \mathrm{L}^1_+ \to \mathbb{R}$. In particular, it yields a smaller value of the functional than the constants function, and thus

$$J_\eta(v) \,\le\, J_\eta(1) \,=\, 1 + 2\eta \,\le\, 3 \,.$$

Combining this with (9.24), where we choose $w := v(\mathrm{i}\eta)$, we see that

$$\widetilde{J}\big(\langle v\rangle_1, \dots, \langle v\rangle_K\big) \,\le\, 3K \,\sim\, 1 \,. \tag{9.26}$$

Now we apply the following lemma which is proven in Appendix B.4.

LEMMA 9.8 (Uniform bound on discrete minimiser). *Assume* $\mathbf{w} := (w_i)_{i=1}^K \in (0,\infty)^K$ *satisfies*

$$\widetilde{J}(\mathbf{w}) \,\le\, \Psi \,,$$

for some $\Psi < \infty$*, where* $\widetilde{J} : (0,\infty)^K \to \mathbb{R}$ *is defined in* (9.23)*. Then there is a constant* $\Phi < \infty$ *depending only on* (Ψ, φ, K)*, such that*

$$\max_{k=1}^{K} w_k \le \Phi \,. \tag{9.27}$$

From (9.26) we see that we can apply Lemma 9.8 to the discretised vector $\mathbf{v} := (\langle v\rangle_1, \dots, \langle v\rangle_K)$ with $\Psi := 3K \sim 1$, and obtain,

$$\max_{i=1}^{K} \langle v\rangle_i \lesssim 1 .$$

Plugging this into (9.21) and the resulting inequality for $\langle v\rangle$ into (9.19) yields the chain of bounds, $\|v\|_{\mathscr{B}} \lesssim \langle v\rangle \le \max_k \langle v\rangle_k \lesssim 1$. This completes the proof of (9.18)

Step 2 (Extension to a neighbourhood): It remains to show that there exists $\delta \sim 1$, such that

$$\|m(\tau + \mathrm{i}\eta) - m(\mathrm{i}\eta)\|_{\mathscr{B}} \lesssim |\operatorname{Re} z| , \qquad \text{when} \quad |\tau| \le \delta . \tag{9.28}$$

Here $\Phi := \sup_\eta \|m(\mathrm{i}\eta)\|_{\mathscr{B}} < \infty$ is considered as a model parameter. In particular, the bound (8.8) on $|m(\mathrm{i}\eta)| = v(\mathrm{i}\eta)$ implies $v(\mathrm{i}\eta) \sim 1$. By (8.41) of Lemma 8.8 we find $\|B(\mathrm{i}\eta)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$. The bound (9.28) follows now from Lemma 8.10 by choosing $z = \mathrm{i}\eta$ and $d_x = \tau$. Indeed, the lemma states that with the abbreviation

$$h(\tau) := m(\tau + \mathrm{i}\eta) - \mathrm{i}v(\mathrm{i}\eta) ,$$

the following holds true. If $\|h(\tau)\|_{\mathscr{B}} \le c_0$ for sufficiently small constant $c_0 \sim 1$, then actually $\|h(\tau)\|_{\mathscr{B}} \le C_1|\tau|$ for some large constant C_1 depending only on Φ and the other model parameters.

The Stieltjes transform representation (6.8) implies that $h(\tau)$ is a continuous function in τ. As $h(0) = 0$, by definition, the bound $\|h(\tau)\|_{\mathscr{B}} \le C_1|\tau|$ applies as long as $C_1|\tau| \le c_0$ remains true. With the choice $\delta := c_0/C_1$ we finish the proof of (9.5). □

10 Regularity in variable z

We will now estimate the complex derivative $\partial_z m$ on the upper half plane $\mathbb{H}$. When $|||m|||_{\mathbb{R}} < \infty$ these bounds turn out to be uniform in z. This makes it possible to extend the domain of the map $z \mapsto m(z)$ to the closure $\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{R}$ of the complex upper half-plane $\mathbb{H}$. Additionally, we prove that the solution and its generating density are 1/3-Hölder continuous (Proposition 10.1), and analytic (Lemma 10.4) away from the special points $\tau \in \operatorname{supp} v$ where $v(\tau) = 0$. Combining these two results we prove Theorem 6.2 at the end of this section. Even if the uniform bound, $|||m|||_{\mathbb{R}} < \infty$, is not available we still obtain weaker regularity for the $\mathfrak{X}$-averaged solution $\langle m\rangle$.

At the technical level, we will consider the QVE at $z = z_2$ as a perturbation of the QVE at $z = z_1$, with $|z_2 - z_1|$ small. The idea is the same as in step 2 of the proof of (ii) from Theorem 9.1, where we extended the uniform bound of m from $\operatorname{Re} z = 0$ to a finite neighbourhood of the imaginary axis.

Proposition 10.1 (Extension to real line and continuity). *Assume* **A1-5.** *Then the average of the solution m of the QVE is uniformly Hölder-continuous away from zero, i.e., for any $\varepsilon > 0$ it satisfies,*

$$\bigl|\langle m(z_1)\rangle - \langle m(z_2)\rangle\bigr| \le C_1|z_1 - z_2|^{1/13} , \qquad z_1, z_2 \in \overline{\mathbb{H}} ,\ \varepsilon \le |z_1|, |z_2| \le 4 , \tag{10.1}$$

where the constant C_1 depends on ε in addition to the model parameters $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}})$.

If, additionally, $|||m|||_I \le \Phi < \infty$, on some interval $I := [\tau_-, \tau_+] \subseteq \mathbb{R}$, then for any $\varepsilon > 0$,

$$\|m(z_1) - m(z_2)\|_{\mathscr{B}} \le C_2|z_1 - z_2|^{1/3} , \quad |z_k| \le 4 ,\ \operatorname{Re} z_k \in [\tau_- + \varepsilon, \tau_+ - \varepsilon] ,\ k = 1, 2 , \tag{10.2}$$

where the constant C_2 depends on $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}}, \varepsilon, \Phi)$.

When m is known to be uniformly bounded everywhere, i.e., $|||m|||_{\mathbb{R}} < \infty$, we will automatically consider m, and all the related quantities as being defined on the extended upper half plane $\overline{\mathbb{H}} := \mathbb{H} \cup \mathbb{R}$. In the case of uniformly bounded solution of the QVE the proof of the proposition below shows the following.

COROLLARY 10.2 (Bound on derivative). *In Proposition 10.1 the bound* (10.2) *generalises to*

$$\big(|\sigma|\langle \operatorname{Im} m\rangle + \langle \operatorname{Im} m\rangle^2\big)\|\partial_z m\|_{\mathscr{B}} \le C_3\,, \quad \text{on} \quad \big\{z \in \mathbb{H} : |z| \le 4\,,\ \tau_- + \varepsilon \le \operatorname{Re} z \le \tau_+ - \varepsilon\big\}\,,$$

where C_3 depends on $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}}, \varepsilon, \Phi)$.

The proof of Proposition 10.1 also yields an analogous regularity of the mean generating measure.

COROLLARY 10.3 (Regularity of mean generating density). *Assume* **A1-5.** *Then the mean generating measure $\nu(\mathrm{d}\tau) = \langle v(\mathrm{d}\tau)\rangle$ satisfies*

$$\nu(\mathrm{d}\tau) \;=\; \widetilde{\nu}(\tau)\,\mathrm{d}\tau \,+\, \nu(\{0\})\,\delta_0(\mathrm{d}\tau)\,.$$

The Lebesgue-absolutely continuous part $\widetilde{\nu}(\tau)$ is symmetric in τ, and locally Hölder-continuous on $\mathbb{R}\backslash\{0\}$. More precisely, for every $\varepsilon > 0$

$$|\widetilde{\nu}(\tau_2) - \widetilde{\nu}(\tau_1)| \;\lesssim\; C_3|\tau_2 - \tau_1|^{1/13}\,, \qquad \forall\, \tau_1, \tau_2 \in [\varepsilon, \infty)\,, \tag{10.3}$$

where C_3 depends on ε in addition to the parameters in **A1-5.**

PROOF OF PROPOSITION 10.1. The solution m is a holomorphic function from $\mathbb{H}$ to $\mathscr{B}$ by Theorem 6.1. Hence, taking the derivative with respect to z on both sides of (6.5) yields

$$(1 - m(z)^2 S)\,\partial_z m(z) \;=\; m(z)^2\,, \qquad \forall\, z \in \mathbb{H}\,.$$

Expressing this in terms of $B = B(z)$, and suppressing the explicit z-dependence, we obtain

$$\mathrm{i}2\,\partial_z v \;=\; \partial_z m \;=\; |m|\, B^{-1}(|m|)\,. \tag{10.4}$$

Here we have also used the general property $\partial_z \phi = \mathrm{i}2\partial_z(\operatorname{Im}\phi)$, valid for all analytic function $\phi : \mathbb{C} \supseteq D \to \mathbb{C}$, to replace m by $v = \operatorname{Im} m$.

CASE 1 (no uniform bound on m): Suppose $z \in \mathbb{H}$ satisfies $\varepsilon/3 \le |z| \le 4$, for some $\varepsilon > 0$. Taking the average of (10.4) yields

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

Cauchy-Schwarz gives,

$$\big|\partial_z\langle v\rangle\big| \;\le\; 2^{-1}\|m\|_2\,\|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2}\|m\|_2 \;\lesssim\; \langle v\rangle^{-12}\,, \qquad \varepsilon/3 \le |z| \le 4\,. \tag{10.5}$$

In order to obtain the last bound we have used (6.9) and (8.39) to estimate $\|m(z)\|_2 \le 2|z|^{-1} \lesssim \varepsilon^{-1} \sim 1$, and $\|B(z)^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \lesssim \langle v(z)\rangle^{-12}$, for $|z| \gtrsim \varepsilon \sim 1$, respectively. The estimate (10.5) implies that $z \mapsto \langle v(z)\rangle$ is uniformly Hölder-continuous with Hölder-exponent $1/13$, and that its continuous extension to $z = \tau \in \mathbb{R}$, with $\varepsilon/3 \le |\tau| \le 4$, is a Lebesgue-density of ν, which has the same modulus of continuity.

It remains to extend this regularity from the mean generating density ν to its Stieltjes transform $\langle m\rangle$. To this end, we split ν, into two non-negative measures, $\nu = \nu_1 + \nu_2$, where $\nu_1(\mathrm{d}\tau) = \varphi(\tau)\,\nu(\mathrm{d}\tau)$, with

$$\varphi(\tau) \;:=\; \begin{cases} 1 & \text{if } |\tau| \,\ge\, 2\varepsilon/3\,; \\ (3/\varepsilon)\,|\tau| - 1 & \text{if } \varepsilon/3 \,<\, |\tau| \,<\, 2\varepsilon/3\,; \\ 0 & \text{if } |\tau| \,\le\, \varepsilon/3\,. \end{cases}$$

Then ν_1 has a Lebegue-density $\widetilde{\nu}_1$ and is supported in $[-2,2]$, since $\operatorname{supp} v \subseteq [-2,2]$ by Theorem 6.1. Furthermore,

$$|\widetilde{\nu}_1(\tau_1) - \widetilde{\nu}_1(\tau_2)| \lesssim |\tau_1 - \tau_2|^{1/13}, \qquad \forall\, \tau_1, \tau_2 \in \mathbb{R}. \tag{10.6}$$

For ν_2 we know that $\nu_2(\mathbb{R}) \leq 1$ and $\operatorname{supp} \nu_2 \subseteq [-2\varepsilon/3, 2\varepsilon/3]$. The Stieltjes transform

$$\langle m(z) \rangle \;=\; \frac{1}{\pi} \int_{\mathfrak{X}} \frac{\nu(\mathrm{d}\tau)}{\tau - z}\,,$$

is a sum of the Stieltjes transforms of ν_1 and ν_2. The Stieltjes transform of ν_1 is Hölder-continuous with Hölder-exponent $1/13$ since this regularity is preserved under the transformation. For the convenience of the reader, we state this simple fact as Lemma B.2 in the appendix and provide a proof as well. On the other hand, the Stieltjes transform of ν_2 satisfies

$$\left|\partial_z \int_{\mathfrak{X}} \frac{\nu_2(\mathrm{d}\tau)}{\tau - z}\right| \;\leq\; \frac{9}{\varepsilon^2} \lesssim 1\,, \qquad \text{when} \quad \varepsilon \leq |z| \leq 4\,,$$

and hence (10.1) follows.

Case 2 (solution uniformly bounded): Now we make the extra assumption $\|\!|m|\!\|_I \leq \Phi \sim 1$, $I := [\tau_-, \tau_+] \subseteq \mathbb{R}$. Taking the $\mathscr{B}$-norm of (10.4) immediately yields

$$|\partial_z v_x(z)| \;\leq\; \|m(z)\|_{\mathscr{B}}^2 \|B(z)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \;\lesssim\; \langle v(z) \rangle^{-2} \;\sim\; v_x(z)^{-2}\,. \tag{10.7}$$

Here we used (8.41) to estimate the norm of B^{-1}, and (ii) of Proposition 8.2 to argue that $v(z) \sim \langle v(z) \rangle$. We see that $z \mapsto v_x(z)$ is uniformly 1/3-Hölder continuous on $I + \mathrm{i}(0,\infty)$. Repeating the localisation argument used to extend the regularity of $\nu = \langle v \rangle$ to the corresponding Stieltjes transform yields (10.2). $\square$

Proof of Corollary 10.2. Using all the terms of (8.41) for the second bound of (10.7) and recalling $|\partial_z m| \sim |\partial_z v|$ yields the derivative bound of the corollary. $\square$

Apart from a set of special points the generating measure v has an analytic density on the real line.

Lemma 10.4 (Real analyticity of generating density). *Assume* **A1-5.** *Suppose $\tau \in \mathbb{R}\backslash\{0\}$ satisfies either $\langle v(\tau) \rangle > 0$ or $\operatorname{dist}(\tau, \operatorname{supp} v) > 0$. Then the generating density v is real analytic around τ. If additionally, $\|\!|m|\!\|_{\{0\}} < \infty$ then the condition $\tau \neq 0$ can be dropped.*

In the following proof we interpret m as a solution of an ODE defined on a subset of the complex Banach space $\mathbb{C} \times \mathscr{B}$. The analycity follows by a straightforward adaptation of standard techniques (cf. Chapter 4 of [61]) to our Banach-space setting.

Proof of Lemma 10.4. From the Stieltjes transform representation (6.8) we see that $m(z)$ can be extended to an analytic function on $\mathbb{C}\backslash\operatorname{supp} v$. Thus it remains to show that m is analytic on the set $\{\tau \in \mathbb{R} \setminus \{0\} : \langle v(\tau) \rangle > 0\}$. To this end we fix τ, with $\langle v(\tau) \rangle > 0$, and consider the ODE,

$$\begin{aligned} \partial_z w \;&=\; (1 - w^2 S)^{-1} w^2\,, \quad \text{on} \quad U(\tau,\delta) \\ w(\tau) &= q\,, \end{aligned} \tag{10.8}$$

where $U(\tau,\delta)$ is the complex disk of radius $\delta > 0$ centred at τ. We will show that if the initial value is chosen to coincide with the solution of the QVE, i.e., $q = m(\tau)$, and the (τ, S)-dependent radius $\delta > 0$ is sufficiently small, then (10.8) has a unique solution $w : U(\tau,\delta) \to \mathscr{B}$ that is analytic, and coincides with the solution of the QVE on their common domain $\overline{\mathbb{H}} \cap U(\tau,\delta)$.

The standard existence and uniqueness theory of ODEs yields the desired results, once we show that the map

$$w \mapsto h(w) := (1-w^2S)^{-1}(w^2)\,,$$

is uniformly bounded and analytic on the set

$$\mathbb{D}(\varepsilon) := \big\{ w \in \mathscr{B} : \|w - m(\tau)\|_{\mathscr{B}} < \varepsilon \big\}\,,$$

provided the $(\tau, m(\tau))$-dependent parameter $\varepsilon > 0$ is sufficiently small. Indeed, from this it follows that the analytic solution exists on the complex disk $U(\tau,\delta)$, with radius $\delta = \varepsilon/\Gamma$, where Γ is the supremum of $\|h(w)\|_{\mathscr{B}}$ among all $w \in \mathbb{D}(\varepsilon)$.

Clearly h is analytic whenever $1 - w^2S$ has an inverse as an operator on $\mathscr{B}$. Moreover, the norm of h factorizes $\|h(w)\|_{\mathscr{B}} \le \|w\|_{\mathscr{B}}^2 \|(1-w^2S)^{-1}\|_{\mathscr{B}\to\mathscr{B}}$, where $\|w\|_{\mathscr{B}} \le \|m(\tau)\|_{\mathscr{B}} + \varepsilon$. Thus we must find $\varepsilon > 0$, such that

$$\sup_{w \in \mathbb{D}(\varepsilon)} \|(1-w^2S)^{-1}\|_{\mathscr{B}\to\mathscr{B}} < \infty\,. \tag{10.9}$$

By using the identity $(A+D)^{-1} = (1+A^{-1}D)^{-1}A^{-1}$, with $A = 1 - m(z_0)^2S$ and $D = (w^2 - m(\tau)^2)S$, and estimating $\|D\|_{\mathscr{B}\to\mathscr{B}} \lesssim \|m(\tau)\|_{\mathscr{B}}\,\varepsilon$, we see that for sufficiently small ε, $\|(1+A^{-1}D)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \le 2$, and hence the first inequality below holds:

$$\|(1-w^2S)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \,\le\, 2\,\|(1-m(\tau)^2S)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \,\le\, \frac{\|m(\tau)\|_{\mathscr{B}}}{\inf_x |m_x(\tau)|}\, \|B(\tau)^{-1}\|_{\mathscr{B}\to\mathscr{B}}\,.$$

For the second bound we have expressed $1 - m(\tau)^2S$ in terms of $B(\tau)$ by sandwiching the latter between bounded multiplication operators, i.e., $1 - m^2S = \mathrm{e}^{\mathrm{i}2a}|m|\,B(|m|^{-1}\cdot\bullet)$.

If $\langle v(\tau)\rangle > 0$ and $\tau \ne 0$, then the basic constraints of Lemma 8.3 imply $0 < |m_x(\tau)| < \infty$ uniformly in x. If $|\!|\!|m|\!|\!|_{\{0\}} < \infty$ then m can be extended to a uniformly bounded function on a real neighbourhood of $\tau = 0$ as in Subsection 9.2. Using Lemma 8.8 we hence see that $\|B(\tau)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim \langle v(\tau)\rangle^{-14} < \infty$. This shows that there is an $(\tau, m(\tau))$-dependent ε such that (10.9) applies, and hence the proof is complete. □

Combining the analyticity and the Hölder regularity we will now prove the next main result.

Proof of Theorem 6.2. Here we assume $|\!|\!|m|\!|\!|_{\mathbb{R}} \le \Phi$, with $\Phi < \infty$, considered as a model parameter. Using the bound (10.2) of Proposition 10.1 we see that m can be extended to a 1/3-Hölder continuous function on the real line. Hence, from (7.24) we read off that the generating measure must have a Lebesgue-density equal to $\mathrm{Im}\, m|_{\mathbb{R}}$. In particular, this density function inherits the Hölder regularity from $m|_{\mathbb{R}}$, i.e., for some $C_1 \sim 1$:

$$|v_x(\tau') - v_x(\tau)| \le C_1 |\tau' - \tau|^{1/3}\,, \qquad \forall\, \tau, \tau' \in \mathbb{R}\,. \tag{10.10}$$

Since $|\!|\!|m|\!|\!|_{\mathbb{R}} \sim 1$, using Lemma 8.3, we see that $v_x(z) \sim v_y(z)$ for $z \in \overline{\mathbb{H}}$. The symmetry $v(-\tau) = v(\tau)$ on the other hand follows directly from (9.12).

By Lemma 10.4 the function v is analytic on $\mathbb{R}\backslash \operatorname{supp} v$. Let $\tau_0 \in \mathbb{R}$ be such that $v(\tau_0) > 0$. In order to bound the derivatives of v at τ_0 we use (10.10) to estimate

$$\mathrm{dist}(\tau_0, \mathbb{R}\backslash\mathrm{supp}\, v) \,\ge\, C_1^{-3}\langle v(\tau_0)\rangle^3 =: \varrho > 0\,.$$

This implies that v is analytic on the ball of radius ϱ centred at τ_0. The elementary bound on analytic functions thus shows that the k-th derivative of v at τ_0 is bounded by $k!\,\varrho^{-k}$. This proves (iii) of the theorem. □

11 Perturbations when generating density is small

We will assume in this and the following sections that S satisfies **A1-5.** and that the solution is uniformly bounded everywhere $\|\|m\|\|_{\mathbb{R}} \leq \Phi < \infty$. The numbers $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}}, \Phi)$ will be considered as model parameters. Due to the uniform boundedness, m and all the related quantities are extended to $\overline{\mathbb{H}}$ (cf. Proposition 10.1). Furthermore, the standing assumptions also imply that Proposition 8.2 is effective, i.e.,

$$|m_x(z)|, f_x(z), \mathrm{Gap}(F(z)) \sim 1\,, \quad \text{and} \quad v_x(z) \sim \langle v(z)\rangle \sim \alpha(z)\,, \quad \forall\, z \in \overline{\mathbb{H}}\,,\ x \in \mathfrak{X}\,. \tag{11.1}$$

In particular, the three quantities $v, \langle v\rangle, \alpha = \langle f, \sin a\rangle$, can be interchanged at will, as long as only their sizes up to constants depending on the model parameters matter.

The stability of the QVE against perturbations deteriorates when the generating density becomes small. This can be seen from the explosion in the estimate

$$\|B^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim \langle v\rangle^{-2}\,, \tag{11.2}$$

(cf. (8.41) and (11.9b) below) for the inverse of the operator B, introduced in (8.37). This norm appears in the estimates (8.60) relating the norm of the difference,

$$u = \frac{g-m}{|m|} \tag{11.3}$$

of the two solutions g and m of the perturbed and the unperturbed QVE,

$$-\frac{1}{g} = z + Sg + d \qquad \text{and} \qquad -\frac{1}{m} = z + Sm\,,$$

respectively, to the size of the perturbation d.

The unboundedness of B^{-1} in (11.2), as $\langle v\rangle \to 0$, is caused by the vanishing of B in a one-dimensional subspace of L^2 corresponding to the eigendirection of the smallest eigenvalue of B. Therefore, in order to extend our analysis to the regime $\langle v\rangle \approx 0$ we decompose the perturbation (11.3) into two parts:

$$u \,=\, \Theta\, b \,+\, r\,. \tag{11.4}$$

Here Θ is a scalar, and b is the eigenfunction corresponding to the smallest eigenvalue of B. The remaining part, $r \in \mathscr{B}$, lies inside a subspace where B^{-1} is bounded due to the spectral gap of F. As B is not symmetric r and b are not orthogonal w.r.t. the standard inner product (6.6) on L^2. The main result of this section is Proposition 11.2 which shows that for sufficiently small $\langle v\rangle \leq \varepsilon_*$, the b-component, Θ, and the other two small quantities d and $\langle v\rangle$ satisfy a cubic equation in Θ and d that is autonomous up to the leading order in appropriate small parameters. We will use the symbol $\varepsilon_* \sim 1$ as the upper threshold for $\langle v\rangle$ and its value will be reduced along the proofs.

11.1 Expansion of operator B

In this subsection we collect necessary information about the operator $B : \mathscr{B} \to \mathscr{B}$, needed to derive and analyse the cubic equation for Θ. Recall, that the **spectral projector** P_λ corresponding to an isolated eigenvalue λ of a compact operator T acting on a Banach space X is obtained (cf. Theorem 6.17 in Chapter 3 of [44]) by integrating the resolvent of T around a loop encircling only the eigenvalue λ:

$$P_\lambda := \frac{-1}{2\pi\mathrm{i}} \oint_\Gamma (T-\zeta)^{-1}\mathrm{d}\zeta\,. \tag{11.5}$$

LEMMA 11.1 (Expansion of B in bad direction). *There exists $\varepsilon_* \sim 1$ such that, uniformly in $z \in \overline{\mathbb{H}}$ with $|z| \leq 4$ the following holds true: If*

$$\alpha \,=\, \alpha(z) \,=\, \Big\langle f(z), \frac{\operatorname{Im} m(z)}{|m(z)|} \Big\rangle \,\leq\, \varepsilon_*\,,$$

then the operator $B = B(z)$, defined in (8.37), has a unique single eigenvalue $\beta = \beta(z)$ of smallest modulus, so that $|\beta'| - |\beta| \gtrsim 1$, $\forall\, \beta' \in \operatorname{Spec}(B)\setminus\{\beta\}$. The corresponding eigenfunction $b = b(z)$, satisfying $Bb = \beta b$, has the properties

$$\langle f, b\rangle \,=\, 1\,, \qquad \text{and} \qquad |b_x| \,\sim\, 1\,, \quad \forall x \in \mathfrak{X}\,. \tag{11.6}$$

The spectral projector $P = P(z) : \mathscr{B} \to \operatorname{Span}\{b(z)\}$, corresponding to β, is given by

$$Pw \,=\, \frac{\langle \bar{b}, w\rangle}{\langle b^2\rangle}\, b\,. \tag{11.7}$$

Denoting, $Q := 1 - P$, we have

$$\|B^{-1}\|_{\mathscr{B}\to\mathscr{B}} \,\lesssim\, \alpha^{-2}, \qquad \text{but} \qquad \|B^{-1}Q\|_{\mathscr{B}\to\mathscr{B}} + \|(B^{-1}Q)^*\|_{\mathscr{B}\to\mathscr{B}} \,\lesssim\, 1\,, \tag{11.8}$$

where $(B^{-1}Q)^*$ is the L^2-adjoint of $B^{-1}Q$.

Furthermore, the following expansions in $\eta = \operatorname{Im} z$ and α apply:

$$B \,=\, 1 \,-\, F \,-\, 2\mathrm{i}pf\,\alpha - 2f^2\alpha^2 \,+\, \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha^3 + \eta)\,, \tag{11.9a}$$
$$\beta \,=\, \langle f|m|\rangle\frac{\eta}{\alpha} \,-\, \mathrm{i}2\sigma\,\alpha \,+\, 2(\psi - \sigma^2)\,\alpha^2 \,+\, \mathcal{O}(\alpha^3 + \eta)\,, \tag{11.9b}$$
$$b \,=\, f \,+\, \mathrm{i}2(1-F)^{-1}Q^{(0)}(pf^2)\,\alpha \,+\, \mathcal{O}_{\mathscr{B}}(\alpha^2 + \eta)\,. \tag{11.9c}$$

If $z \in \mathbb{R}$, then the ratio η/α is defined through its limit $\eta \downarrow 0$. The real valued auxiliary functions $\sigma = \sigma(z)$ and $\psi = \psi(z) \geq 0$ in (11.9), are defined by

$$\sigma \,:=\, \langle pf^3\rangle \qquad \text{and} \qquad \psi \,:=\, \mathcal{D}(pf^2)\,, \tag{11.10}$$

where the sign function $p = p(z)$, and the positive quadratic form $\mathcal{D} = \mathcal{D}(\,\bullet\,; z)$, are given by

$$p \,:=\, \operatorname{sign}\operatorname{Re} m \tag{11.11}$$

and

$$\mathcal{D}(w) \,:=\, \Big\langle Q^{(0)}w, \Big[(1 + \|F\|_{\mathrm{L}^2\to\mathrm{L}^2})(1-F)^{-1} - 1\Big]Q^{(0)}w\Big\rangle \,\geq\, \frac{\operatorname{Gap}(F)}{2}\,\|Q^{(0)}w\|_2^2\,, \tag{11.12}$$

respectively. The orthogonal projector $Q^{(0)} = Q^{(0)}(z) = 1 - f(z)\langle f(z), \bullet\,\rangle$, is the leading order term of Q, i.e., $Q = Q^{(0)} + \mathcal{O}_{\mathrm{L}^2\to\mathrm{L}^2}(\alpha)$. Furthermore, $\operatorname{Gap}(F) \sim 1$.

Finally, $\lambda(z) = \|F(z)\|_{\mathrm{L}^2\to\mathrm{L}^2}$, $\beta(z), \sigma(z), \psi(z)$, as well as the vectors $f(z), b(z) \in \mathscr{B}$, are all uniformly $1/3$-Hölder continuous functions of z on connected components of the domain

$$\big\{z \in \overline{\mathbb{H}} : \alpha(z) \leq \varepsilon_*,\, |z| \leq 4\big\}\,.$$

The function p stays constant on these connected components.

The notation $\mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\varphi)$ for some non negative φ means that the expression is bounded by φ, up to a constant, after taking the operator norm for operators from $\mathscr{B}$ to $\mathscr{B}$. The notation $\mathcal{O}_{\mathrm{L}^2\to\mathrm{L}^2}(\varphi)$ is interpreted analogously.

Although, P is not an orthogonal projection (unless $\bar{b} = b$), it follows from (11.6) and (11.7),

$$\|P\|_{\mathscr{B}\to\mathscr{B}},\ \|P^*\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1\,. \tag{11.13}$$

Here $P^* = \langle b, \bullet\rangle/\langle \bar{b}^2\rangle$ is the Hilbert space adjoint of P.

Proof. Recall that $\sin a = (\operatorname{Im} m)/|m|$ (cf. (8.37)), and

$$B \,=\, \mathrm{e}^{-\mathrm{i}2a} - F \,=\, (1-F) + D\,, \tag{11.14}$$

where D is the multiplication operator

$$D = -\mathrm{i}2\cos a\,\sin a - 2\sin^2 a\,. \tag{11.15}$$

From the definition of $\alpha = \langle f \operatorname{Im} m/|m|\rangle$, and $f, |m| \sim 1$, we see that $|\sin a| \sim \alpha$, and thus

$$\|D\|_{\mathrm{L}^2\to\mathrm{L}^2} + \|D\|_{\mathscr{B}\to\mathscr{B}} \leq C_0\alpha\,. \tag{11.16}$$

The formula (11.9a) for B follows by expanding D in α and η using the representations (8.51) and (8.54) of $\sin a$ and $\cos a$, respectively. In particular, from (8.52) we know that $\|t\|_{\mathscr{B}} \lesssim 1$, and thus $\sin a = \alpha f + \mathcal{O}_{\mathscr{B}}(\eta)$.

Let us first consider the operators as acting on the space L^2. By Proposition 8.2 the operator $1-F$ has an isolated single eigenvalue of smallest modulus equal to

$$1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \,=\, \frac{\eta}{\alpha}\langle |m| f\rangle\,, \tag{11.17}$$

and the L^2-spectrum of $1-F$ lies inside the set

$$\mathbb{L} \,:=\, \big\{1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\big\} \cup \big[1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} + \mathrm{Gap}(F),\, 2\big]\,. \tag{11.18}$$

Here the upper spectral gap of F satisfies $\mathrm{Gap}(F) \sim 1$ by (iv) of Proposition 8.2.

The properties of β and b, etc., are deduced from the resolvent of B by using the analytic perturbation theory (cf. Chapter 7 of [44]). To this end denote $R(\zeta) := (1-F-\zeta)^{-1}$, so that

$$(B-\zeta)^{-1} \,=\, (1 + R(\zeta)D)^{-1}R(\zeta)\,.$$

We will now bound $R(\zeta) = -(\widehat{F}(|m|) - (1-\zeta))^{-1}$ as an operator on $\mathscr{B}$, using the property (8.26) of the resolvent of the F-like operators $\widehat{F}$ (cf. (8.21))

$$\|R(\zeta)\|_{\mathscr{B}\to\mathscr{B}} \,\leq\, \frac{1 + \Phi^2\|R(\zeta)\|_{\mathrm{L}^2\to\mathrm{L}^2}}{|\zeta - 1|}\,. \tag{11.19}$$

Thus there exists a constant $\delta \sim 1$,

$$\|R(\zeta)\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1\,, \qquad \operatorname{dist}(\zeta, \mathbb{L}) \geq \delta\,.$$

Here we have used the fact that the set $\mathbb{L}$ contains both the L^2-spectrum of $1-F$, and the point $\zeta = 1$. Thus (11.19) shows that $\mathbb{L}$ contains also the $\mathscr{B}$-spectrum of $1-F$.

By requiring ε_* to be sufficiently small it follows from (11.16) that $\|(1+R(\zeta)D)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$ provided ζ is at least a distance δ away from $\mathbb{L}$, and thus

$$\|(B-\zeta)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1\,, \qquad \mathrm{dist}(\zeta,\mathbb{L}) \geq \delta\,. \tag{11.20}$$

By (iv) of Proposition 8.2 we see that $\mathrm{Gap}(F) \gtrsim 1$. By taking ε_* sufficiently small the perturbation $\|D\|_{\mathscr{B}\to\mathscr{B}}$ becomes so small that we may take $\delta \leq \mathrm{Gap}(F)/3$. It then follows that the eigenvalue β is separated from the rest of the $\mathscr{B}$-spectrum of B by a gap of size $\delta \sim 1$.

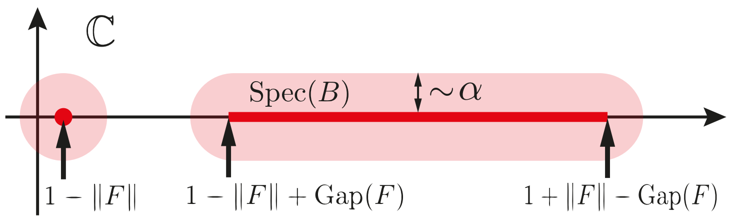

Figure 11.1: The spectrum of $1-F$ lies inside the union of an interval with one isolated point. The perturbation B of $1-F$ has spectrum in the indicated area.

Knowing that β is separated from the rest of the spectrum by a distance $\delta \sim 1$, the standard resolvent contour integral representation formulas (cf. (11.5)) imply that $\|b\|_{\mathscr{B}} \lesssim 1$ and $\|P\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$, $\|B^{-1}Q\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$, etc., provided the threshold $\varepsilon_* \sim 1$ for α is sufficiently small. Similar bounds hold for the adjoints, e.g., $\|(B^{-1}Q)^*\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$. For an illustration of how the spectrum of the perturbation B differs from the spectrum of $1-F$, see Figure 11.1.

Setting $\beta^{(0)} = 1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}$ and $b^{(0)} = f$, the formulas (11.9b) and (11.9c) amount to determining the subleading order terms of

$$\begin{aligned} \beta &= \beta^{(0)} + \beta^{(1)}\alpha + \beta^{(2)}\alpha^2 + \mathcal{O}(\alpha^3+\eta) \\ b &= b^{(0)} + b^{(1)}\alpha + \mathcal{O}_{\mathscr{B}}(\alpha^2+\eta)\,, \end{aligned} \tag{11.21}$$

using the standard perturbation formulas. Writing (11.9a) as,

$$B = B^{(0)} + \alpha\, B^{(1)} + \alpha^2 B^{(2)} + \dots\,,$$

with $B^{(0)} = 1-F$, $B^{(1)} = -2\mathrm{i}pf$, $B^{(2)} := -2f^2$, we obtain

$$\begin{aligned} \beta^{(1)} &= \big\langle b^{(0)}, B^{(1)}b^{(0)}\big\rangle = -\mathrm{i}2\langle pf^3\rangle\,, \\ \beta^{(2)} &= \big\langle b^{(0)}, B^{(2)}b^{(0)}\big\rangle - \Big\langle b^{(0)},\, B^{(1)}Q^{(0)}(B^{(0)}-\beta^{(0)})^{-1}Q^{(0)}B^{(1)}b^{(0)}\Big\rangle \\ &= 2\big(1+\|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\big)\Big\langle Q^{(0)}(pf^2), (1-F)^{-1}Q^{(0)}(pf^2)\Big\rangle - 2\big\langle f^4\big\rangle + \mathcal{O}\Big(\frac{\eta}{\alpha}\Big)\,. \end{aligned} \tag{11.22}$$

These expressions match (11.9). To get the last expression of $\beta^{(2)}$ in (11.22) we have used $\|Q^{(0)}R(\zeta)Q^{(0)}\|_{\mathrm{L}^2\to\mathrm{L}^2} \sim 1$, $\zeta \in [0,\beta^{(0)}]$, and $\beta^{(0)} \sim \eta/\alpha$, to approximate

$$(B^{(0)}-\beta^{(0)})^{-1}Q^{(0)} = (1-F)^{-1}Q^{(0)} + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}\Big(\frac{\eta}{\alpha}\Big)\,.$$

The formula (11.9c) follows similarly

$$b^{(1)} = -(B^{(0)}-\beta^{(0)})^{-1}Q^{(0)}B^{(1)}b^{(0)} = \mathrm{i}2(1-F)^{-1}Q^{(0)}(pf^2) + \mathcal{O}_{\mathscr{B}}\Big(\frac{\eta}{\alpha}\Big)\,.$$

In order, to see that $\psi \geq 0$, we use $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \leq 1$, to estimate

$$(1+\|F\|_{\mathrm{L}^2\to\mathrm{L}^2})(1-F)^{-1} \geq 1 + \frac{\mathrm{Gap}(F)}{2}\,.$$

This yields the estimate in (11.12).

It remains to prove the 1/3-Hölder continuity of the various quantities in the lemma. To this end we write

$$B(z) = \mathrm{e}^{-2a(z)} - \widehat{F}(|m(z)|)\,, \tag{11.23}$$

where the operator $\widehat{F}(r) : \mathscr{B} \to \mathscr{B}$ is defined in (8.21). Since $\|S\|_{\mathscr{B}\to\mathscr{B}} \le \|S\|_{\mathrm{L}^2\to\mathscr{B}} \sim 1$ it is easy to see from (8.21) that the map $r \mapsto \widehat{F}(r)$ is uniformly continuous when restricted on the domain of arguments $r \in \mathscr{B}_+$ satisfying $c/\Phi \le r_x \le \Phi$. Furthermore, the exponent $\mathrm{e}^{-\mathrm{i}2a} = (|m|/m)^2$, has the same regularity as m because $|m| \sim 1$. Since $m(z)$ is uniformly 1/3-Hölder continuous in z (cf. (10.2)) we thus have

$$\|B(z') - B(z)\|_{\mathscr{B}\to\mathscr{B}} \lesssim |z'-z|^{1/3}\,, \tag{11.24}$$

for any sufficiently close points z and z'. The resolvent $(B(z)-\zeta)^{-1}$ inherits this regularity in z.

The continuity of $\beta(z), b(z), P(z)$ in z is proven by representing them as contour integrals of the resolvent $(B(z)-\zeta)^{-1}$ around a contour enclosing the isolated eigenvalue $\beta(z)$. The functions σ and ψ inherit the 1/3-Hölder regularity from their building blocks, $1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}$, f, $Q^{(0)}$, and the function p. The continuity of the first three follows similarly as that of β, b and Q, using the continuity of the resolvent of $1 - F(z)$ in z. Also the continuity of the largest eigenvalue $\lambda(z)$ of $F(z)$ is proven this way. In particular, we see from (11.17) that the limit $\eta/\alpha(z)$ exists as z approaches the real line.

The function $p(z) = \operatorname{sign}\operatorname{Re} m(z)$ in z, on the other hand, is handled differently. We show that if $\varepsilon_* > 0$ sufficiently small then the restriction of p to a connected component J of the set $\{z : \alpha(z) \le \varepsilon_*\}$ is a constant, i.e., $p(z') = p(z)$, for any $z, z' \in J$. Indeed, since $\inf_x |m_x(z)| \ge c_0$, and $\sup_x \operatorname{Im} m_x(z) \le C_1\varepsilon_*$, for some $c_0, C_1 \sim 1$, we get

$$(\operatorname{Re} m_x)^2 = |m_x|^2 - (\operatorname{Im} m_x)^2 \ge c_0^2 - (C_1\varepsilon_*)^2\,, \qquad \forall x \in \mathfrak{X}\,. \tag{11.25}$$

Clearly, for a sufficiently small ε_* the real part $\operatorname{Re} m_x(z)$ cannot vanish. Consequently, the continuity of $m : \overline{\mathbb{H}} \to \mathscr{B}$ means that the components $p_x(z) = \operatorname{sign}\,\operatorname{Re} m_x(z) \in \{-1,+1\}$, may change values only when $\alpha(z) > \varepsilon_*$.

The explicit representation (11.7) of the spectral projector P follows from an elementary property of compact integral operators: If the integral kernel $(T^*)_{xy}$ of the Hilbert-space adjoint of an operator $T : \mathrm{L}^2 \to \mathrm{L}^2$, defined by $(Tw)_y = \int T_{xy} w_y \pi(\mathrm{d}y)$, has the symmetry $(T^*)_{xy} = \overline{T_{xy}}$, then the right and left eigenvectors v and v' corresponding to the right and left eigenvalues λ and $\overline{\lambda}$, respectively, are also related by the simple component wise complex conjugation: $(v')_x = \overline{v_x}\,$. □

11.2 Cubic equation

We are now ready to show that the projection of u in the b-direction satisfies a cubic equation (up to the leading order) provided α and η are sufficiently small. Recall, that T^* denotes the L^2-adjoint of a linear operator T on L^2.

Proposition 11.2 (General cubic equation). *Suppose $g \in \mathscr{B}$ solves the perturbed QVE (8.34) at $z \in \overline{\mathbb{H}}$, with $|z| \le 4$. Set*

$$u := \frac{g-m}{|m|}\,, \tag{11.26a}$$

and define $\Theta \in \mathbb{C}$ *and* $r \in \mathscr{B}$ *by*

$$\Theta := \frac{\langle \bar{b}, u \rangle}{\langle b^2 \rangle} \qquad \textit{and} \qquad r := Qu\,. \tag{11.26b}$$

There exists $\varepsilon_* \sim 1$ *such that if*

$$\langle v \rangle \le \varepsilon_*\,, \qquad \textit{and} \qquad \|g - m\|_{\mathscr{B}} \le \varepsilon_*\,, \tag{11.27}$$

then the following holds: The component r *is dominated by* d *and* Θ*,*

$$r \;=\; Rd \,+\, \mathcal{O}_{\mathscr{B}}\big(|\Theta|^2 + \|d\|_{\mathscr{B}}^2\big)\,, \tag{11.28}$$

where $R = R(z)$ *denotes the bounded linear operator* $w \mapsto B^{-1}Q(|m|w)$*, that satisfies*

$$\|R\|_{\mathscr{B}\to\mathscr{B}} + \|R^*\|_{\mathscr{B}\to\mathscr{B}} \;\sim\; 1\,. \tag{11.29}$$

The coefficient Θ *in* (11.26) *is a root of the complex cubic polynomial,*

$$\mu_3\Theta^3 \,+\, \mu_2\Theta^2 \,+\, \mu_1\Theta \,+\, \langle |m|\bar{b}, d\rangle \;=\; \kappa(u,d)\,, \tag{11.30}$$

perturbed by the function $\kappa(u,d)$ *of sub-leading order. This perturbation satisfies*

$$|\kappa(u,d)| \;\lesssim\; |\Theta|^4 \,+\, \|d\|_{\mathscr{B}}^2 \,+\, |\Theta|\,|\langle e, d\rangle|\,, \tag{11.31}$$

where $e : \overline{\mathbb{H}} \to \mathscr{B}$ *is a uniformly bounded function,* $\|e(z)\|_{\mathscr{B}} \lesssim 1$*, determined by* S*. The coefficient functions* $\mu_k : \overline{\mathbb{H}} \to \mathbb{C}$ *are determined by* S *alone, and satisfy*

$$\mu_3 \;:=\; \Big(1 - \langle f\,|m|\rangle\frac{\eta}{\alpha}\Big)\,\psi \,+\, \mathcal{O}(\alpha) \tag{11.32a}$$

$$\mu_2 \;:=\; \Big(1 - \langle f\,|m|\rangle\frac{\eta}{\alpha}\Big)\,\sigma \,+\, \mathrm{i}(3\psi \,-\, \sigma^2)\,\alpha \,+\, \mathcal{O}\big(\alpha^2 + \eta\big) \tag{11.32b}$$

$$\mu_1 \;:=\; -\langle f\,|m|\rangle\frac{\eta}{\alpha} \,+\, \mathrm{i}2\sigma\,\alpha \,-\, 2(\psi - \sigma^2)\,\alpha^2 \,+\, \mathcal{O}\big(\alpha^3 + \eta\big)\,. \tag{11.32c}$$

If $z \in \mathbb{R}$*, then the ratio* η/α *is defined through its limit as* $\eta \to 0$*.*

Finally, the cubic is **stable** *in the sense that*

$$|\mu_3(z)| + |\mu_2(z)| \;\sim\; 1\,. \tag{11.33}$$

Note that from (11.26b) and (11.7) we see that Θ is just the component of u in the one-dimensional subspace spanned by b, i.e, $Pu = \Theta b$. From (11.26) and (11.13) we see that $|\Theta| \le C_1\varepsilon_*$ is a small parameter along with α and η. Therefore, we needed to expand μ_1 to a higher order than μ_2, which is in turn expanded to a higher order than μ_3 in the variables α and η in (11.32).

Proof. The proof is split into two separate parts. First, we derive formulas for the μ_k's in terms of B, β and b (cf. (11.43) below). Second, we use the formulas (11.9) from Lemma 11.1 to expand μ_k's further in α and η.

First, we write the equation (8.36) in the form

$$Bu = \mathcal{A}(u,u) + |m|(1 + \mathrm{e}^{-\mathrm{i}a}u)\,d\,, \tag{11.34}$$

where $a = a(z) := \arg\, m(z)$, and the symmetric bilinear map $\mathcal{A} : \mathscr{B}^2 \to \mathscr{B}$, is defined by

$$\mathcal{A}_x(q,w) := \tfrac{1}{2}\mathrm{e}^{-\mathrm{i}a_x}\big(q_x\,(Fw)_x + (Fq)_x w_x\big)\,.$$

Clearly, $\|\mathcal{A}(q,w)\|_{\mathscr{B}} \lesssim \|q\|_{\mathscr{B}}\|w\|_{\mathscr{B}}$, since $\|F\|_{\mathscr{B}\to\mathscr{B}} \le \|m\|_{\mathscr{B}}^2 \lesssim 1$. Applying Q on (11.34) gives

$$r = B^{-1}Q\mathcal{A}(u,u) + B^{-1}Q\big[\,|m|(1+\mathrm{e}^{-\mathrm{i}a}u)d\,\big]\,. \tag{11.35}$$

From Lemma 11.1 we know that $\|QB^{-1}Q\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$, and hence the boundedness of $\mathcal{A}$ implies

$$\|B^{-1}Q\mathcal{A}(u,u)\|_{\mathscr{B}} \;\lesssim\; \|u\|_{\mathscr{B}}^2 \;\lesssim\; |\Theta|^2 + \|r\|_{\mathscr{B}}^2\,.$$

From the boundedness of the projections (11.13)

$$\|r\|_{\mathscr{B}} \;=\; \|Qu\|_{\mathscr{B}} \;\lesssim\; \|u\|_{\mathscr{B}} \;\le\; \frac{\|g-m\|_{\mathscr{B}}}{\inf_x |m_x|} \;\lesssim\; \varepsilon_*\,,$$

where in the second to last inequality we have used $|m|\sim 1$. Plugging this back into (11.35), we find

$$\|r\|_{\mathscr{B}} \;\le\; C_0\big(\,|\Theta|^2 + \varepsilon_*\,\|r\|_{\mathscr{B}} + \|d\|_{\mathscr{B}}\big)\,.$$

Now we require ε_* to be so small that $2C_0\varepsilon_* \le 1$, and get

$$\|r\|_{\mathscr{B}} \;\lesssim\; |\Theta|^2 + \|d\|_{\mathscr{B}}\,. \tag{11.36}$$

Applying this on the right hand side of $u = \Theta\, b + r$ yields a uniform bound on u,

$$\|u\|_{\mathscr{B}} \;\lesssim\; |\Theta| + \|d\|_{\mathscr{B}}\,. \tag{11.37}$$

Using the bilinearity and the symmetry of $\mathcal{A}$ we decompose r into three parts

$$r \;=\; B^{-1}Q\mathcal{A}(b,b)\,\Theta^2 \;+\; Rd \;+\; \widetilde{r}\,, \tag{11.38}$$

where we have identified the operator R from (11.28), and introduced the subleading order part,

$$\begin{aligned} \widetilde{r} \;&:=\; 2B^{-1}Q\mathcal{A}(b,r)\,\Theta + B^{-1}Q\mathcal{A}(r,r) + B^{-1}Q(|m|\mathrm{e}^{-\mathrm{i}a}ud\,) \\ &=\; \mathcal{O}_{\mathscr{B}}\big(\,|\Theta|^3 + |\Theta|\|d\|_{\mathscr{B}} + \|d\|_{\mathscr{B}}^2\big)\,. \end{aligned} \tag{11.39}$$

Applying the last estimate in (11.38) yields (11.28). We know that $B^{-1}Q$ is bounded as an operators on $\mathscr{B}$ from (11.8). A direct calculation using (11.7) shows that also its L^2-Hilbert-space adjoint satisfies a similar bound, $\|(B^{-1}Q)^*\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$. From this and $\|m\|_{\mathscr{B}} \lesssim 1$ the bound (11.29) follows.

From (11.7) we see that applying $\langle \overline{b}, \bullet\,\rangle$ to (11.34) corresponds to projecting onto the b-direction

$$\begin{aligned} \beta\langle b^2\rangle\,\Theta \;&=\; \langle \overline{b}, \mathcal{A}(b,b)\rangle\,\Theta^2 \;+\; 2\,\langle \overline{b}, \mathcal{A}(b,r)\rangle\,\Theta \;+\; \langle \overline{b}, \mathcal{A}(r,r)\rangle \;+\; \big\langle \overline{b}, |m|(1+\mathrm{e}^{-\mathrm{i}a}u)d\big\rangle \\ &=\; \langle b\mathcal{A}(b,b)\rangle\,\Theta^2 \;+\; 2\big\langle b\mathcal{A}(b, B^{-1}Q\mathcal{A}(b,b))\big\rangle\,\Theta^3 \;+\; \langle b|m|d\rangle \;+\; \kappa(u,d)\,, \end{aligned} \tag{11.40}$$

where the cubic term corresponds to the part $B^{-1}Q\mathcal{A}(b,b)\,\Theta^2$ of r in (11.38), while the other parts of $\langle \overline{b}, \mathcal{A}(b,r)\rangle\,\Theta$, have been absorbed into the remainder term, alongside other small terms:

$$\begin{aligned} \kappa(u,d) \;&:=\; 2\,\langle b\mathcal{A}(b, Rd+\widetilde{r}\,)\rangle\,\Theta \;+\; \langle b\mathcal{A}(r,r)\rangle \;+\; \big\langle b|m|\mathrm{e}^{-\mathrm{i}a}ud\big\rangle \\ &=\; \langle e,\, d\rangle\,\Theta \;+\; \mathcal{O}\big(|\Theta|^4 + \|d\|_{\mathscr{B}}^2\big)\,, \end{aligned} \tag{11.41}$$

where in the second line we have defined $e\in\mathscr{B}$ in (11.31) such that

$$\langle e, w\rangle \;:=\; 2\,\langle b\mathcal{A}(b,Rw)\rangle + \langle b^2|m|\mathrm{e}^{-\mathrm{i}a}w\rangle\,, \qquad \forall\, w\in \mathrm{L}^2\,.$$

For the error estimate in (11.41) we have also used (11.36), (11.37), and $\|b\|_{\mathscr{B}} \sim 1$. This completes the proof of (11.31).

From the definitions of $\mathcal{A}$, B, b and β, it follows

$$\begin{aligned} \mathcal{A}(b,b) &= \mathrm{e}^{-\mathrm{i}a} bFb \;=\; \mathrm{e}^{-\mathrm{i}a} b(\mathrm{e}^{-\mathrm{i}2a} - B)b \;=\; (\mathrm{e}^{-\mathrm{i}3a} - \beta \mathrm{e}^{-\mathrm{i}a})\, b^2 \\ 2\mathcal{A}(b,w) &= \mathrm{e}^{-\mathrm{i}a}\big(bFw - w(\mathrm{e}^{-\mathrm{i}2a} - \beta)b\big) \;=\; b\,\mathrm{e}^{-\mathrm{i}a}(\mathrm{e}^{-\mathrm{i}2a} + F - \beta)w\,. \end{aligned} \tag{11.42}$$

Using these formulas in (11.40) we see that the cubic (11.30) holds with the coefficients,

$$\mu_3 \;=\; \Big\langle b^2\, \mathrm{e}^{-\mathrm{i}a}(\mathrm{e}^{-\mathrm{i}2a} + F - \beta)\, B^{-1} Q\big[b^2 \mathrm{e}^{-\mathrm{i}a}(\mathrm{e}^{-\mathrm{i}2a} - \beta)\big]\Big\rangle \tag{11.43a}$$

$$\mu_2 \;=\; \big\langle (\mathrm{e}^{-\mathrm{i}3a} - \beta \mathrm{e}^{-\mathrm{i}a}) b^3 \big\rangle \tag{11.43b}$$

$$\mu_1 \;=\; -\beta\,\langle b^2\rangle \tag{11.43c}$$

that are determined by S and z alone.

The final expressions (11.32) follow from these formulas by expanding B, β and b, w.r.t. the small parameters α and η using the expansions (11.9). Let us write

$$w := (1-F)^{-1} Q^{(0)}(pf^2)\,,$$

so that $b = f + (\mathrm{i}2w)\alpha + \mathcal{O}_{\mathscr{B}}(\alpha^2+\eta)$, and $\langle f, w\rangle = 0$. Using (8.51) and (8.54) we also obtain a useful representation $\mathrm{e}^{-\mathrm{i}a} = p - \mathrm{i}\,f\alpha + \mathcal{O}_{\mathscr{B}}(\alpha^2+\eta)$.

First we expand the coefficient μ_1. Using $\langle f^2\rangle = 1$ and $\langle f, w\rangle = 0$ we obtain $\langle b^2\rangle = 1 + \mathcal{O}(\alpha^2+\eta)$. Hence, only the expansion of β contributes at the level of desired accuracy to μ_1,

$$\mu_1 \;=\; -\beta\langle b^2\rangle \;=\; -\beta \,+\, \mathcal{O}(\alpha^3+\eta) \;=\; -\langle f|m|\rangle\frac{\eta}{\alpha} + \mathrm{i}2\sigma\alpha - 2(\psi-\sigma^2)\alpha^2 \,+\, \mathcal{O}(\alpha^3+\eta)\,.$$

Now we expand the second coefficient, μ_2. Let us first write

$$\mu_2 \;=\; \big\langle (\mathrm{e}^{-\mathrm{i}3a} - \beta\mathrm{e}^{-\mathrm{i}a})b^3\big\rangle \;=\; \big\langle (\mathrm{e}^{-\mathrm{i}a}b)^3\big\rangle - \beta\big\langle \mathrm{e}^{-\mathrm{i}a} b^3\big\rangle\,. \tag{11.44}$$

Using the expansions we see that $\mathrm{e}^{-\mathrm{i}a}b = pf + \mathrm{i}(2pw - f^2)\alpha + \mathcal{O}_{\mathscr{B}}(\alpha^2+\eta)$, and thus, taking this to the third power, we find $(\mathrm{e}^{-\mathrm{i}a}b)^3 = pf^3 + \mathrm{i}3(2pf^2w - f^4) + \mathcal{O}_{\mathscr{B}}(\alpha^2+\eta)$. Consequently,

$$\begin{aligned} \big\langle (\mathrm{e}^{-\mathrm{i}a}b)^3\big\rangle &= \langle pf^3\rangle + \mathrm{i}3\big[2\langle pf^2w\rangle - \langle f^4\rangle\big]\,\alpha + \mathcal{O}(\alpha^2+\eta) \\ &= \sigma + \mathrm{i}3(\psi-\sigma^2)\alpha + \mathcal{O}(\alpha^2+\eta)\,. \end{aligned} \tag{11.45}$$

In order to obtain expressions in terms of σ and $\psi = \mathcal{D}(pf^2)$, where the bilinear positive form $\mathcal{D}$ is defined in (11.12), we have used

$$2\big\langle pf^2 w\big\rangle \;=\; (1+\|F\|_{\mathrm{L}^2\to\mathrm{L}^2})\big\langle Q^{(0)}(pf^2), (1-F)^{-1}Q^{(0)}(pf^2)\big\rangle + \mathcal{O}(\eta/\alpha)\,,$$

as well as the following consequence of $P^{(0)}(pf^2) = \sigma f$ and $\|f\|_2 = 1$:

$$\langle f^4\rangle \;=\; \|pf^2\|_2^2 \;=\; \|P^{(0)}(pf^2)\|_2^2 + \|Q^{(0)}(pf^2)\|_2^2 \;=\; \sigma^2 + \big\langle Q^{(0)}(pf^2), Q^{(0)}(pf^2)\big\rangle\,. \tag{11.46}$$

The expansion of the last term of (11.44) is easy since only β has to be expanded beyond the leading order. Indeed, directly from (11.9b) we obtain

$$\beta\big\langle \mathrm{e}^{-\mathrm{i}a}b^3\big\rangle = \Big(\langle f|m|\rangle\frac{\eta}{\alpha} - \mathrm{i}2\sigma\alpha + \mathcal{O}(\alpha^2+\eta)\Big)\Big(\langle pf^3\rangle + \mathcal{O}(\alpha+\eta)\Big) = -\mathrm{i}2\sigma^2\alpha + \langle f|m|\rangle\frac{\eta}{\alpha}\sigma + \mathcal{O}\big(\alpha^2+\eta\big).$$

Plugging this together with (11.45) into (11.44) yields the desired expansion of μ_2.

Finally, μ_3, is expanded. By the definitions and the identity (8.18) for $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2}$ we have

$$\mathrm{e}^{-\mathrm{i}2a} + F - \beta \;=\; 2 - \langle f|m|\rangle \frac{\eta}{\alpha} - B + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha) \;=\; 1 + \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} - B + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha)\,.$$

Recalling $\|B^{-1}Q\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$ and $\eta \lesssim \alpha$, we thus obtain

$$(\mathrm{e}^{-\mathrm{i}2a} + F - \beta)\, B^{-1}Q \;=\; (1 + \|F\|_{\mathrm{L}^2\to\mathrm{L}^2})\, B^{-1}Q - Q + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha)\,. \tag{11.47}$$

Directly from the definition (11.7) of $P = 1 - Q$, we see that $Q = Q^{(0)} + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha)$. Thus,

$$BQ \;=\; (1-F)Q^{(0)} + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha)\,.$$

Using the general identity $(A+D)^{-1} = A^{-1} - A^{-1}D(A+D)^{-1}$, with $A := (1-F)Q^{(0)}$ and $A + D := BQ$, yields

$$B^{-1}Q \;=\; (1-F)^{-1}Q^{(0)} + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha)\,, \tag{11.48}$$

since $B^{-1}Q$ and $(1-F)^{-1}Q^{(0)}$ are both $\mathcal{O}_{\mathscr{B}\to\mathscr{B}}(1)$. By applying (11.48) in (11.47) we get

$$(\mathrm{e}^{-\mathrm{i}2a} + F - \beta)\,(QBQ)^{-1} \;=\; Q^{(0)}\big[(1+\|F\|_{\mathrm{L}^2\to\mathrm{L}^2})\,(1-F)^{-1} - 1\big]\,Q^{(0)} + \mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\alpha)\,.$$

Using this in the first formula of μ_3 below yields

$$\begin{aligned}
\mu_3 \;&=\; \Big\langle b^2\,\mathrm{e}^{-\mathrm{i}a}(\mathrm{e}^{-\mathrm{i}2a} + F - \beta)\, B^{-1}Q\big(b^2\mathrm{e}^{-\mathrm{i}a}(\mathrm{e}^{-\mathrm{i}2a} - \beta)\big)\Big\rangle \\
&=\; \Big(1 - \langle f|m|\rangle\frac{\eta}{\alpha}\Big)\Big\langle Q^{(0)}(pf^2), \big[(1+\|F\|_{\mathrm{L}^2\to\mathrm{L}^2})\,(1-F)^{-1} - 1\big]\,Q^{(0)}(pf^2)\Big\rangle + \mathcal{O}(\alpha)\,,
\end{aligned}$$

which equals the second expression (11.32a) because the first term above is $\mathcal{D}(pf^2)$.

Finally, we show that $|\mu_2| + |\mu_3| \sim 1$. From the expansion of μ_2, we get

$$|\mu_2| \;=\; \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}\,|\sigma| + \mathcal{O}(\alpha) \;\gtrsim\; |\sigma| + \mathcal{O}(\alpha)\,.$$

Similarly, we estimate from below $|\mu_3| \gtrsim \psi + \mathcal{O}(\alpha)$. Therefore, we find that

$$|\mu_3| + |\mu_2|^2 \;\gtrsim\; \psi + \sigma^2 + \mathcal{O}(\alpha)\,.$$

We will now show that $\psi + \sigma^2 \gtrsim 1$, which implies $|\mu_2|^2 + |\mu_3| \gtrsim 1$, provided the upper bound ε_* of α is small enough. Indeed, from the lower bound (11.12) on the quadratic form $\mathcal{D}$, $\mathrm{Gap}(F) \sim 1$ and the identity $|\sigma| = |\langle f, pf^2\rangle| = \|P^{(0)}(pf^2)\|_2$ we conclude that

$$\psi + \sigma^2 \;\geq\; \frac{\mathrm{Gap}(F)}{2}\,\|Q^{(0)}(pf^2)\|_2^2 + \|P^{(0)}(pf^2)\|_2^2 \;\gtrsim\; \|pf^2\|_2^2\,. \tag{11.49}$$

Since $\inf_x f_x \sim 1$ and $|p| = 1$ it follows that $\|pf^2\|_2 \sim 1$. □

12 Behaviour of generating density where it is small

In this section we prove Theorem 6.4. We will assume that S satisfies **A1-5.** and $|\!|\!|m|\!|\!|_{\mathbb{R}} \leq \Phi < \infty$ as in the previous section. In particular, we have $v_x \sim \langle v\rangle$ and thus the support of the components of the generating densities satisfy $\operatorname{supp} v = \operatorname{supp}\langle v\rangle$ (cf. Definition 7.4). As we are interested in the generating density $\operatorname{Im} m|_{\mathbb{R}}$ we will consider m and all the related quantities as functions on $\mathbb{R}$ instead of on $\mathbb{H}$ or $\overline{\mathbb{H}}$ in this section.

Consider the domain

$$\mathbb{D}_\varepsilon := \big\{\tau \in \operatorname{supp} v : \langle v(\tau)\rangle \le \varepsilon\big\} \tag{12.1}$$

for some sufficiently small threshold $\varepsilon \sim 1$. We decompose it into disjoint connected components labelled by k and pick a local minimum τ_k of $\langle v\rangle$ in each component. Denote by $\mathbb{D}_\varepsilon(\tau_k)$ the component containing τ_k, so that

$$\mathbb{D}_\varepsilon = \bigcup_k \mathbb{D}_\varepsilon(\tau_k)\,.$$

Later on we will see that there are only finitely many components.

The goal is to show (cf. (6.18)) that to leading order we can write

$$v(\tau) = \widehat{v}(\tau) + \mathcal{O}\big(\widehat{v}(\tau)^2\big)\,, \qquad \tau \in \mathbb{D}_\varepsilon\,, \tag{12.2a}$$

where the leading order part factorises on each connected components,

$$\widehat{v}_x(\tau) = v_x(\tau_k) + h_x(\tau_k)\,\Psi(\tau - \tau_k;\tau_k)\,, \qquad \tau \in \mathbb{D}_\varepsilon(\tau_k)\,, \tag{12.2b}$$

with $h_x(\tau_k) \sim 1$ and $\Psi(\omega;\tau_k) \ge 0$. We show that the function $\Psi(\omega;\tau_k)$ determining the shape of $\omega \to \langle v(\tau_k + \omega)\rangle$ is **universal** in the sense that it depends on τ_k only through a single scalar parameter (cf. (6.19)).

Let τ_0 denote one of the minima τ_k. We consider $m(\tau_0 + \omega)$ as the solution of the perturbed QVE (8.34) at $z = \tau_0$ with the scalar perturbation

$$d_x(\omega) := \omega\,, \qquad \forall\, x \in \mathfrak{X}\,, \tag{12.3}$$

and apply Proposition 11.2. The leading order behaviour of $m(\tau_0 + \omega)$ is determined by expressing

$$u(\omega;\tau_0) := \frac{m(\tau_0 + \omega) - m(\tau_0)}{|m(\tau_0)|}\,, \tag{12.4}$$

as a sum of its projections,

$$\Theta(\omega;\tau_0)b(\tau_0) := P(\tau_0)u(\omega;\tau_0) \qquad \text{and} \qquad r(\omega;\tau_0) := Q(\tau_0)u(\omega;\tau_0)\,, \tag{12.5}$$

where $P = P(\tau_0)$ is defined in (11.7) and $Q(\tau_0) = 1 - P(\tau_0)$. The coefficient $\Theta(\omega;\tau_0)$ is then computed as a root of the cubic equation (11.30) corresponding to the scalar perturbation (12.3). Its imaginary part will give $\Psi(\omega,\tau_0)$. Finally, the part $r(\omega;\tau_0)$ is shown to be much smaller than $\Theta(\omega;\tau_0)$ so that it can be considered as an error term. The next lemma collects necessary information needed to carry out this analysis rigorously.

LEMMA 12.1 (Cubic for shape analysis). *There are two constants $\varepsilon_*, \delta \sim 1$, such that if*

$$\tau_0 \in \operatorname{supp} v \qquad \text{and} \qquad \langle v(\tau_0)\rangle \le \varepsilon_*\,, \tag{12.6}$$

holds for some fixed **base point** $\tau_0 \in \operatorname{supp} v$, *then*

$$\Theta(\omega) = \Theta(\omega;\tau_0) = \bigg\langle \frac{b(\tau_0)}{\langle b(\tau_0)^2\rangle}\,\frac{m(\tau_0 + \omega) - m(\tau_0)}{|m(\tau_0)|}\bigg\rangle\,, \tag{12.7}$$

satisfies the perturbed cubic equation

$$\mu_3\Theta(\omega)^3 + \mu_2\Theta(\omega)^2 + \mu_1\Theta(\omega) + \Xi(\omega)\,\omega = 0\,, \qquad |\omega| \le \delta\,. \tag{12.8}$$

The coefficients $\mu_k = \mu_k(\tau_0) \in \mathbb{C}$ are independent of ω and have expansions in α:

$$\mu_3 \;:=\; \psi + \kappa_3\alpha \tag{12.9a}$$
$$\mu_2 \;:=\; \sigma + \mathrm{i}\,(3\psi - \sigma^2)\,\alpha + \kappa_2\alpha^2 \tag{12.9b}$$
$$\mu_1 \;:=\; \mathrm{i}2\sigma\,\alpha - 2(\psi - \sigma^2)\,\alpha^2 + \kappa_1\alpha^3\,, \tag{12.9c}$$

and $\Xi(\omega) = \Xi(\omega;\tau_0) \in \mathbb{C}$ is close to a real constant:

$$\Xi(\omega) \;:=\; \langle f|m|\rangle\,(1 + \kappa_0\alpha + \nu(\omega))\,. \tag{12.10}$$

The scalars $\alpha = \langle f, v/|m|\rangle$, $\sigma = \langle f, pf^2\rangle$ and $\psi = \mathcal{D}(pf^2)$ are defined in (8.19), (11.10) *and* (11.12)*, respectively. They are uniformly $1/3$-Hölder continuous functions of τ_0 on the connected components of the set $\{\tau : \langle v(\tau)\rangle \le \varepsilon_*, |\tau| \le 4\}$. The cubic* (12.8) *is* **stable** *(cf.* (11.33)*) in the sense that*

$$|\mu_3| + |\mu_2| \;\sim\; \psi \,+\, \sigma^2 \;\sim\; 1\,. \tag{12.11}$$

Both the rest term $r(\omega) = r(\omega;\tau_0)$ (cf. (12.5)*) and $\Theta(\omega)$ are differentiable as functions of ω on the domain $\{\omega : \langle v(\tau_0 + \omega)\rangle > 0\}$, and they satisfy:*

$$|\Theta(\omega)| \;\lesssim\; \min\biggl\{\frac{|\omega|}{\alpha^2}\,,\, |\omega|^{1/3}\biggr\} \tag{12.12a}$$
$$\|r(\omega)\|_{\mathscr{B}} \;\lesssim\; |\Theta(\omega)|^2 + |\omega|\,. \tag{12.12b}$$

The constants $\kappa_j = \kappa_j(\tau_0) \in \mathbb{C}$, $j = 0,1,2,3$, and $\nu(\omega) = \nu(\omega;\tau_0) \in \mathbb{C}$ in (12.9) *and* (12.10) *satisfy*

$$|\kappa_0|, \dots, |\kappa_3| \;\lesssim\; 1 \tag{12.13a}$$
$$|\nu(\omega)| \;\lesssim |\Theta(\omega)| + |\omega| \;\lesssim\; |\omega|^{1/3}\,, \tag{12.13b}$$

and $\nu(\omega)$ is $1/3$-Hölder continuous in ω.

Consequently, the leading behaviour of m on $[\tau_0 - \delta, \tau_0 + \delta]$ is determined by $\Theta(\omega;\tau_0)$:

$$m_x(\tau_0 + \omega) \;=\; m_x(\tau_0) \,+\, |m_x(\tau_0)|\,b_x(\tau_0)\,\Theta(\omega;\tau_0) \,+\, \mathcal{O}\bigl(\Theta(\omega;\tau_0)^2 + |\omega|\bigr) \tag{12.14a}$$
$$\;=\; m_x(\tau_0) \,+\, |m_x(\tau_0)|\,f_x(\tau_0)\,\Theta(\omega;\tau_0) \,+\, \mathcal{O}\bigl(\alpha(\tau_0)|\omega|^{1/3} + |\omega|^{2/3}\bigr)\,. \tag{12.14b}$$

All comparison relations hold w.r.t. the model parameters $(\rho, L, \|S\|_{\mathrm{L}^2\to\mathscr{B}}, \Phi)$.

The expansion (6.18) will be obtained by studying the imaginary parts of (12.14). The factorisation (12.2b) corresponds to the factorisation of the second terms on the right hand side of (12.14). In particular, $\Psi(\omega;\tau_k) = \operatorname{Im}\Theta(\omega;\tau_k)$. The universality of the function $\Psi(\omega;\tau_k)$ corresponds to $\Theta(\omega)$ being close to the solution of the **ideal cubic** obtained from (12.8) and by setting $\kappa_1 = \kappa_2 = \kappa_3 = 0$ and $\kappa_0 = \nu(\omega) = 0$ in (12.9) and (12.10), respectively.

Proof of Lemma 12.1. The present lemma is an application of Proposition 11.2 in the case where $z = \tau_0 \in \operatorname{supp} v$ and the perturbation is a real number (12.3). Then the solution to (8.34) is $g = m(\tau_0 + \omega)$. As for the assumptions of Proposition 11.2, we need to verify the second inequality of (11.27), i.e.,

$$\|m(\tau_0 + \omega) - m(\tau_0)\|_{\mathscr{B}} \;\le\; \varepsilon_*\,, \qquad |\omega| \le \delta\,.$$

This follows from the uniform 1/3-Hölder continuity of the solution of the QVE (cf. Theorem 6.2), provided we choose $\delta \sim \varepsilon_*^3$ sufficiently small. By Theorem 6.2 the solution m is also smooth on the set where $\alpha > 0$. By Lemma 11.1 and (11.13) the projectors P and Q are uniformly bounded on the connected components of the set where $\alpha \leq \varepsilon_*$. This boundedness extends to the real line. Since $|m| \sim 1$, the functions $u(\omega)$ and $r(\omega)$ have the same regularity in ω as $m(\tau)$ has in τ. In particular, (12.12a) follows this way (cf. Corollary 10.2) using $\alpha = \alpha(\tau_0) \sim v(\tau_0)$. Lemma 11.1 implies the Hölder regularity of α, σ, ψ. The estimate (12.11) follows from (11.33), provided ε_* is sufficiently small. The a priori bound (12.12b) for r follows from the analogous general estimate (11.28).

The formulas (12.9) for the coefficients μ_k follow from the general formulas (11.32) by letting $\eta = \operatorname{Im} z$ go to zero. The only non-trivial part is to establish

$$\lim_{\eta \to 0} \frac{\eta}{\alpha(\tau_0 + \mathrm{i}\eta)} \;=\; 0\,, \qquad \forall\, \tau_0 \in \operatorname{supp} v\,. \tag{12.15}$$

Since $m(z) \in \mathscr{B}$ is continuous in z, $F(z)$ is also continuous as an operator on L^2. Thus taking the limit $\operatorname{Im} z \to 0$ of the identity (7.37) shows that

$$\frac{v}{|m|} \;=\; F\frac{v}{|m|}\,,$$

since $|m| \sim 1$. If $\operatorname{Re} z = \tau_0$, with $v(\tau_0) \neq 0$, then the vector $v(\tau_0)/|m(\tau_0)| \in \mathrm{L}^2$ is non-zero, and thus an eigenvector of F corresponding to the eigenvalue 1. In particular, we get

$$\|F(\tau_0)\|_{\mathrm{L}^2 \to \mathrm{L}^2} \;=\; 1\,, \qquad \tau_0 \in \operatorname{supp} v\,. \tag{12.16}$$

If $\tau_0 \in \operatorname{supp} v$ is such that $v(\tau_0) = 0$ then (12.16) follows from a limiting argument $\tau \to \tau_0$, with $v(\tau) \neq 0$, and the continuity of F. Comparing (12.16) with (8.18) implies (12.15).

The cubic equation (12.8) in Θ is a rewriting of (11.30). In particular, we have

$$1 + \kappa_0\alpha + \nu(\omega) \;=\; \frac{\Xi(\omega)}{\langle |m| f\rangle} \;=\; 1 + \frac{\langle |m|(b-f)\rangle}{\langle |m| f\rangle} + \frac{1}{\langle |m| f\rangle}\frac{\kappa(u(\omega),\omega)}{\omega} \tag{12.17}$$

where $\kappa(u,d)$ is from (11.30). We set the ω-independent term $\kappa_0\alpha$ equal to the second term on the right hand side of (12.17). We set $\nu(\omega)$ equal to the last term in (12.17). Clearly, $|\kappa_0| \lesssim 1$ because $b = f + \mathcal{O}_{\mathscr{B}}(\alpha)$ and $|m|, f \sim 1$. The bound (11.31) and the Hölder continuity of Θ yield

$$\left|\frac{\kappa(u(\omega),\omega)}{\omega}\right| \;\lesssim\; \frac{|\Theta(\omega)|^4 + |\omega||\Theta(\omega)| + |\omega|^2}{|\omega|} \;\lesssim\; |\Theta(\omega)| + |\omega| \;\lesssim\; |\omega|^{1/3}\,.$$

This proves (12.13b). The expansions (12.14) follow by expressing $m(\tau_0+\omega)$ in terms of $\Theta(\omega;\tau_0)$ and $r(\omega;\tau_0)$, and approximating the latter with (11.28). □

The following ratio,

$$\Pi(\tau) \;:=\; \frac{|\sigma(\tau)|}{\langle v(\tau)\rangle^2}\,, \tag{12.18}$$

will play a key role in the classification of the points in $\mathbb{D}_\varepsilon$ when $\varepsilon > 0$ is small. The next result shows that if Π is sufficiently large then v grows at least like a square root in the direction $\operatorname{sign}\sigma$.

Lemma 12.2 (Monotonicity). *There exist $\varepsilon_*, \Pi_* \sim 1$ such that*

$$\operatorname{sign}\sigma(\tau)\,|\partial_\tau v(\tau)| \;\gtrsim\; \frac{\mathbb{1}\{\Pi(\tau) \geq \Pi_*\}}{|\sigma(\tau)|\langle v(\tau)\rangle + \langle v(\tau)\rangle^2}\,, \qquad \tau \in \mathbb{D}_{\varepsilon_*}\,. \tag{12.19}$$

PROOF. By Lemma 12.1 both $\Theta(\omega;\tau)$ and $r(\omega;\tau)$ are differentiable functions in ω, and thus,

$$\partial_\tau m(\tau) \;=\; |m(\tau)|\, b(\tau)\, \partial_\omega \Theta(0;\tau) \;+\; |m(\tau)|\, \partial_\omega r(0;\tau)\,. \tag{12.20}$$

Let us drop the fixed argument τ to simplify notations. Taking imaginary parts of (12.20) yields:

$$\partial_\tau v \;=\; \operatorname{Im} \partial_\tau m \;=\; |m| \operatorname{Im}\big[b\, \partial_\omega \Theta(0)\big] \;+\; |m| \operatorname{Im} \partial_\omega r(0)\,. \tag{12.21}$$

By dividing (12.12b) by ω, and using (12.12a), we see that

$$\Big|\frac{r_x(\omega)}{\omega}\Big| \;\lesssim\; 1 + \Big|\frac{\Theta(\omega)^2}{\omega}\Big| \;\lesssim\; 1 + \frac{|\omega|}{\alpha^4}\,, \qquad \forall x \in \mathfrak{X}\,.$$

Letting $\omega \to 0$, and recalling $r(0) = 0$, we see that the last term in (12.21) is uniformly bounded,

$$\|\operatorname{Im} \partial_\omega r(0)\|_{\mathscr{B}} \;\lesssim\; 1\,. \tag{12.22}$$

We will now show that $\operatorname{Im}[b\, \partial_\omega \Theta]$ dominates the second term in (12.20) provided α is sufficiently small and $|\sigma|/\alpha^2 \sim \Pi$ is sufficiently large. To this end we first rewrite the cubic (12.8),

$$\bigg(1 + \frac{\mu_2}{\mu_1}\Theta(\omega) + \frac{\mu_3}{\mu_1}\Theta(\omega)^2\bigg)\, \frac{\Theta(\omega)}{\omega} \;=\; -\frac{\Xi(\omega)}{\mu_1}\,. \tag{12.23}$$

From the definition (12.9c) we obtain

$$|\mu_1| \;\sim\; \alpha\, \big|\sigma + \mathcal{O}(\alpha^2)\big| + \alpha^2 \big|\psi - \sigma^2 + \mathcal{O}(\alpha)\big|\,,$$

by distinguishing the cases $2\sigma^2 \le \psi$ and $2\sigma^2 > \psi$, and using (12.11). Applying (12.13b) to estimate $\Xi(\omega)$ we see that the right hand side of (12.23) satisfies:

$$\frac{\Xi(\omega)}{\mu_1} \;=\; \frac{\langle f|m|\rangle}{2}\, \frac{1 + \mathcal{O}(\alpha + |\omega|^{1/3})}{\mathrm{i}\alpha\sigma - \alpha^2(\psi - \sigma^2) + \mathcal{O}(\alpha^3)}\,. \tag{12.24}$$

From (12.12a) we see that $\Theta(\omega) \to 0$ as $\omega \to 0$. Hence taking the limit $\omega \to 0$ in (12.23) and recalling $|\mu_2|, |\mu_3| \lesssim 1$, yields

$$\partial_\omega \Theta(0) \;=\; \frac{\mathrm{d}\Theta}{\mathrm{d}\omega}\bigg|_{\omega=0} \;=\; \frac{\langle f|m|\rangle}{2}\, \frac{\alpha^2(\psi - \sigma^2) + \mathrm{i}\,\alpha\,\sigma + \mathcal{O}(\alpha^3 + |\sigma|\alpha^2)}{\alpha^2|\sigma + \mathcal{O}(\alpha^2)|^2 + \alpha^4|\psi - \sigma^2 + \mathcal{O}(\alpha)|^2}\,. \tag{12.25}$$

Using $b = f + \mathcal{O}_{\mathscr{B}}(\alpha)$ and $\langle f|m|\rangle \sim 1$, we conclude from (12.25) that

$$(\operatorname{sign} \sigma) \operatorname{Im}\big[b\, \partial_\omega \Theta(0)\big] \;\sim\; \frac{|\sigma| \,+ \mathcal{O}_{\mathscr{B}}(\alpha^2 + |\sigma|\alpha)}{|\sigma + \mathcal{O}(\alpha^2)|^2 + \alpha^2|\psi - \sigma^2 + \mathcal{O}(\alpha)|^2}\, \frac{1}{\alpha}\,, \tag{12.26}$$

By definitions $|\sigma|/\alpha^2 \sim \Pi \ge \Pi_*$. Hence, if $\Pi_* \sim 1$ is sufficiently large, then the factor multiplying $1/\alpha$ on the right hand side of (12.26) scales like $\min\big\{|\sigma|^{-1}, \alpha^{-2}|\sigma|\big\}$. Here we used again (12.11). Using (12.21), (12.22), and $\alpha \sim \langle v\rangle$, from (12.26) we obtain

$$(\operatorname{sign} \sigma)\, \partial_\tau v \;\gtrsim\; \min\bigg\{\frac{1}{|\sigma|}, \frac{|\sigma|}{\langle v\rangle^2}\bigg\} \frac{1}{\langle v\rangle} \;+\; \mathcal{O}_{\mathscr{B}}(1)\,.$$

By taking $\Pi_* \sim 1$ sufficiently large and $\varepsilon_* \sim 1$ sufficiently small the term $\mathcal{O}_{\mathscr{B}}(1)$ can be ignored and (12.19) follows. $\square$

12.1 Expansion around non-zero minima of generating density

Lemma 12.2 shows that if $\tau_0 \in \mathbb{D}_{\varepsilon_*}$ is a non-zero minimum of $\tau \mapsto \langle v(\tau)\rangle$, i.e., $\langle v(\tau_0)\rangle > 0$, then $\partial_\tau \langle v(\tau_0)\rangle = 0$, and hence $\Pi(\tau_0) < \Pi_*$. Now we show that any point τ_0 satisfying $\Pi(\tau_0) < \Pi_*$ is an approximate minimum of $\langle v\rangle$, and its shape is described by the universal shape function $\Psi_{\min} : [0,\infty) \to [0,\infty)$ introduced in Definition 6.3.

Proposition 12.3 (Non-zero local minimum). *If $\tau_0 \in \mathbb{D}_\varepsilon$ satisfies*

$$\Pi(\tau_0) \,\le\, \Pi_* \tag{12.27}$$

where $\Pi_* \sim 1$ is from Lemma 12.2 (in particular if τ_0 is a non-zero local minimum of $\langle v\rangle$), then

$$v_x(\tau_0 + \omega) \,-\, v_x(\tau_0) \,=\, h_x \langle v\rangle\, \Psi_{\min}\biggl(\Gamma \frac{\omega}{\langle v\rangle^3}\biggr) \,+\, \mathcal{O}\biggl(\min\biggl\{\frac{|\omega|}{\langle v\rangle}, |\omega|^{2/3}\biggr\}\biggr) \tag{12.28}$$

for some ω-independent constants $h_x = h_x(\tau_0) \sim 1$ and $\Gamma = \Gamma(\tau_0) \sim 1$. Here $\langle v\rangle = \langle v(\tau_0)\rangle$, $\sigma = \sigma(\tau_0)$, etc. are evaluated at τ_0.

Using (6.13b) we see that the first term on the right hand side of (12.28) satisfies

$$\langle v\rangle\, \Psi_{\min}\biggl(\Gamma \frac{\omega}{\langle v\rangle^3}\biggr) \,\sim\, \min\biggl\{\frac{|\omega|^2}{\langle v\rangle^5}, |\omega|^{1/3}\biggr\}, \qquad \omega \in \mathbb{R}\,. \tag{12.29}$$

Comparing this with the last term of (12.28) we see that the first term dominates the error on the right hand side of (12.28), provided $\langle v\rangle^4 \lesssim |\omega| \lesssim 1$. Applying the lemma at two distinct base points hence yields the following property of the non-zero minima.

Corollary 12.4 (Location of non-zero minima). *Suppose two points $\tau_1, \tau_2 \in \mathbb{D}_\varepsilon$ satisfy the hypotheses of Proposition 12.3. Then, either*

$$|\tau_1 - \tau_2| \gtrsim 1\,, \qquad \textit{or} \qquad |\tau_1 - \tau_2| \lesssim \min\bigl\{\langle v(\tau_1)\rangle, \langle v(\tau_2)\rangle\bigr\}^4\,. \tag{12.30}$$

Proof. Suppose the points τ_1 and τ_2 qualify as the base points for Proposition 12.3. Then the corresponding expansions (12.28) are compatible only if the base points satisfy the dichotomy (12.30). For the second bound in (12.30) we use (12.29). $\square$

We will use the standard convention on complex powers.

Definition 12.5 (Complex powers). *We define complex powers $\zeta \mapsto \zeta^\gamma$, $\gamma \in \mathbb{C}$, on $\mathbb{C}\backslash(-\infty, 0)$, by setting $\zeta^\gamma := \exp(\gamma \log \zeta)$, where $\log : \mathbb{C}\backslash(-\infty,0) \to \mathbb{C}$ is a continuous branch of the complex logarithm with $\log 1 = 0$. We denote by $\arg : \mathbb{C}\backslash\{0\} \to (-\pi, \pi)$, the corresponding angle function.*

Proof of Proposition 12.3. Without loss of generality it suffices to prove (12.28) in the case $|\omega| \le \delta$ for some sufficiently small constant $\delta \sim 1$. Indeed, when $|\omega| \gtrsim 1$ the expansion (12.28) becomes trivial since the last term is $\mathcal{O}(1)$ and therefore dominates all the other terms, including $|v_x(\tau)| \le \|m\|_{\mathbb{R}} \sim 1$. Similarly, we may restrict ourselves to the setting where the quantity

$$\chi \,:=\, \alpha + \frac{|\sigma|}{\alpha}\,, \tag{12.31}$$

satisfies $\chi \le \chi_*$, for some sufficiently small threshold $\chi_* \sim 1$. In particular, we assume that χ_* is so small that $\chi \le \chi_*$ implies $\langle v\rangle \le \varepsilon_*$.

Let us denote by $\gamma_k \in \mathbb{C}$, $k = 1, 2, 3, ...$, generic ω-independent numbers, satisfying

$$|\gamma_k| \lesssim \chi\,. \tag{12.32}$$

Since $\Pi \sim |\sigma|/\alpha^2$ and $\Pi \leq \Pi_*$ we have $|\sigma| \leq \Pi_* \chi_*^2$. From (12.11) it hence follows that $\psi \sim 1$ for sufficiently small $\chi_* \sim 1$. Thus the cubic (12.8) takes the form

$$\Theta(\omega)^3 \,+\, \mathrm{i}3\alpha(1+\gamma_2)\Theta(\omega)^2 \,-\, 2\alpha^2(1+\gamma_1)\Theta(\omega) \,+\, (1+\gamma_0+\nu(\omega))\,\frac{\langle f|m|\rangle}{\psi}\omega \;=\; 0\,. \tag{12.33}$$

Using the following *normal coordinates*,

$$\begin{aligned} \lambda &:= \Gamma\,\frac{\omega}{\alpha^3} \\ \Omega(\lambda) &:= \sqrt{3}\,\biggl[(1+\gamma_3)\,\frac{1}{\alpha}\Theta\Bigl(\frac{\alpha^3}{\Gamma}\lambda\Bigr) \,+\, \mathrm{i} \,+\, \gamma_4\biggr]\,, \end{aligned} \tag{12.34}$$

where $\Gamma := (\sqrt{27}/2)\langle|m|f\rangle/\psi \sim 1$, (12.33) reduces to

$$\Omega(\lambda)^3 + 3\Omega(\lambda) \,+\, 2\Lambda(\lambda) \;=\; 0\,. \tag{12.35}$$

Here the constant term $\Lambda : \mathbb{R} \to \mathbb{C}$ is given by

$$\begin{aligned} \Lambda(\lambda) &:= (1+\gamma_5+\mu(\lambda))\,\lambda \,+\, \gamma_6 \\ \mu(\lambda) &:= \nu\Bigl(\frac{\alpha^3}{\Gamma}\lambda\Bigr)\,. \end{aligned} \tag{12.36}$$

The following lemma presents Cardano's solution for the reduced cubic (12.35) in a form that is convenient for our analysis. We omit the proof of this well know result.

LEMMA 12.6 (Roots of reduced cubic with positive linear coefficient). *The following holds*

$$\Omega^3 \,+\, 3\Omega \,+\, 2\zeta \;=\; (\Omega-\widehat{\Omega}_+(\zeta))(\Omega-\widehat{\Omega}_0(\zeta))(\Omega-\widehat{\Omega}_-(\zeta))\,, \qquad \forall \zeta \in \mathbb{C}\,, \tag{12.37}$$

where the three root functions $\widehat{\Omega}_a : \mathbb{C} \to \mathbb{C}$, $a = 0, \pm$, are given by

$$\begin{aligned} \widehat{\Omega}_0 &:= -2\Phi_{\mathrm{odd}} \\ \widehat{\Omega}_\pm &:= \Phi_{\mathrm{odd}} \,\pm\, \mathrm{i}\sqrt{3}\,\Phi_{\mathrm{even}}\,, \end{aligned} \tag{12.38a}$$

with Φ_{even} and Φ_{odd} denoting the even and odd parts of the function $\Phi : \mathbb{C} \to \mathbb{C}$,

$$\Phi(\zeta) \,:=\, \bigl(\sqrt{1+\zeta^2} \,+\, \zeta\bigr)^{1/3}\,, \tag{12.38b}$$

respectively. The roots (12.38) *are analytic and distinct on the set,*

$$\widehat{\mathbb{C}} \,:=\, \mathbb{C}\backslash\{\,\mathrm{i}\xi : \xi \in \mathbb{R}, |\xi| > 1\}\,. \tag{12.39}$$

Indeed, if $\widehat{\Omega}_a(\zeta) = \widehat{\Omega}_b(\zeta)$, for $a \neq b$, then $\zeta = \pm\mathrm{i}$.

Since $\Omega(\lambda)$, defined in (12.34), solves the cubic (12.35), there exists $A : \mathbb{R} \to \{0, \pm\}$, such that

$$\Omega(\lambda) = \widehat{\Omega}_{A(\lambda)}(\Lambda(\lambda))\,, \qquad \lambda \in \mathbb{R}\,. \tag{12.40}$$

In the normal coordinates the restriction $|\omega| \le \delta$ becomes $|\lambda| \le \lambda_*$, where

$$|\lambda| \le \lambda_* := \Gamma \frac{\delta}{\alpha^3}. \tag{12.41}$$

Nevertheless, for sufficiently small $\delta \sim 1$ the function Λ in (12.36) is a small perturbation of the identity function. Indeed, from (12.36) and the bound (12.13b) on ν, we get

$$\begin{aligned} |\mu(\lambda)| &\lesssim \Big|\Theta\Big(\frac{\alpha^3}{\Gamma}\lambda\Big)\Big| + \alpha^3|\lambda| \\ &\lesssim \alpha|\lambda|^{1/3} \lesssim \delta^{1/3}, \qquad \text{when} \quad |\lambda| \le \lambda_* \,. \end{aligned} \tag{12.42}$$

Hence, if the thresholds $\delta, \chi_* \lesssim 1$ are sufficiently small, then

$$\Lambda(\lambda) \in \mathbb{G}\,, \quad \text{and} \quad |\Lambda(\lambda)| \sim |\lambda|\,, \qquad |\lambda| \le \lambda_*\,, \tag{12.43}$$

where

$$\mathbb{G} := \Big\{\zeta \in \mathbb{C} : \operatorname{dist}\big(\zeta, \mathrm{i}(-\infty,-1) \cup \mathrm{i}(+1,+\infty)\big) \ge 1/2\Big\}. \tag{12.44}$$

By Lemma 12.6 the root functions have uniformly bounded derivatives on this subset of $\widehat{\mathbb{C}}$.

The following lemma which is proven in Appendix B.5 provides a useful bound to replace $\Lambda(\lambda)$ by λ in (12.40).

LEMMA 12.7 (Stability of roots). *There exist positive constants c_1, C_1 such that if $\zeta \in \mathbb{G}$ and $\xi \in \mathbb{C}$ satisfy*

$$|\xi| \le c_1(1+|\zeta|)\,, \tag{12.45}$$

then the roots (12.38) *are stable in the sense that*

$$\big|\widehat{\Omega}_a(\zeta+\xi) - \widehat{\Omega}_a(\zeta)\big| \le \frac{C_1|\xi|}{1+|\zeta|^{2/3}}\,, \qquad a = 0, \pm\,. \tag{12.46}$$

From (12.43) we see that $\Lambda(\lambda) \neq \pm\mathrm{i}$ and hence the roots do not coincide. Moreover, we know from Lemma 12.1 and (12.34):

SP-1. $\lambda \mapsto \Omega(\lambda)$ is continuous.

This simple fact will be the first of the four selection principles (**SP**) used for determining the correct roots of the cubic (12.8) in the following (cf. Lemma 12.9). Since the roots $\widehat{\Omega}_a|_{\mathbb{G}}$ are also continuous by Lemma 12.6, we conclude that the labelling function A in (12.40) stays constant on the interval $[-\lambda_*, \lambda_*]$. In order to determine this constant, $a := A(\lambda)$, we use the second selection principle:

SP-2. $\Omega(0)$ must match the initial condition $\Theta(0) = 0$.

Plugging $\Theta(0) = 0$ into (12.34) yields

$$\Omega(0) = \mathrm{i}\sqrt{3}\,(1+\gamma_4) = \mathrm{i}\sqrt{3} + \mathcal{O}\Big(\alpha + \frac{|\sigma|}{\alpha}\Big). \tag{12.47}$$

On the other hand, using Lemma 12.7 and (12.36) we get

$$\widehat{\Omega}_a(\Lambda(0)) = \widehat{\Omega}_a(\gamma_6) = \widehat{\Omega}_a(0) + \mathcal{O}\Big(\alpha + \frac{|\sigma|}{\alpha}\Big), \tag{12.48}$$

where

$$\widehat{\Omega}_0(0) \;=\; 0 \qquad \text{and} \qquad \widehat{\Omega}_\pm(0) \;=\; \pm\mathrm{i}\sqrt{3}\,.$$

Comparing this with (12.47) and (12.48), we see that for sufficiently small $\alpha+|\sigma|/\alpha \lesssim \chi_*$, only the the choice $A(0) = +$ satisfies **SP-2.**

As the last step we derive the expansion (12.28) using the formula

$$v_x(\tau_0+\omega) - v_x(\tau_0) \;=\; |m_x|\, f_x \operatorname{Im}\Theta(\omega) \;+\; \mathcal{O}\Big(\,\alpha\,|\Theta(\omega)| + |\Theta(\omega)|^2 + |\omega|\,\Big)\,, \tag{12.49}$$

which follows by taking the imaginary part of (12.14a). We also used $b_x = (1+\mathcal{O}(\alpha))f_x$ and $f_x \sim 1$ here. Let us express Θ in terms of the normal coordinates using (12.34)

$$\Theta(\omega) \;=\; \frac{\alpha}{1+\gamma_3}\left[\frac{\widehat{\Omega}_+(\Lambda(\lambda))}{\sqrt{3}} \;-\; \mathrm{i} \;-\; \gamma_4\right]. \tag{12.50}$$

Here, ω and λ are related by (12.34). Since $\Theta(0) = 0$, and $\Lambda(0) = \gamma_6$ (cf. (12.36)), we get

$$\mathrm{i}+\gamma_4 = \frac{\widehat{\Omega}_+(\gamma_6)}{\sqrt{3}}\,.$$

Using this identity and

$$\Lambda(\lambda) = \gamma_6 + \Lambda_0(\lambda) \qquad \text{with} \qquad \Lambda_0(\lambda) := (1+\gamma_5+\mu(\lambda))\,\lambda\,,$$

we rewrite the formula (12.50) as

$$\Theta(\omega) \;=\; (1+\mathcal{O}(\chi))\,\frac{\alpha}{\sqrt{3}}\Big[\widehat{\Omega}_+(\gamma_6+\Lambda_0(\lambda)) - \widehat{\Omega}_+(\gamma_6)\Big]. \tag{12.51}$$

From (12.43) we know that the arguments of $\widehat{\Omega}_+$ in (12.51) are in $\mathbb{G}$. Using the uniform boundedness of the derivatives of $\Omega|_{\mathbb{G}}$, and the bound $|\Phi(\zeta)| \lesssim 1+|\zeta|^{1/3}$, we get

$$\big|\widehat{\Omega}_+(\gamma_6+\Lambda_0(\lambda)) - \widehat{\Omega}_+(\gamma_6)\big| \;\lesssim\; \min\big\{|\lambda|,|\lambda|^{1/3}\big\}\,, \qquad |\lambda| \le \lambda_*\,. \tag{12.52}$$

By using (12.52) in (12.42) and (12.51) we estimate the sizes of both $\mu(\lambda)$ and $\Theta(\omega)$,

$$|\mu(\lambda)| \;+\; \left|\Theta\Big(\frac{\alpha^3}{\Gamma}\lambda\Big)\right| \;\lesssim\; \alpha\min\big\{|\lambda|,|\lambda|^{1/3}\big\}\,, \qquad |\lambda| \le \lambda_*\,. \tag{12.53}$$

In order to extract the exact leading order terms, we express the difference on the right hand side of (12.51) using the mean value theorem

$$\begin{aligned}\widehat{\Omega}_+(\gamma_6+\Lambda_0(\lambda)) - \widehat{\Omega}_+(\gamma_6) \;&=\; \widehat{\Omega}_+(\Lambda_0(\lambda)) - \widehat{\Omega}_+(0)\\ &\quad+\;\gamma_6\,\frac{\partial}{\partial\zeta}\Big[\widehat{\Omega}_+(\zeta+\Lambda_0(\lambda)) - \widehat{\Omega}_+(\zeta)\Big]_{\zeta=\gamma}\,,\end{aligned} \tag{12.54}$$

where $\gamma \in \mathbb{G}$ is some point on the line segment connecting 0 and γ_6. Using (12.53) and Lemma 12.7 on the first term on the right hand side of (12.54) shows

$$\widehat{\Omega}_+(\Lambda_0(\lambda)) - \widehat{\Omega}_+(0) \;=\; \widehat{\Omega}_+(\lambda) - \widehat{\Omega}_+(0) \;+\; \mathcal{O}\Big(\chi\,\min\big\{|\lambda|,|\lambda|^{2/3}\big\}\Big)\,. \tag{12.55}$$

From an explicit calculation we get $|\partial_\zeta\widehat{\Omega}_+(\zeta)| \lesssim 1$, for $\zeta \in \mathbb{G}$. Thus

$$\left|\frac{\partial}{\partial\zeta}\Big[\widehat{\Omega}_+(\zeta+\Lambda_0(\lambda)) - \widehat{\Omega}_+(\zeta)\Big]_{\zeta=\gamma}\right| \;\lesssim\; \min\big\{\,|\lambda|,1\big\}\,.$$

Plugging this and (12.55) into (12.54) yields

$$\widehat{\Omega}_+(\gamma_6 + \Lambda_0(\lambda)) - \widehat{\Omega}_+(\gamma_6) \,=\, \widehat{\Omega}_+(\lambda) - \widehat{\Omega}_+(0) \,+\, \mathcal{O}\Big(\chi \min\big\{|\lambda|, |\lambda|^{2/3}\big\}\Big)\,. \tag{12.56}$$

Via (12.51) we use this to represent the leading order term in (12.49). By approximating all the other terms in (12.49) with (12.53) we obtain

$$v_x(\tau_0+\omega) - v_x(\tau_0) \,=\, |m|_x\, f_x\, \alpha\, \frac{\mathrm{Im}\big[\widehat{\Omega}_+(\lambda) - \widehat{\Omega}_+(0)\big]}{\sqrt{3}} \,+\, \mathcal{O}\Big(\big(\alpha^2 + |\sigma|\big) \min\big\{|\lambda|, |\lambda|^{2/3}\big\}\Big)\,. \tag{12.57}$$

Using the formulas (12.38) and (12.38b), we identify the universal shape function from (6.13b),

$$\Psi_{\min}(\lambda) \,=\, \frac{\mathrm{Im}\big[\widehat{\Omega}_+(\lambda) - \widehat{\Omega}_+(0)\big]}{\sqrt{3}}\,.$$

Denoting $h_x := (\alpha/\langle v\rangle)\, f_x$ and writing λ in terms of ω in (12.57), the expansion (12.28) follows. □

12.2 Expansions around minima where generating density vanishes

Together with Proposition 12.3 the next result covers the behaviour of $v|_{\mathbb{D}_\varepsilon}$ around its minima for sufficiently small $\varepsilon \sim 1$. For each $\tau_0 \in \partial\operatorname{supp} v$, satisfying $\sigma(\tau_0) \neq 0$, we associate the **gap length**,

$$\Delta(\tau_0) \,:=\, \inf\big\{\xi \in (0,2] : \big\langle v(\tau_0 - \operatorname{sign}\sigma(\tau_0)\xi)\big\rangle > 0\,\big\}\,, \tag{12.58}$$

with the convention $\Delta(\tau_0) := 1$ in case the infimum is infinite. We will see below that if $\tau_0 \in \partial\operatorname{supp} v$, then $\sigma(\tau_0) \neq 0$ and $\operatorname{sign}\sigma(\tau_0)$ is indeed the direction in which the set $\operatorname{supp} v$ continues from τ_0. Because $\operatorname{supp} v \subset [-2,2]$ and $0 \in \operatorname{supp} v$, the number $\Delta(\tau_0)$ thus defines the length of the actual gap in $\operatorname{supp} v$ starting at τ_0, with the convention that the gap length is 1 for the extreme edges.

Recall the definition (6.13a) of the universal edge shape function $\Psi_{\mathrm{edge}} : [0,\infty) \to [0,\infty)$.

PROPOSITION 12.8 (Vanishing local minimum). *Suppose $\tau_0 \in \operatorname{supp} v$ with $v(\tau_0) = 0$. Depending on the value of $\sigma = \sigma(\tau_0)$ either of the following holds:*

(i) *If $\sigma(\tau_0) \neq 0$, then $\tau_0 \in \partial\operatorname{supp} v$ and $\operatorname{supp} v$ continues in the direction $\operatorname{sign}\sigma$, such that*

$$v_x(\tau_0+\omega) \,=\, h_x\, \Delta^{1/3}\, \Psi_{\mathrm{edge}}\Big(\frac{|\omega|}{\Delta}\Big) \,+\, \mathcal{O}\bigg(\min\Big\{\frac{|\omega|}{\Delta^{1/3}}, |\omega|^{2/3}\Big\}\bigg)\,, \qquad (\operatorname{sign}\sigma)\,\omega \geq 0\,, \tag{12.59}$$

where $h_x = h_x(\tau_0) \sim 1$, and $\Delta = \Delta(\tau_0)$ is the length of the gap in $\operatorname{supp} v$ in the direction $-\operatorname{sign}\sigma$ from τ_0 (cf. (12.58)). Furthermore, the gap length satisfies

$$\Delta(\tau_0) \,\sim\, |\sigma(\tau_0)|^3\,, \tag{12.60}$$

while the shapes in the x-direction match at the opposite edges of the gap in the sense that $h(\tau_1) = h(\tau_0) + \mathcal{O}_{\mathscr{B}}(\Delta^{1/3})$, for $\tau_1 = \tau_0 - \operatorname{sign}\sigma(\tau_0)\,\Delta$.

(ii) *If $\sigma(\tau_0) = 0$ then $\operatorname{dist}(\tau_0, \partial\operatorname{supp} v) \sim 1$, and for some $h_x = h_x(\tau_0) \sim 1$:*

$$v_x(\tau_0+\omega) \,=\, h_x\, |\omega|^{1/3} + \mathcal{O}\big(|\omega|^{2/3}\big)\,. \tag{12.61}$$

From the explicit formula (6.13a) one sees that the leading order term in (12.59) satisfies

$$\Delta^{1/3}\,\Psi_{\mathrm{edge}}\biggl(\frac{\omega}{\Delta}\biggr) \;\sim\; \begin{cases} \dfrac{\omega^{1/2}}{\Delta^{1/6}} & \text{when } 0 \le \omega \lesssim \Delta\,; \\ \omega^{1/3} & \text{when } \omega \gtrsim \Delta\,. \end{cases} \tag{12.62}$$

In particular, when an edge τ_0 is separated by a gap of length $\Delta(\tau_0) \sim 1$ from the opposite edge of the gap then v grows like a square root.

Proposition 12.8 is proven at the end of this section by combining various auxiliary results which we prove in the following two subsections. What is common with these intermediate results is that the underlying cubic (12.8) is always of the form

$$\psi\,\Theta(\omega)^3 + \sigma\,\Theta(\omega)^2 + (1+\nu(\omega))\langle |m|\,f\rangle\,\omega \;=\; 0\,, \qquad \psi + |\sigma|^2 \sim 1\,, \tag{12.63}$$

since $\alpha(\tau_0) = v(\tau_0) = 0$ at the base point τ_0. In order to analyse (12.63) we bring it to a **normal form** by an affine transformation. This corresponds to expressing the variables ω and Θ in terms of **normal variables** Ω and λ, such that

$$\begin{aligned} \Omega(\lambda) \;&=\; \kappa\,\Theta(\Gamma\lambda) + \Omega_0 \\ &=\; \bigl\langle h, m(\tau_0 + \Gamma\lambda) - m(\tau_0)\bigr\rangle \,+\, \Omega_0\,, \end{aligned} \tag{12.64}$$

with some λ-independent parameters $\kappa = \kappa(\tau_0), \Gamma = \Gamma(\tau_0) > 0$, $h_x = h(\tau_0) \sim 1$, and $\Omega_0 = \Omega(0) \in \mathbb{C}$. These parameters will be determined on a case by case basis. We remark, that in the proof of Proposition 12.3 the coordinate transformations (12.34) were of the form (12.64).

The variable $\Omega(\lambda)$ satisfies an equation which has typically multiple solutions since a generic cubic has three distinct roots. In order to choose the correct solution Ω we use the following **selection principles**.

LEMMA 12.9 (Selection principles). *If $v(\tau_0) = 0$ at the base point $\tau_0 \in \operatorname{supp} v$ of the expansion* (12.64)*, then $\Omega(\lambda) = \Omega(\lambda;\tau_0)$ defined in* (12.64) *has the following four properties:*

SP-1. $\lambda \mapsto \Omega(\lambda)$ *is continuous;*

SP-2. $\Omega(0) = \Omega_0$*;*

SP-3. $\operatorname{Im}\bigl[\Omega(\lambda) - \Omega(0)\bigr] \ge 0$, $\forall \lambda \in \mathbb{R}$*;*

SP-4. *If the imaginary part of Ω grows slower than a square root in a direction $\theta \in \{\pm 1\}$,*

$$\lim_{\xi \to 0_+} \xi^{-1/2}\,\operatorname{Im}\Omega(\theta\xi) = 0\,,$$

then $\Omega|_I$ is real and non-decreasing on an interval $I := \{\theta\xi : 0 < \xi < \Delta\}$, with some $\Delta > 0$.

The first three selection principles follow trivially from the corresponding properties of m and Θ. The property **SP-4.** is just the following lemma stated in the normal variables (12.64).

LEMMA 12.10 (Growth condition). *Suppose $v(\tau_0) = 0$ and that $\langle v\rangle$ grows slower than any square-root in a direction $\theta \in \{\pm\}$, i.e.,*

$$\liminf_{\xi \to 0_+} \frac{\langle v(\tau_0 + \theta\xi)\rangle}{\xi^{1/2}} \;=\; 0\,. \tag{12.65}$$

Then $\langle v\rangle$ actually vanishes, $\operatorname{Im}\langle m\rangle|_I = 0$, while $\operatorname{Re}\langle m\rangle$ is non-decreasing on some interval $I = \{\theta\xi : 0 \le \xi \le \Delta\}$, for some $\Delta > 0$.

If the $\liminf$ *in* (12.65) *is non-zero, then either $\theta = \operatorname{sign}\sigma(\tau_0)$ or $\sigma(\tau_0) = 0$.*

PROOF. We will prove below that if $v(\tau_0) = 0$, and

$$\inf\bigl\{\, \xi > 0 : \langle v(\tau_0 + \theta\xi)\rangle > 0 \bigr\} \,=\, 0\,, \tag{12.66}$$

for some direction $\theta \in \{\pm 1\}$, then

$$\liminf_{\xi \to 0_+} \frac{\langle v(\tau_0 + \theta\xi)\rangle}{\xi^{1/2}} \,>\, 0\,. \tag{12.67}$$

Assuming this implication, the lemma follows easily: If (12.65) holds, then (12.66) is not true, i.e., there is a non-trivial interval $I = \{\theta\xi : 0 \le \xi \le \Delta\}$, $\Delta > 0$, such that $v|_I = 0$. As the negative of a Hilbert-transform of v_x (cf. (6.8)), the function $\tau \mapsto \operatorname{Re} m_x(\tau)$, is non-decreasing on I. This proves the first part of the lemma.

We will now prove that (12.66) implies (12.67). The key idea is to use Lemma 12.2 to prove that $\langle v\rangle$ grows at least like a square root. However, first we use Proposition 12.3 to argue that the indicator function on the right hand side of (12.19) is non-zero in a non-trivial neighbourhood of τ_0. To this end, assume $0 < \langle v(\tau)\rangle \le \varepsilon$ and $\Pi(\tau) < \Pi_*$. If $\varepsilon, \delta > 0$ are sufficiently small, then Proposition 12.3 can be applied with τ as the base point. In particular, (12.28) and (12.29) imply

$$\langle v(\tau + \omega)\rangle \,\sim\, \langle v(\tau)\rangle + |\omega|^{1/3} \,>\, 0\,, \qquad |\omega| \le \delta\,. \tag{12.68}$$

Suppose τ_0 satisfies (12.66). Since $v(\tau_0) = 0$ the lower bound in (12.68), applied to $\omega = \tau_0 - \tau$, implies $|\tau - \tau_0| > \delta$. As τ was arbitrary we conclude $\Pi(\tau) \ge \Pi_*$ for every τ in the set

$$I := \bigl\{\tau \in \mathbb{R} : |\tau - \tau_0| \le \delta\,,\; 0 < \langle v(\tau)\rangle \le \varepsilon\bigr\}\,.$$

Applying Lemma 12.2 on I, recalling the upper bound on $|\partial_z m|$ from Corollary 10.2, yields

$$\langle v\rangle^{-1} \lesssim\, (\operatorname{sign}\sigma)\,\partial_\tau\langle v\rangle \,\lesssim\, \langle v\rangle^{-2}\,, \quad \text{on} \quad I\,. \tag{12.69}$$

Since v is analytic when non-zero, and $\operatorname{dist}(\tau_0, I) = 0$ by (12.66), we conclude that I equals the interval with end points τ_0 and $\tau_1 := \tau_0 + \theta\delta$. Here we set $\delta \lesssim \varepsilon^3$ so small that the 1/3-Hölder continuity of m guarantees $\langle v\rangle \le \varepsilon$ on I. Moreover, $\operatorname{sign}\sigma(\tau)$ must be equal to the constant θ for every $\tau \in I$: If σ changed its sign at some point $\tau_* \in I$ this would violate $\Pi(\tau_*) \ge \Pi_*$ as $\langle v\rangle$ is a continuous function.

Integrating (12.69) from τ_0 to τ_1 we see that $\langle v(\tau_0 + \theta\xi)\rangle^2 \gtrsim \xi$ for any $\xi \le |\tau_1 - \tau_0|$. This proves the limit (12.67), and hence the first part of the lemma. The second part of the lemma follows from (12.69). □

12.3 Simple edge and sharp cusp

When $|\sigma| > 0$ and $|\omega|$ is sufficiently small compared to $|\sigma|$ the cubic term $\psi\,\Theta(\omega)^3$ in (12.63) can be ignored. In this regime the following simple expansion holds showing the square root behaviour of v near an edge of its support.

LEMMA 12.11 (Simple edge). *If $\tau_0 \in \operatorname{supp} v$ satisfies $v(\tau_0) = 0$ and $\sigma = \sigma(\tau_0) \neq 0$, then*

$$v_x(\tau_0 + \omega) \;=\; \begin{cases} h'_x \left|\dfrac{\omega}{\sigma}\right|^{1/2} + \mathcal{O}\Bigl(\dfrac{\omega}{\sigma^2}\Bigr) & \text{if} \quad 0 \le (\operatorname{sign}\sigma)\,\omega \le c_*|\sigma|^3\,; \\ 0 & \text{if} \;\; -\, c_*|\sigma|^3 \le (\operatorname{sign}\sigma)\,\omega \le 0\,; \end{cases} \tag{12.70}$$

for some sufficiently small $c_* \sim 1$. Here $h' = h'(\tau_0) \in \mathscr{B}$ satisfies $h'_x \sim 1$.

This result already shows that $\operatorname{supp} v$ continues in the direction $\operatorname{sign}\sigma(\tau_0)$ and in the opposite direction there is a gap of length $\Delta(\tau_0) \gtrsim |\sigma(\tau_0)|^3$ in the set $\operatorname{supp} v$. We will see later (cf. Lemma 12.17) that for small $|\sigma(\tau_0)|$ there is an asymptotically sharp correspondence between $\Delta(\tau_0)$ and $|\sigma(\tau_0)|^3$, as $\Delta(\tau_0)$ becomes very small.

Proof. Treating the cubic term $\psi\Theta^3$ in (12.63) as a perturbation, (12.63) takes the form

$$\Omega(\lambda)^2 + \Lambda(\lambda) \,=\, 0\,, \tag{12.71}$$

in the normal coordinates,

$$\begin{aligned} \lambda \,&:=\, \frac{\omega}{\sigma} \\ \Omega(\lambda) \,&:=\, \frac{\Theta(\sigma\lambda)}{\sqrt{\langle |m| f\rangle}} \end{aligned} \tag{12.72}$$

where $\Lambda : \mathbb{R} \to \mathbb{C}$ is a multiplicative perturbation of λ:

$$\begin{aligned} \Lambda(\lambda) \,&:=\, (1+\mu(\lambda))\,\lambda \\ 1 \,+\, \mu(\lambda) \,&:=\, \frac{1 \,+\, \nu(\sigma\lambda)}{1+(\psi/\sigma)\,\Theta(\sigma\lambda)}\,. \end{aligned} \tag{12.73}$$

Let $\lambda_* = c_*|\sigma|^2$, with some $c_* \sim 1$, so that the constraint $|\omega| \le c_*|\sigma|^3$ in (12.70) translates into $|\lambda| \le \lambda_*$.

Using the a priori bounds (12.12a) and (12.13b) for Θ and ν yields

$$|\mu(\lambda)| \,\lesssim\, \Big(1+\frac{\psi}{|\sigma|}\Big)\big|\Theta(\sigma\lambda)\big| + |\sigma|\,|\lambda| \,\lesssim\, c_*^{1/3}\,. \tag{12.74}$$

Hence, for sufficiently small $c_* \sim 1$ we get $|\mu(\lambda)| < 1$, provided $|\lambda| \le \lambda_*$.

Let us define two root functions $\widehat{\Omega}_a : \mathbb{C} \to \mathbb{C}$, $a = \pm$, such that

$$\widehat{\Omega}_a(\zeta)^2 + \zeta \,=\, 0\,, \tag{12.75}$$

by setting

$$\widehat{\Omega}_\pm(\zeta) \,:=\, \pm\begin{cases} \mathrm{i}\,\zeta^{1/2} & \text{if } \operatorname{Re}\zeta \ge 0\,; \\ -(-\zeta)^{1/2} & \text{if } \operatorname{Re}\zeta < 0\,. \end{cases} \tag{12.76}$$

Note that we use the same symbol $\widehat{\Omega}_a$ for the roots as in (12.38) for different functions. In each expansion $\widehat{\Omega}_a$ will denote the root function of the appropriate normal form of the cubic.

Comparing (12.71) and (12.75) we see that there exists a *labelling function* $A : \mathbb{R} \to \{\pm\}$, such that

$$\Omega(\lambda) = \widehat{\Omega}_{A(\lambda)}(\Lambda(\lambda))\,, \tag{12.77}$$

for every $\lambda \in \mathbb{R}$. The function $A|_{[-\lambda_*,\lambda_*]}$ will now be determined using the selection principles **SP-1.** and **SP-3.**

The restrictions of the root functions onto the half spaces $\operatorname{Re}\zeta > 0$ and $\operatorname{Re}\zeta < 0$ are continuous (analytic) and distinct, i.e., $\widehat{\Omega}_+(\zeta) \neq \widehat{\Omega}_-(\zeta)$ for $\zeta \neq 0$. Since $\Omega : \mathbb{R} \to \mathbb{C}$ is also continuous by **SP-1.**, $A(\lambda)$ may change its value at some point $\lambda = \lambda_0$ only if $\Lambda(\lambda_0) = 0$. Since

$|\mu(\lambda)| < 1$ for $|\lambda| \le \lambda_*$, we conclude that $\Lambda(\lambda) = 0$ only for $\lambda = 0$. Thus, there exist two labels $a_+, a_- \in \{\pm\}$, such that

$$A(\lambda) \;=\; a_\pm \qquad \forall\, \lambda \in \pm[0, \lambda_*]\,. \tag{12.78}$$

Let us first consider the case $\lambda \ge 0$, and show that $a_+ = +$. Indeed, the choice $a_+ = -$ is ruled out, since

$$\operatorname{Im} \widehat{\Omega}_-(\Lambda(\lambda)) \;=\; \operatorname{Im}\Big[-\mathrm{i}\,(1 + \mu(\lambda))^{1/2}\,\lambda^{1/2}\Big] \;=\; -\lambda^{1/2} + \mathcal{O}\Big(\mu(\lambda)\,\lambda^{1/2}\Big)\,, \tag{12.79}$$

is negative for sufficiently small $c_* \sim 1$ in (12.74), and this violates the selection principle **SP-3.**

By definitions,

$$|\Theta(\sigma\lambda)| \;\sim\; |\widehat{\Omega}_+(\lambda)| \;\lesssim\; |\Lambda(\lambda)|^{1/2} \sim |\lambda|^{1/2}\,.$$

Using $\psi/|\sigma| \lesssim |\sigma|^{-1}$, with $|\sigma| \gtrsim 1$, we write (12.74) in the form $|\mu(\lambda)| \lesssim |\sigma|^{-1}|\lambda|^{1/2}$. Similarly, as (12.79) we obtain

$$\Omega(\lambda) \;=\; \widehat{\Omega}_+(\lambda) + \mathcal{O}\Big(\mu(\lambda)\,\lambda^{1/2}\Big) \;=\; \mathrm{i}\,\lambda^{1/2} + \mathcal{O}\Big(\frac{\lambda}{\sigma}\Big)\,, \qquad \lambda \in [0, \lambda_*]\,.$$

Inverting (12.72) we obtain

$$\operatorname{Im} \Theta(\omega) \;=\; \langle |m|\, f\rangle^{1/2}\Big|\frac{\omega}{\sigma}\Big|^{1/2} + \mathcal{O}\Big(\frac{\omega}{\sigma^2}\Big)\,, \qquad \operatorname{sign}\sigma = \operatorname{sign}\omega\,. \tag{12.80}$$

Taking the imaginary part of (12.14a) and using (12.80) yields the first line of (12.70), with $h'_x = |m_x|\, f_x/\langle |m|\, f\rangle^{1/2}$. Since $|m_x|, f_x \sim 1$, we also have $h'_x \sim 1$.

In order to prove the second line of (12.70) we show that the gap length (cf. (12.58)) satisfies

$$\Delta(\tau_0) \;\gtrsim\; |\sigma(\tau_0)|^3\,. \tag{12.81}$$

At the opposite edge of the gap $\tau_1 := \tau_0 - \operatorname{sign}\sigma(\tau_0)\,\Delta(\tau_0)$, the density $\langle v\rangle$ increases, by definition, in the opposite direction than at τ_0. By Lemma 12.10 the average generating density $\langle v\rangle$ increases at least like a square root function and either $\operatorname{sign}\sigma(\tau_1) = -\operatorname{sign}\sigma(\tau_0)$ or $\sigma(\tau_1) = 0$. Since σ is 1/3-Hölder continuous, σ can not change arbitrarily fast. Namely, we have $\Delta(\tau_0) \gtrsim |\sigma(\tau_0)|^3$, and this proves (12.81). □

Although not necessary for the proof of the present lemma, it can be shown that $a_- := \operatorname{sign}\sigma$ using the selection principle **SP-4.** The same reasoning will be used in the proofs of the next two lemmas (cf. (12.90) and discussion after that).

Next we consider the marginal case where the term $\sigma\,\Theta(\omega)^2$ is absent in the cubic (12.63). In this case $\langle v\rangle$ has a cubic root cusp shape around the base point.

Lemma 12.12 (Vanishing quadratic term). *If $\tau_0 \in \operatorname{supp} v$ is such that $v(\tau_0) = \sigma(\tau_0) = 0$, then*

$$v_x(\tau_0 + \omega) \;=\; h_x\, |\omega|^{1/3} + \mathcal{O}\Big(|\omega|^{2/3}\Big)\,, \tag{12.82}$$

where $h = h(\tau_0) \in \mathscr{B}$ satisfies $h_x \sim 1$.

Contrasting this with Lemma 12.11 shows that $\sigma(\tau_0) \ne 0$ for $\tau_0 \in \partial \operatorname{supp} v$. In particular, the gap length $\Delta(\tau_0)$ is always well defined for $\tau_0 \in \partial \operatorname{supp} v$ (cf. (12.58)).

Proof. First we note that it suffices to prove (12.82) only for $|\omega| \le \delta$, where $\delta \sim 1$ can be chosen to be sufficiently small. When $|\omega| > \delta$ the last term may dominate the first term on the right hand side of (12.82), and thus we have nothing prove. Since $\sigma = 0$, the quadratic term is missing in (12.63), and thus the cubic reduces to

$$\Omega(\omega)^3 + \Lambda(\omega) = 0\,, \tag{12.83}$$

using the normal coordinates

$$\begin{aligned} \lambda &:= \omega \\ \Omega(\lambda) &:= \Big(\frac{\psi}{\langle |m| f\rangle}\Big)^{1/3}\Theta(\lambda)\,. \end{aligned} \tag{12.84}$$

Here $\Lambda : \mathbb{R} \to \mathbb{C}$ is a perturbation of the identity function,

$$\Lambda(\lambda) := (1 + \nu(\lambda))\,\lambda\,. \tag{12.85}$$

Note that $\psi \sim 1$ because of (12.11).

Let us define three root functions $\widehat{\Omega}_a : \mathbb{C} \to \mathbb{C}$, $a = 0, \pm$, satisfying

$$\widehat{\Omega}_a(\zeta)^3 + \zeta = 0\,,$$

by the explicit formulas

$$\begin{aligned} \widehat{\Omega}_0(\zeta) &:= -p_3(\zeta) \\ \widehat{\Omega}_\pm(\zeta) &:= \frac{-1 \pm \mathrm{i}\sqrt{3}}{2}\,p_3(\zeta)\,, \end{aligned} \tag{12.86}$$

where $p_3 : \mathbb{C} \to \mathbb{C}$ is a (non-standard) branch of the complex cubic root,

$$p_3(\zeta) := \begin{cases} \zeta^{1/3} & \text{when } \mathrm{Re}\,\zeta > 0\,; \\ -(-\zeta)^{1/3} & \text{when } \mathrm{Re}\,\zeta < 0\,. \end{cases}\,. \tag{12.87}$$

From (12.83) we see that there exists a labelling $A : \mathbb{R} \to \{0, \pm\}$, such that

$$\Omega(\lambda) = \widehat{\Omega}_{A(\omega)}(\Lambda(\lambda))\,. \tag{12.88}$$

Similarly as before, we conclude that Ω and the roots are continuous (cf. **SP-1.**) on $\mathbb{R}$ and on the half-spaces $\{\zeta \in \mathbb{C} : \pm\mathrm{Re}\,\zeta > 0\}$, respectively. This implies that $A(\lambda_0 - 0) \neq A(\lambda_0 + 0)$ if and only if $\Lambda(\lambda_0) = 0$. From the a priori estimate $|\nu(\lambda)| \lesssim |\lambda|^{1/3}$ (cf. (12.13b)) we see that there exists $\delta \sim 1$ such that $\Lambda(\lambda) \neq 0$, for $0 < |\lambda| \le \delta$. Hence, we conclude

$$A(\lambda) = a_\pm\,, \qquad \forall\, \lambda \in \pm(0, \delta]\,. \tag{12.89}$$

The choices $a_+ = -$ and $a_- = +$ are excluded by the selection principle **SP-3.**: Similarly as (12.79), we get

$$\pm\,(\mathrm{sign}\,\lambda)\,\mathrm{Im}\,\widehat{\Omega}_\pm(\Lambda(\lambda)) = \frac{\sqrt{3}}{2}|\lambda|^{1/3} + \mathcal{O}\Big(\mu(\lambda)\,\lambda^{1/3}\Big) \ge |\lambda|^{1/3} - C|\lambda|^{2/3}\,. \tag{12.90}$$

From this it follows that $\mathrm{Im}\,\widehat{\Omega}_-(\Lambda(\lambda)) < 0$ for small $|\lambda| > 0$. Thus **SP-3.** implies $a_\pm \neq \mp$.

We will now exclude the choices $a_\pm = 0$. Similarly as (12.90) we use (12.13b) to get

$$\begin{aligned} \operatorname{Re}\widehat{\Omega}_0(\Lambda(\lambda)) \,&\leq\, -\lambda^{1/3} + C\lambda^{2/3} \\ \operatorname{Im}\widehat{\Omega}_0(\Lambda(\lambda)) \,&\lesssim\, |\nu(\lambda)|\,|\lambda|^{1/3} \lesssim |\lambda|^{2/3}\,, \end{aligned} \tag{12.91}$$

for $\lambda \geq 0$. If $a_+ = 0$, then these two bounds together would violate **SP-4.** The choice $a_- = 0$ is excluded similarly. Thus we are left with the unique choices $a_+ = +$ and $a_- = -$.

The expansion (12.82) is obtained similarly as in the proof of Lemma 12.11. First, we use (12.84) and (12.90) to solve for $\operatorname{Im}\Theta(\omega)$. Then we take the imaginary part of (12.14b) to express $v_x(\tau_0 + \omega)$ in terms of $\operatorname{Im}\Theta(\omega)$. We identify

$$h_x \,:=\, \frac{\sqrt{3}}{2}\Big(\frac{\langle |m| f\rangle}{\psi}\Big)^{1/3}|m_x| f_x\,,$$

in the expansion (12.82). From $\psi, |m|, f \sim 1$ it follows that $h_x \sim 1$. □

12.4 Two nearby edges

In this subsection we consider the generic case of the cubic (12.63), where neither the cubic nor the quadratic term can be neglected. First, we remark that Lemma 12.11 becomes ineffective as $|\sigma|$ approaches zero since the cubic term of

$$\psi\,\Theta(\omega)^3 + \sigma\,\Theta(\omega)^2 + (1+\nu(\omega))\langle |m| f\rangle\,\omega \,=\, 0\,, \qquad \psi, \sigma \neq 0\,, \tag{12.92}$$

was treated as a perturbation of a quadratic equation along with $\nu(\omega)$ in the proof. Thus we need to consider the case where $|\sigma|$ is small. Indeed, we will assume that $|\sigma| \leq \sigma_*$, where $\sigma_* \sim 1$ is a threshold parameter that will be adjusted so that the analysis of the cubic (12.92) simplifies sufficiently. In particular, we will choose σ_* so small that the number $\widehat{\Delta} = \widehat{\Delta}(\tau_0) > 0$ defined by

$$\widehat{\Delta} \,:=\, \frac{4}{27\langle |m| f\rangle}\,\frac{|\sigma|^3}{\psi^2}\,, \tag{12.93}$$

satisfies

$$\widehat{\Delta} \,\sim\, |\sigma|^3\,, \qquad \text{provided} \quad |\sigma| \leq \sigma_*\,. \tag{12.94}$$

Note that the existence of $\sigma_* \sim 1$ such that (12.94) holds follows from $f_x, |m_x| \sim 1$ and the stability of the cubic, (12.11). Indeed, (12.11) shows that $\psi \sim 1$ when $|\sigma| \leq \sigma_*$ for some small enough $\sigma_* \sim 1$. We will see below (cf. Lemma 12.17) that $\widehat{\Delta}(\tau_0)$ approximates the gap length $\Delta(\tau_0)$ when the latter is small.

Introducing the normal coordinates,

$$\begin{aligned} \lambda \,&:=\, 2\,\frac{\omega}{\widehat{\Delta}} \\ \Omega(\lambda) \,&:=\, 3\,\frac{\psi}{|\sigma|}\Theta\Big(\frac{\widehat{\Delta}}{2}\lambda\Big) \,+\, \operatorname{sign}\sigma\,, \end{aligned} \tag{12.95}$$

the generic cubic (12.92) reduces to

$$\Omega(\lambda)^3 - 3\,\Omega(\lambda) \,+\, 2\,\Lambda(\lambda) \,=\, 0\,, \tag{12.96}$$

with the constant term

$$\Lambda(\lambda) := \operatorname{sign}\sigma + (1+\mu(\lambda))\lambda, \tag{12.97}$$

$$\mu(\lambda) := \nu\Big(\frac{\widehat{\Delta}}{2}\lambda\Big). \tag{12.98}$$

Here, $\Lambda(\lambda)$ is considered as a perturbation of $\operatorname{sign}\sigma + \lambda$. Indeed, from (12.13b) and (12.98) we see that $|\mu(\lambda)| \lesssim \delta^{1/3}$.

The left hand side of equation (12.96) is a cubic polynomial of $\Omega(\lambda)$ with a constant term $\Lambda(\lambda)$. It is very similar to (12.35) but with an opposite sign in the linear term. Cardano's formula in this case read as follows.

LEMMA 12.13 (Roots of reduced cubic with negative linear coefficient). *For any $\zeta \in \mathbb{C}$,*

$$\Omega^3 - 3\Omega + 2\zeta = (\Omega - \widehat{\Omega}_+(\zeta))(\Omega - \widehat{\Omega}_0(\zeta))(\Omega - \widehat{\Omega}_-(\zeta)), \tag{12.99}$$

where the three root functions $\widehat{\Omega}_\varpi : \mathbb{C} \to \mathbb{C}$, $\varpi = 0, \pm$, have the form

$$\begin{aligned} \widehat{\Omega}_0 &:= -(\Phi_+ + \Phi_-) \\ \widehat{\Omega}_\pm &:= \frac{1}{2}(\Phi_+ + \Phi_-) \pm \mathrm{i}\frac{\sqrt{3}}{2}(\Phi_+ - \Phi_-). \end{aligned} \tag{12.100a}$$

The auxiliary functions $\Phi_\pm : \mathbb{C} \to \mathbb{C}$, are defined by (recall Definition 12.5)

$$\Phi_\pm(\zeta) := \begin{cases} \big(\zeta \pm \sqrt{\zeta^2 - 1}\,\big)^{1/3} & \textit{if} \quad \operatorname{Re}\zeta \geq 1, \\ \big(\zeta \pm \mathrm{i}\sqrt{1-\zeta^2}\,\big)^{1/3} & \textit{if} \quad |\operatorname{Re}\zeta| < 1, \\ -\big(-\zeta \mp \sqrt{\zeta^2 - 1}\,\big)^{1/3} & \textit{if} \quad \operatorname{Re}\zeta \leq -1. \end{cases} \tag{12.100b}$$

On the simply connected complex domains

$$\widehat{\mathbb{C}}_0 := \big\{\zeta \in \mathbb{C} : |\operatorname{Re}\zeta| < 1\big\}, \qquad \textit{and} \qquad \widehat{\mathbb{C}}_\pm := \big\{\zeta \in \mathbb{C} : \pm\operatorname{Re}\zeta > 1\big\}, \tag{12.101}$$

the respective restrictions of $\widehat{\Omega}_a$ are analytic and distinct. Indeed, if $\widehat{\Omega}_a(\zeta) = \widehat{\Omega}_b(\zeta)$ holds for some $a \neq b$ and $\zeta \in \mathbb{C}$, then $\zeta = \pm 1$.

This lemma is analogue of Lemma 12.6 but for (12.96) instead of (12.35). As before the meaning of the symbols $\widehat{\Omega}_a$, λ, etc., is changed accordingly.

Comparing (12.96) and (12.99) we see that there exists a function $A : \mathbb{R} \to \{0, \pm\}$ such that

$$\Omega(\lambda) = \widehat{\Omega}_{A(\lambda)}(\Lambda(\lambda)). \tag{12.102}$$

We will determine the values of A inside the following three intervals

$$\begin{aligned} I_1 &:= -(\operatorname{sign}\sigma)[-\lambda_1, 0), \\ I_2 &:= -(\operatorname{sign}\sigma)(0, \lambda_2], \\ I_3 &:= -(\operatorname{sign}\sigma)[\lambda_3, \lambda_1], \end{aligned} \tag{12.103}$$

which are defined by their boundary points,

$$\lambda_1 := 2\frac{\delta}{\widehat{\Delta}}, \qquad \lambda_2 := 2 - \varrho|\sigma|, \qquad \lambda_3 := 2 + \varrho|\sigma|, \tag{12.104}$$

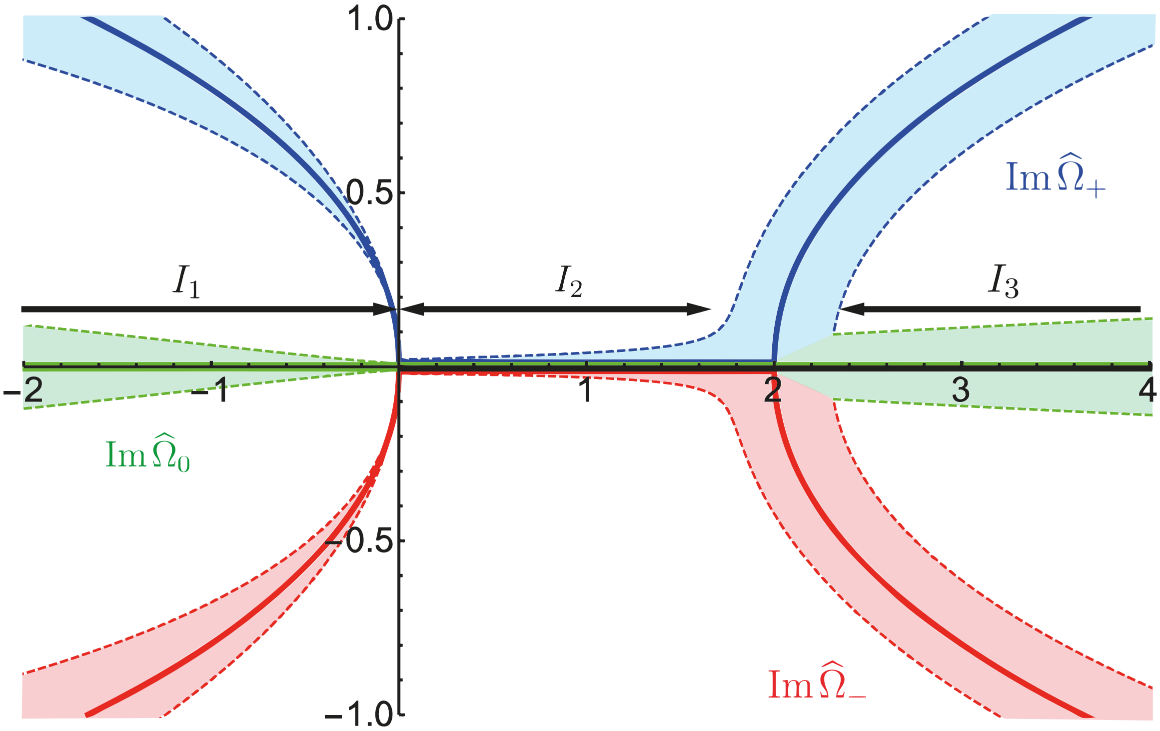

Figure 12.1: Imaginary parts of the three branches of the roots of the cubic equation. The true solution remains within the allowed error margin indicated by the dashed lines.

for some $\varrho \sim 1$. The shape of the imaginary parts of the roots $\widehat{\Omega}_a$ on the intervals I_1, I_2 and I_3 is shown in Figure 12.1. The number λ_1 is the expansion range δ in the normal coordinates. From (12.94) it follows that

$$c_1 \frac{\delta}{|\sigma|^3} \le \lambda_1 \le C_1 \frac{\delta}{|\sigma|^3}, \qquad \text{provided} \quad |\sigma| \le \sigma_* . \tag{12.105}$$

The points λ_2 and λ_3 will act as a lower and an upper bounds for the size of the gap in $\operatorname{supp} v$ associated to the edge τ_0, respectively. Given any $\delta, \varrho \sim 1$ we can choose $\sigma_* \sim 1$ so small that

$$\lambda_1 \ge 4\,, \quad \text{and} \quad 1 \le \lambda_2 < 2 < \lambda_3 \le 3\,, \qquad \text{provided} \quad |\sigma| \le \sigma_* . \tag{12.106}$$

In particular, the intervals (12.103) are disjoint and non-trivial for a triple $(\delta, \varrho, \sigma_*)$ chosen this way. The value $A(\lambda)$ can be uniquely determined using the selection principles if λ lies inside one of the intervals (12.103).

Lemma 12.14 (Choice of roots). *There exist $\delta, \varrho, \sigma_* \sim 1$, such that* (12.106) *holds, and if*

$$|\sigma| \le \sigma_* ,$$

then the restrictions of Ω on the intervals $I_k := I_k(\delta, \varrho, \sigma, \widehat{\Delta})$, defined in (12.103)*, satisfy:*

$$\begin{aligned} \Omega|_{I_1} &= \widehat{\Omega}_+ \circ \Lambda|_{I_1} \\ \Omega|_{I_2} &= \widehat{\Omega}_- \circ \Lambda|_{I_2} \\ \Omega|_{I_3} &= \widehat{\Omega}_+ \circ \Lambda|_{I_3} . \end{aligned} \tag{12.107}$$

Moreover, we have

$$\operatorname{Im} \Omega(-\operatorname{sign}\sigma\, \lambda_3) > 0 . \tag{12.108}$$

The proof of the following simple result is given in Appendix B.5.

LEMMA 12.15 (Stability of roots). *In a connected component of* $\widehat{\mathbb{C}}$, *the roots* (12.100a) *satisfy*

$$\big|\widehat{\Omega}_a(\zeta)-\widehat{\Omega}_a(\xi)\big| \lesssim \min\big\{|\zeta-\xi|^{1/2},|\zeta-\xi|^{1/3}\big\}\,,\qquad \forall\,(\zeta,\xi)\in\widehat{\mathbb{C}}_-^2\cup\widehat{\mathbb{C}}_0^2\cup\widehat{\mathbb{C}}_+^2\,, \tag{12.109}$$

for each $a=-,0,+$.

In particular, suppose ζ *and* ξ *are of the following special form*

$$\begin{aligned}\xi &= -\theta+\lambda\\ \zeta &= -\theta+(1+\mu')\lambda\,,\end{aligned}$$

where $\theta=\pm1$, $\lambda\in\mathbb{R}$ *and* $\mu'\in\mathbb{C}$. *Suppose also that* $|\lambda-2\theta|\geq 6\kappa$, *and* $|\mu'|\leq\kappa$, *for some* $\kappa\in(0,1/2)$. *Then for each* $a=-,0,+$ *the function* $\widehat{\Omega}_a$ *satisfies*

$$\big|\widehat{\Omega}_a(\zeta)-\widehat{\Omega}_a(\xi)\big| \lesssim \frac{\min\big\{|\lambda|^{1/2},\,|\lambda|^{1/3}\big\}}{\kappa^{1/2}}\,|\mu'|\,. \tag{12.110}$$

Using Lemma 12.15 we may treat $\Lambda(\lambda)$ as a perturbation of $\operatorname{sign}\sigma+\lambda$ by a small error term $\lambda\mu(\lambda)$. By expressing the a priori bounds (12.13b) for $\nu(\omega)$ in the normal coordinates (12.95), and recalling that $|\lambda|\leq\lambda_1$ is equivalent to $|\omega|\leq\delta$, we obtain estimates for this error term,

$$|\mu(\lambda)| \,\leq\, C_2|\sigma||\lambda|^{1/3} \tag{12.111a}$$

$$\leq\, C_3\,\delta^{1/3}\,,\qquad \text{provided}\quad |\sigma|\leq\sigma_*\,,\quad |\lambda|\leq\lambda_1\,. \tag{12.111b}$$

In the following we will assume that $\delta\leq(2C_3)^{-3}\sim1$, so that

$$\sup_{\lambda:|\lambda|\leq\lambda_1}|\mu(\lambda)| \,\leq\, \frac{1}{2}\,,\qquad \text{provided}\quad |\sigma|\leq\sigma_*\,. \tag{12.111c}$$

The a priori bound in the middle of (12.13b) also yields the third estimate of μ in terms of Ω and λ. Indeed, inverting (12.95) and using $\Omega(0)=\operatorname{sign}\sigma=1$ (also from (12.95)), we get

$$|\mu(\lambda)| \,\lesssim\, |\sigma||\Omega(\lambda)-\Omega(0)|\,+\,|\sigma|^3|\lambda|\,,\qquad \text{provided}\quad |\sigma|\leq\sigma_0\,. \tag{12.111d}$$

Recall that the solution of the QVE has the symmetry $m(-\tau)=-\overline{m(\tau)}$. Since the sign of $\sigma(\tau)=\langle(\operatorname{sign}\operatorname{Re}m(\tau))\,f(\tau)^3\rangle$ changes under this transformation, we may restrict our analysis to the case $\operatorname{sign}\sigma=-1$ without loss of generality.

We will use the notations $\varphi(\tau{+}0)$ and $\varphi(\tau{-}0)$, for the right and the left limits $\lim_{\xi\to\tau:\xi>\tau}\varphi(\xi)$ and $\lim_{\xi\to\tau:\xi>\tau}\varphi(\xi)$, respectively.

PROOF OF LEMMA 12.14. Let us assume $\operatorname{sign}\sigma=-1$. We will consider $\delta\sim1$ and $\varrho\sim1$ as free parameters which can be adjusted to be as small and large as we need, respectively. Given $\delta\sim1$ and $\varrho\sim1$ the threshold $\sigma_*\sim1$ is then chosen so small that (12.106) holds.

First we show that $A(\lambda)$ is constant on each I_k, i.e., there are three labels $a_k\in\{0,\pm\}$ such that

$$A(\lambda)\,=\,a_k\,,\qquad \forall\,\lambda\in I_k\,,\qquad k=1,2,3\,. \tag{12.112}$$

In order to prove this we first recall that the root functions $\zeta\mapsto\widehat{\Omega}_a(\zeta)$, $a,b=0,\pm$, are continuous on the domains $\widehat{\mathbb{C}}_b$, $b=0,\pm$, and that they may coincide only at points $\operatorname{Re}\zeta=\pm1$ (Indeed, the roots coincide only at the two points $\zeta=\pm1$.). From Lemma 12.1 and **SP-1.** we see that $\Lambda,\Omega:\mathbb{R}\to\mathbb{C}$ are continuous. Hence, (12.112) will follow from

$$\Lambda(I_1)\subset\widehat{\mathbb{C}}_-\,,\qquad \Lambda(I_2)\subset\widehat{\mathbb{C}}_0\,,\qquad \Lambda(I_3)\subset\widehat{\mathbb{C}}_+\,, \tag{12.113}$$

since $|\mathrm{Re}\,\zeta| \neq 1$ for $\zeta \in \cup_a \widehat{\mathbb{C}}_a$ (cf. (12.101)).

From (12.97) and (12.111c) we get

$$\mathrm{Re}\,\Lambda(\lambda) \;=\; -1-(1+\mu(\lambda))|\lambda| \;\leq\; -1-\frac{1}{2}|\lambda| \;<\; -1\,, \qquad \lambda \in I_1\,, \tag{12.114}$$

and thus $\Lambda(I_1) \subset \widehat{\mathbb{C}}_-$. Similarly, we get the first estimate below:

$$\begin{aligned} -1+\frac{1}{2}|\lambda| \;\leq\; \mathrm{Re}\,\Lambda(\lambda) \;&\leq\; -1+(1+C_2|\sigma||\lambda|^{1/3})\,|\lambda| \\ &\leq\; 1-(\varrho-2^{4/3}C_2)|\sigma|\,, \qquad \lambda \in I_2\,. \end{aligned} \tag{12.115}$$

For the second inequality we have used (12.111a), while for the last inequality we have estimated $\lambda \leq \lambda_2 = 2-\varrho|\sigma|$. Taking ϱ sufficiently large yields $\Lambda(I_2) \subset \widehat{\mathbb{C}}_0$.

In order to show $\Lambda(I_3) \subset \widehat{\mathbb{C}}_+$ we split $I_3 = [\lambda_3, \lambda_1]$ into two parts, $[\lambda_3, 4]$ and $(4, \lambda_1]$ (note that $[\lambda_3, 4] \subset I_3$ by (12.106)). In the first part we estimate similarly as in (12.115) to get

$$\mathrm{Re}\,\Lambda(\lambda) \;\geq\; -1+(1-C_2|\sigma|\lambda^{1/3})\,\lambda \;\geq\; 1+(\varrho-4^{4/3}C_2)|\sigma|\,, \qquad \lambda_3 \leq \lambda \leq 4\,. \tag{12.116}$$

Taking $\varrho \sim 1$ large enough, the right most expression is larger than 1. If $\lambda_1 > 4$, we use the rough bound (12.111c) similarly as in (12.114) to obtain

$$\mathrm{Re}\,\Lambda(\lambda) \;=\; -1-(1+\mu(\lambda))\,\lambda \;\geq\; -1+\frac{\lambda}{2} \;>\; 1\,, \qquad 4 < \lambda \leq \lambda_1\,.$$

Together with (12.116) this shows that $\Lambda(I_3) \subset \widehat{\mathbb{C}}_+$.

Next, we will determine the three values a_k using the four selection principles of Lemma 12.9.

Choice of a_1: The initial condition, i.e., **SP-2.**, must be satisfied,

$$\widehat{\Omega}_{a_1}(-1-0) \;=\; \widehat{\Omega}_{a_1}(\Lambda(0-0)) \;=\; \Omega(0) \;=\; -1\,.$$

This excludes the choice $a_1 = 0$ since $\widehat{\Omega}_0(-1-0) = 2$. The choice $a_1 = -$ is excluded using 1/2-Hölder continuity (12.109) of the roots (12.100a) inside the domain $\widehat{\mathbb{C}}_-$, and (12.111b):

$$\mathrm{Im}\,\widehat{\Omega}_-(\Lambda(-\xi)) \;=\; \mathrm{Im}\Big[\widehat{\Omega}_-(-1-\xi)+\mathcal{O}\big(|\mu(-\xi)\,\xi|^{1/2}\big)\Big] \;\leq\; -c\,\xi^{1/2}\,, \qquad 0 \leq \xi \leq 1\,. \tag{12.117}$$

For the last bound we have used (12.111a) and the bound

$$\pm\mathrm{Im}\,\widehat{\Omega}_\pm(1+\xi) \;=\; \pm\mathrm{Im}\,\widehat{\Omega}_\pm(-1-\xi) \;\geq\; c_3\,\xi^{1/2}\,, \qquad 0 \leq \xi \leq 1\,, \tag{12.118}$$

which follows from the explicit formulas (12.100a). Since (12.117) violates **SP-3.** we are left with only one choice: $a_1 = +$.

Choice of a_2: Since $\widehat{\Omega}_+(-1+0) = 2$, while $\Omega(0) = -1$, we exclude the choice $a_2 = +$ using **SP-2.** Moreover, from the explicit formulas of the roots (12.100a) it is easy to see that $\mathrm{Im}\,\widehat{\Omega}_a|_{(-1,1)} = 0$ for each of the three roots $a = \pm, 0$. Similarly as in (12.117) we estimate for small enough $\lambda > 0$ the real and imaginary part of $\widehat{\Omega}_0 \circ \Lambda$ by

$$\begin{aligned} \mathrm{Re}\,\widehat{\Omega}_0(\Lambda(\lambda)) &\leq -1-c\lambda^{1/2}+C|\sigma|^{1/2}\lambda^{2/3} \\ \big|\mathrm{Im}\,\widehat{\Omega}_0(\Lambda(\lambda))\big| \;&=\; \big|\,0\,+\,\mathcal{O}\big(|\mu(\lambda)\lambda|^{1/2}\big)\big| \;\lesssim\; |\sigma|^{1/2}\lambda^{2/3}\,. \end{aligned} \tag{12.119}$$

If $a_2 = 0$, then (12.119) would violate **SP-4.** for small $\lambda > 0$. We are left with only one choice: $a_2 = -$.

CHOICE OF a_3: Using the formulas (12.100a) we get

$$\big\{\widehat{\Omega}_0(1\pm 0)\,,\,\widehat{\Omega}_+(1\pm 0)\,,\,\widehat{\Omega}_-(1\pm 0)\big\} \,=\, \{1,-2\}\,.$$

Thus, the 1/2-Hölder regularity (12.109) of the roots (outside the branch cuts) implies

$$\operatorname{dist}\big(\widehat{\Omega}_a(\zeta),\{1,-2\}\big) \,\lesssim\, |\zeta-1|^{1/2}, \qquad \zeta\in\mathbb{C}\,,\; a=0,\pm\,. \tag{12.120}$$

We will apply this estimate for

$$\zeta=\Lambda(\lambda)=1+\mathcal{O}\big(|\lambda-2|+|\sigma|\big)\,, \qquad \lambda\in[\lambda_2,\lambda_3]\,.$$

Using (12.111a) to estimate $\mu(\lambda)$, and recalling that $|\lambda-2|\lesssim|\sigma|$ for $\lambda\in[\lambda_2,\lambda_3]$, (12.102) and (12.120) yield

$$\operatorname{dist}\big(\Omega(\lambda),\{1,-2\}\big) \,\le\, \max_a \operatorname{dist}\big(\widehat{\Omega}_a(\Lambda(\lambda)),\{1,-2\}\big) \,\lesssim\, |\sigma|^{1/2}\,, \qquad \lambda\in[\lambda_2,\lambda_3]\,. \tag{12.121}$$

In particular, taking $\sigma_*\sim 1$ sufficiently small (12.121) implies for every $|\sigma|\le\sigma_*$,

$$\Omega([\lambda_2,\lambda_3]) \,\subset\, \mathbb{B}(1,1)\cup\mathbb{B}(-2,1)\,,$$

where $\mathbb{B}(\zeta,\rho)\subset\mathbb{C}$ is a complex ball of radius ρ centred at ζ. Since $a_2=-$ and $\widehat{\Omega}_-(1-0)=1$ we see that $\Omega(\lambda_2-0)\in\mathbb{B}(1,1)$. The continuity of Ω (cf. **SP-1**) thus implies

$$\Omega([\lambda_2,\lambda_3]) \,\subset\, \mathbb{B}(1,1)\,.$$

In particular, $|\Omega(\lambda_3)-1|\le 1$, while $|\widehat{\Omega}_0(\Lambda(\lambda_3))-1|\ge 2$, since $\widehat{\Omega}_0(1+0)=2$ and $\Lambda(\lambda_3)\in\widehat{\mathbb{C}}_+$ is close to 1. This shows that $a_3\neq 0$.

In order to choose a_3 among $\pm$, we use (12.109) and the symmetry $\operatorname{Im}\widehat{\Omega}_- = -\operatorname{Im}\widehat{\Omega}_+$ to get

$$\pm\operatorname{Im}\widehat{\Omega}_\pm(\Lambda(\lambda)) \,\ge\, \operatorname{Im}\widehat{\Omega}_+(-1+\lambda) - C|\lambda\mu(\lambda)|^{1/2}\,, \qquad \lambda\in I_3\,. \tag{12.122}$$

Since $\lambda_3=2+\varrho|\sigma|\le 4$ combining (12.118) and (12.111a) yields

$$\pm\operatorname{Im}\widehat{\Omega}_\pm(\Lambda(\lambda_3)) \,\ge\, c\,(\lambda_3-2)^{1/2} - C|\sigma|^{1/2} \,=\, (c\,\varrho^{1/2}-C)|\sigma|^{1/2}\,. \tag{12.123}$$

Taking $\varrho\sim 1$ sufficiently large, the last lower bound becomes positive. Thus, the choice $a_3=-$ is excluded by **SP-3.**. We are left with only one choice: $a_3=+$. The estimate (12.108) follows from (12.123). □

For the rest of the analysis we always assume that the triple $(\delta,\varrho,\sigma_*)$ is from Lemma 12.14. Next we determine the shape of the general edge when the associated gap in $\operatorname{supp} v$ is small.

LEMMA 12.16 (Edge shape). *Let $\tau_0\in\partial\operatorname{supp} v$ and suppose $|\sigma(\tau_0)|\le\sigma_*$, where $\sigma_*\sim 1$ is from Lemma 12.14. Then $\sigma=\sigma(\tau_0)\neq 0$, and $\operatorname{supp} v$ continues in the direction $\operatorname{sign}\sigma$ such that*

$$\big|\Omega(\lambda) - \widehat{\Omega}_+(1+|\lambda|)\big| \,\lesssim\, |\sigma|\min\big\{|\lambda|,|\lambda|^{2/3}\big\}\,, \qquad \operatorname{sign}\lambda=\operatorname{sign}\sigma\,. \tag{12.124}$$

In particular,

$$\operatorname{Im}\Omega(\lambda) \,=\, \Psi_{\mathrm{edge}}\Big(\frac{|\lambda|}{2}\Big) \,+\, \mathcal{O}\Big(|\sigma|\min\big\{|\lambda|,|\lambda|^{2/3}\big\}\Big)\,, \qquad \operatorname{sign}\lambda=\operatorname{sign}\sigma\,, \tag{12.125}$$

where the function $\Psi_{\mathrm{edge}}:[0,\infty)\to[0,\infty)$, defined in (6.13a)*, satisfies*

$$\Psi_{\mathrm{edge}}(\lambda) \,=\, \operatorname{Im}\widehat{\Omega}_+(1+2\lambda)\,, \qquad \lambda\ge 0\,. \tag{12.126}$$

We remark that from (6.13a) one obtains:

$$\Psi_{\mathrm{edge}}(\lambda) \,\sim\, \min\bigl\{\lambda^{1/2}, \lambda^{1/3}\bigr\}\,, \qquad \lambda \ge 0\,. \tag{12.127}$$

PROOF OF LEMMA 12.16. The bound $\sigma \neq 0$ follows from Lemma 12.12. The statement concerning the direction of $\operatorname{supp} v$ follows from Lemma 12.11. Without loss of generality we assume $\sigma > 0$. Let $\delta, \sigma_* \sim 1$ be from Lemma 12.14. The relation (12.124) is trivial when $|\lambda| \gtrsim \delta/|\sigma|^3$ since $\Omega(\lambda)$ and $\widehat{\Omega}_+(1+\lambda)$ are both $\mathcal{O}(\lambda^{1/3})$ by (12.12a) and (12.100), respectively. Thus we consider only the case $\lambda \in I_1 = (0, \lambda_1]$. Using (12.107) and the stability estimate (12.110), with $\rho = 1$, we get

$$\begin{aligned} \Omega(\lambda) \,&=\, \widehat{\Omega}_+(1+\lambda+\mu(\lambda)\lambda) \\ &=\, \widehat{\Omega}_+(1+\lambda) \,+\, \mathcal{O}\Bigl(\mu(\lambda)\min\bigl\{\lambda^{1/2}, \lambda^{1/3}\bigr\}\Bigr)\,, \qquad \lambda \in I_1 = (0,\lambda_1]\,. \end{aligned} \tag{12.128}$$

From (12.111d) we obtain

$$|\mu(\lambda)| \,\lesssim\, |\sigma|\, \bigl|\widehat{\Omega}_+(1+(1+\mu(\lambda))\lambda) - \widehat{\Omega}_+(1+0)\bigr| \,+\, |\sigma|^3\lambda\,. \tag{12.129}$$

The stability estimate (12.109) then yields

$$\begin{aligned} \bigl|\widehat{\Omega}_+(1+(1+\mu(\lambda))\lambda) - \widehat{\Omega}_+(1+0)\bigr| \,&\lesssim\, \min\Bigl\{\bigl|(1+\mu(\lambda))\lambda\bigr|^{1/2}, \bigl|(1+\mu(\lambda))\lambda\bigr|^{1/3}\Bigr\} \\ &\lesssim\, \min\bigl\{\lambda^{1/2}, \lambda^{1/3}\bigr\}\,, \end{aligned} \tag{12.130}$$

where we have used the first estimate of (12.108) to obtain $|(1+\mu(\lambda))\lambda| \sim \lambda$. Plugging (12.130) into (12.129) and using the resulting bound in (12.128) to estimate $\mu(\lambda)$ yields (12.124). The formula (12.125) follows by taking the imaginary part of (12.124) and using (12.126). In order to see that (12.126) is equivalent to our original definition (6.13a) of $\Psi_{\mathrm{edge}}(\lambda)$ we rewrite the right hand side of (12.126) using (12.100a) and (12.100b). □

We know now already from Lemma 12.14 that $\operatorname{Im}\Omega$ is small in I_2 since $a_2 = -$ and $\operatorname{Im}\widehat{\Omega}_-(-1+\lambda) = 0$, $\lambda \in I_2$. The next result shows that actually $\operatorname{Im}\Omega|_{I_2} = 0$ which bounds the size of the gap $\Delta(\tau_0)$ from below.

LEMMA 12.17 (Size of small gap). *Suppose $\tau_0 \in \partial \operatorname{supp} v$. Then the gap length $\Delta(\tau_0)$ (cf. (12.58)) is approximated by $\widehat{\Delta}(\tau_0)$ for small $|\sigma(\tau_0)|$, such that*

$$\frac{\Delta(\tau_0)}{\widehat{\Delta}(\tau_0)} \,=\, 1 \,+\, \mathcal{O}\bigl(\sigma(\tau_0)\bigr)\,. \tag{12.131}$$

In general $\Delta(\tau_0) \sim |\sigma(\tau_0)|^3 \lesssim \widehat{\Delta}(\tau_0)$.

PROOF. Let $(\delta, \varrho, \sigma_*)$ be from Lemma 12.14. If $\sigma = \sigma(\tau_0)$ satisfies $|\sigma| \ge \sigma_*$, then $\Delta = \Delta(\tau_0) \gtrsim |\sigma|^3$ by the second line of (12.70). On the other hand, $\Delta \le 2$ and $|\sigma| \lesssim 1$ by definitions (12.58) and (11.10), respectively. Thus, we find $\Delta \sim |\sigma|^3$. Since $\psi = \psi(\tau_0) \lesssim 1$, we see from (12.93) that $\widehat{\Delta} = \widehat{\Delta}(\tau_0) \gtrsim |\sigma|^3$. Thus, the lemma holds for $|\sigma| \ge \sigma_*$. Therefore, from now on we will assume $0 < |\sigma| \le \sigma_*$ ($\sigma \neq 0$ by Lemma 12.16). Moreover, it suffices to consider only the case $\sigma < 0$ without loss of generality.

Let us define the gap length $\lambda_0 = \lambda_0(\tau_0)$ in the normal coordinates as

$$\lambda_0 \,:=\, \inf\bigl\{\lambda > 0 : \operatorname{Im}\Omega(\lambda) > 0\bigr\}\,. \tag{12.132}$$

Comparing this with (12.58) shows

$$\lambda_0 \,=\, 2\,\frac{\Delta}{\widehat{\Delta}}\,. \tag{12.133}$$

From (12.108) we already see that $\lambda_0 \le \lambda_3$, which is equivalent to

$$\Delta \,\le\, \bigl(1 + \frac{\varrho}{2}|\sigma|\bigr)\widehat{\Delta}\,. \tag{12.134}$$

Since $\varrho \sim 1$ the estimate (12.131) hence follows if we prove the lower bound,

$$\Delta \,\ge\, (1 - C|\sigma|)\,\widehat{\Delta}\,. \tag{12.135}$$

Using the representation (12.102) and the perturbation bound (12.109) we get

$$\operatorname{Im}\Omega(\lambda) \,=\, \operatorname{Im}\widehat{\Omega}_-(-1+\lambda) + \mathcal{O}\bigl(|\lambda\mu(\lambda)|^{1/2}\bigr) \,\le\, 0 \,+\, C_1|\sigma|^{1/2}\,, \qquad \forall\,\lambda \in I_2\,. \tag{12.136}$$

We will show that $\lambda \mapsto \operatorname{Im}\Omega(\lambda)$, grows at least like a square root function on the domain $\{\lambda : \operatorname{Im}\Omega(\lambda) \le c\varepsilon\}$. More precisely, we will show that if $\lambda_0 \le 2$, then

$$\operatorname{Im}\Omega(\lambda_0 + \xi) \,\gtrsim\, \xi^{1/2}\,, \qquad 0 \le \xi \le 1\,. \tag{12.137}$$

Assuming that (12.137) is known, the estimate (12.135) follows from (12.136) and (12.137). Indeed, if $\lambda_0 \ge \lambda_2 = 2 - \varrho|\sigma|$ then (12.136) is immediate as $\varrho \sim 1$. On the other hand, if $\lambda_0 < \lambda_2$, then

$$c_0\,(\lambda_2 - \lambda_0)^{1/2} \,\le\, \operatorname{Im}\Omega(\lambda_2) \,\le\, C_1|\sigma|^{1/2}\,.$$

Solving this for λ_0 yields

$$\lambda_0 \,\ge\, \lambda_2 - (C_1/c_0)^2|\sigma| \,\ge\, 2 - C\,|\sigma|\,,$$

where $\lambda_2 = 2 - \varrho|\sigma|$ with $\varrho \sim 1$ (cf. (12.104)) has been used to get the last estimate. Using (12.133) we see that this equals (12.135). Together with (12.134) this proves (12.131).

In order to prove the growth estimate (12.137), we express it in the original coordinates $(\omega, v(\tau_0+\omega))$ using (12.95), (12.7), $v(\tau_0+\Delta) = 0$, and $f, |m| \sim 1$ (Note that $b = f$ since $v(\tau_0) = 0$):

$$v(\tau_0 + \Delta + \widetilde{\omega}) \,\gtrsim\, \min\Bigl\{\bigl(1 + \widehat{\Delta}(\tau_0)^{-1/6}\bigr)\,\widetilde{\omega}^{1/2}, \widetilde{\omega}^{1/3}\Bigr\}\,, \qquad 0 \le \widetilde{\omega} \le \delta\,. \tag{12.138}$$

Applying Lemma 12.16 with $\tau_0 + \Delta$ as the base point yields

$$v(\tau_0 + \Delta + \widetilde{\omega}) \,\sim\, \min\Bigl\{\bigl(1 + \widehat{\Delta}(\tau_0 + \Delta\,)^{-1/6}\bigr)\,\widetilde{\omega}^{1/2}, \widetilde{\omega}^{1/3}\Bigr\}\,, \qquad 0 \le \widetilde{\omega} \le \delta\,. \tag{12.139}$$

The relation (12.139) implies (12.138), provided we show

$$\widehat{\Delta}(\tau_0 + \Delta) \,\lesssim\, \widehat{\Delta}(\tau_0)\,, \qquad \text{for} \quad \Delta \lesssim \widehat{\Delta}(\tau_0)\,. \tag{12.140}$$

From the definition (12.93) we get

$$\widehat{\Delta}(\tau_0 + \Delta) \,\sim\, \frac{|\sigma(\tau_0 + \Delta)|^3}{\psi(\tau_0 + \Delta)^2}\,. \tag{12.141}$$

Using the upper bound (12.134) and (12.94) we see that

$$\Delta \,\lesssim\, \widehat{\Delta}(\tau_0) \,\sim\, |\sigma(\tau_0)|^3\,,$$

for sufficiently small $\sigma_* \sim 1$. Since $\sigma(\tau)$ is 1/3-Hölder continuous in τ, we get

$$|\sigma(\tau_0 + \Delta)| \le |\sigma(\tau_0)| + C\Delta^{1/3} \lesssim |\sigma(\tau_0)|. \tag{12.142}$$

From the stability of the cubic, (12.11), it follows that for small enough $\sigma_* \sim 1$ we have

$$\psi(\tau_0 + \Delta) \sim \psi(\tau_0) \sim 1.$$

Plugging this together with (12.142) into (12.141) yields (12.140). □

We have now covered all the parameter regimes of σ and ψ satisfying (12.11). Combining the preceding lemmas yields the expansion around general base points τ_0 where $v(\tau_0) = 0$. We will need the following representation of the edge shape function (6.13a) below:

$$\Psi_{\mathrm{edge}}(\lambda) = \frac{\lambda^{1/2}}{\sqrt{3}}(1 + \widetilde{\Psi}(\lambda)), \qquad \lambda \ge 0, \tag{12.143}$$

where the smooth function $\widetilde{\Psi} : [0,\infty) \to \mathbb{R}$ has uniformly bounded derivatives, and $\widetilde{\Psi}(0) = 0$.

Proof of Proposition 12.8. Let $\tau_0 \in \operatorname{supp} v$ satisfy $v(\tau_0) = 0$. If $\sigma(\tau_0) = 0$, then the expansion (12.61) follows directly from Lemma 12.12.

In the case $0 < |\sigma(\tau_0)| \le \sigma_*$ (12.125) in Lemma 12.16 yields (12.59) with $\widehat{\Delta} = \widehat{\Delta}(\tau_0)$ in place of $\Delta = \Delta(\tau_0)$. Here, the threshold $\sigma_* \sim 1$ is fixed by Lemma 12.14. We will show that replacing $\widehat{\Delta}$ with Δ in (12.59) yields an error that is so small that it can be absorbed into the sub-leading order correction of (12.59). Since the smooth auxiliary function $\widetilde{\Psi}$ in the representation (12.143) of Ψ_{edge} has uniformly bounded derivatives, we get for every $0 \le \lambda \lesssim 1$,

$$\Psi_{\mathrm{edge}}((1+\epsilon)\lambda) = (1+\epsilon)^{1/2}\Psi_{\mathrm{edge}}(\lambda) + \mathcal{O}\big(\epsilon \min\{\lambda^{3/2}, \lambda^{1/3}\}\big), \qquad \lambda \ge 0, \tag{12.144}$$

provided the size $|\epsilon| \lesssim 1$ of $\epsilon \in \mathbb{R}$ is sufficiently small. On the other hand, if $|\lambda| \gtrsim 1$ then (12.144) follows from (12.110) of Lemma 12.15. Now by Lemma 12.17 we have $\widehat{\Delta} = (1 + |\sigma|\kappa)\Delta$, where $\Delta = \Delta(\tau_0)$ and the constant $\kappa \in \mathbb{R}$ is independent of λ, and can be assumed to satisfy $|\kappa| \le 1/2$ (otherwise we reduce $\sigma_* \sim 1$). Thus applying (12.144) with $\epsilon = |\sigma|\kappa = \mathcal{O}(\Delta^{1/3})$, yields

$$|\sigma|\,\Psi_{\mathrm{edge}}\bigg(\frac{\omega}{\widehat{\Delta}}\bigg) = \frac{(1+|\sigma|\kappa)^{1/2}|\sigma|}{\Delta^{1/3}}\,\Delta^{1/3}\,\Psi_{\mathrm{edge}}\bigg(\frac{\omega}{\Delta}\bigg) + \mathcal{O}\bigg(\min\bigg\{\frac{|\omega|^{3/2}}{\Delta^{5/6}}, |\omega|^{1/3}\bigg\}\bigg), \qquad \omega \ge 0.$$

Here, the error on the right hand side is of smaller size than the subleading order term in the expansion (12.59).

From (12.14) we identify the formula for h_x, in the case $0 < |\sigma| \le \sigma_*$:

$$h_x := \begin{cases} \frac{(1+|\sigma|\kappa)^{1/2}}{3\psi}\frac{|\sigma|}{\Delta^{1/3}}|m_x| f_x & \text{when} \quad 0 < |\sigma| \le \sigma_*\,; \\ \sqrt{\frac{3\,\Delta^{1/3}}{|\sigma|}}\,h_x' & \text{when} \quad |\sigma| > \sigma_*\,. \end{cases} \tag{12.145}$$

For $|\sigma| \le \sigma_*$ we used (12.125). In the case $|\sigma| > \sigma_*$, the function h_x' is from (12.70), and the function h is defined such that

$$h_x'\bigg|\frac{\omega}{\sigma}\bigg|^{1/2} = h_x\,\Delta^{1/3}\,\Psi_{\mathrm{edge}}\bigg(\frac{\omega}{\Delta}\bigg) + \mathcal{O}\bigg(\frac{|\omega|^{3/2}}{\Delta^{7/6}}\bigg). \tag{12.146}$$

Here, the second term originates from the representation (12.143) of Ψ_{edge}. This proves (12.59).

Finally, suppose τ_0 and τ_1 are the opposite edges of $\operatorname{supp} v$ separated by a small gap of length $\Delta \lesssim \sigma_*^3$, between them. Now, $f(\tau)$, $|m(\tau)|$ and $\psi(\tau)$ are 1/3-Hölder continuous in τ, and satisfy $f, |m|, \psi \sim 1$. Thus, the terms constituting h_x in the case $|\sigma| \le \sigma_*$ in (12.145) satisfy

$$\frac{f_x(\tau_1)}{f_x(\tau_0)} = 1 + \mathcal{O}(\Delta^{1/3}), \qquad \frac{|m_x(\tau_1)|}{|m_x(\tau_0)|} = 1 + \mathcal{O}(\Delta^{1/3}), \qquad \frac{\psi(\tau_1)}{\psi(\tau_0)} = 1 + \mathcal{O}(\Delta^{1/3}). \quad (12.147)$$

Of course, $\Delta = \Delta(\tau_0) = \Delta(\tau_1)$. Moreover, by Lemma 12.17,

$$\frac{\widehat{\Delta}(\tau_1)}{\widehat{\Delta}(\tau_0)} = 1 + \mathcal{O}(\Delta^{1/3}). \quad (12.148)$$

Using (12.93) we express $|\sigma|$ in terms of $\widehat{\Delta}, f, |m|, \psi$, and hence (12.147) and (12.148) imply

$$\frac{|\sigma(\tau_1)|}{|\sigma(\tau_0)|} = 1 + \mathcal{O}(\Delta^{1/3}). \quad (12.149)$$

Thus, combining (12.147), (12.148), and (12.149) we see from (12.145) that $h(\tau_1) = h(\tau_0) + \mathcal{O}_{\mathscr{B}}(\Delta^{1/3})$. This proves the last remaining claim of the proposition. □

12.5 Proofs of Theorems 6.4 and 6.9

Let us recall the definition (12.1) of $\mathbb{D}_\varepsilon$, and define for every $\varepsilon > 0$,

$$\mathbb{M}_\varepsilon := \big\{\tau_0 \in \mathbb{D}_\varepsilon : \tau_0 \text{ is a local minimum of } \tau \mapsto \langle v(\tau)\rangle\big\}. \quad (12.150)$$

In the following we split $\mathbb{M}_\varepsilon$ into two parts,

$$\begin{aligned} \mathbb{M}^{(1)} &:= \partial \operatorname{supp} v \\ \mathbb{M}^{(2)}_\varepsilon &:= \mathbb{M}_\varepsilon \backslash \partial \operatorname{supp} v. \end{aligned} \quad (12.151)$$

Proof of Theorem 6.4. Combining Proposition 12.3 and Proposition 12.8 shows that there are constants $\varepsilon_*, \delta_1, \delta_2 \sim 1$ such that the following hold:

1. If $\tau_0 \in \mathbb{M}^{(1)}$, then $\sigma(\tau_0) \neq 0$ and $v_x(\tau_0 + \omega) \ge c_1 |\omega|^{1/2}$, for $0 \le \operatorname{sign} \sigma(\tau_0)\, \omega \le \delta_1$.

2. If $\tau_0 \in \mathbb{M}^{(2)}_{\varepsilon_*}$, then $v_x(\tau_0 + \omega) \ge c_2 \big(v_x(\tau_0) + |\omega|^{1/3}\big)$, for $-\delta_2 \le \omega \le \delta_2$.

In the case 1. we see that each connected component of $\operatorname{supp} v$ must be at least of length $2\delta_1 \sim 1$. This implies (6.14). In particular, since $\sup \operatorname{supp} v \le 2$ (cf. Theorem 6.1), the number of these components K' satisfies the bound $K' \sim 1$.

In order to prove (6.17) and (6.18) we may assume that $\varepsilon \le \varepsilon_*$ and $|\omega| \le \delta$ for some $\varepsilon_*, \delta \sim 1$. Indeed, (6.17) becomes trivial when $C\varepsilon^3 \ge 2 \ge \sup \operatorname{supp} v$. Similarly, if $\langle v(\tau_0)\rangle + |\omega| \gtrsim 1$, then $\langle v(\tau_0)\rangle + \Psi(\omega) \sim 1$ and thus the $\mathcal{O}(\cdots)$-term in (6.18) is $\mathcal{O}(1)$. Since $v \le \|m\|_{\mathbb{R}} \sim 1$, the expansion (6.18) is hence trivial.

Obviously the bounds in the cases 1. and 2. continue to hold if we reduce the parameters $\varepsilon_*, \delta_1, \delta_2$. We choose $\varepsilon_* \sim 1$ so small that $(\varepsilon_*/c_1)^2 \le \delta_1$ and $(\varepsilon_*/c_2)^3 \le \delta_2$. Let us define expansion radius around $\tau_0 \in \mathbb{M}_\varepsilon$ for every $\varepsilon \le \varepsilon_*$

$$\delta_\varepsilon(\tau_0) := \begin{cases} (\varepsilon/c_1)^2 & \text{if } \tau_0 \in \mathbb{M}^{(1)}, \\ (\varepsilon/c_2)^3 & \text{if } \tau_0 \in \mathbb{M}^{(2)}_\varepsilon, \end{cases} \quad (12.152)$$

and the corresponding expansion domains

$$I_\varepsilon(\tau_0) := \begin{cases} \bigl\{\tau_0 + \operatorname{sign}\sigma(\tau_0)\,\xi : 0 \le \xi \le \delta_\varepsilon(\tau_0)\bigr\} & \text{if } \tau_0 \in \mathbb{M}^{(1)}, \\ \bigl[\tau_0 - \delta_\varepsilon(\tau_0),\, \tau_0 + \delta_\varepsilon(\tau_0)\bigr] & \text{if } \tau_0 \in \mathbb{M}_\varepsilon^{(2)}. \end{cases} \tag{12.153}$$

If $\tau \in I_\varepsilon(\tau_0)$ for some $\tau_0 \in \mathbb{M}_\varepsilon$, then either $v_x(\tau) \ge c_1|\tau - \tau_0|^{1/2}$ or $v_x(\tau) \ge c_2|\tau - \tau_0|^{1/3}$ depending on whether τ_0 is an edge or not. In particular, it follows that

$$\langle v(\tau)\rangle \,\ge\, \varepsilon\,, \qquad \forall\, \tau \in \partial I_\varepsilon(\tau_0)\backslash\partial\,\mathrm{supp}\, v\,. \tag{12.154}$$

This implies that each connected component of $\mathbb{D}_\varepsilon$ is contained in the expansion domain $I_\varepsilon(\tau_0)$ of some $\tau_0 \in \mathbb{M}_\varepsilon$, i.e.,

$$\mathbb{D}_\varepsilon \,\subset \bigcup_{\tau_0 \in \mathbb{M}_\varepsilon} I_\varepsilon(\tau_0)\,. \tag{12.155}$$

In order to see this formally let $\tau \in \mathbb{D}_\varepsilon\backslash\mathbb{M}_\varepsilon$ be arbitrary, and define $\tau_0 \in \mathbb{M}_\varepsilon$ as the nearest point of $\mathbb{M}_\varepsilon$ from τ, in the direction,

$$\theta := -\operatorname{sign}\partial_\tau\langle v(\tau)\rangle\,,$$

where $\langle v\rangle$ decreases. In other words, we set

$$\tau_0 := \tau + \theta\,\xi_0\,, \qquad \text{where} \qquad \xi_0 := \inf\bigl\{\xi > 0 : \tau + \theta\,\xi \in \mathbb{M}_\varepsilon\bigr\}\,. \tag{12.156}$$

From (12.156) it follows that if $\tau_0 \in \partial\,\mathrm{supp}\, v$ then $\mathrm{supp}\, v$ continues in the direction $\operatorname{sign}(\tau - \tau_0) = -\theta$ from τ_0. We show that $|\tau - \tau_0| \le \delta_\varepsilon(\tau_0)$. To this end, suppose $|\tau - \tau_0| > \delta_\varepsilon(\tau_0)$, and define

$$\tau_1 \,:=\, \tau_0 \,+\, \operatorname{sign}(\tau - \tau_0)\,\delta_\varepsilon(\tau_0)\,, \tag{12.157}$$

as the point between τ and τ_0 exactly at the distance $\delta_\varepsilon(\tau_0)$ away from τ_0. Now, $\tau_1 \notin \partial\,\mathrm{supp}\, v$ as otherwise τ_0 would not be the nearest point of $\mathbb{M}_\varepsilon$ (cf. (12.156)). On the other hand, by definition we have $\tau_1 \in \partial I(\tau_0)$. Thus, the estimate (12.154) with τ_1 in place of τ_0 yields

$$\langle v(\tau_1)\rangle \,\ge\, \varepsilon \,\ge\, \langle v(\tau)\rangle\,.$$

Since $\langle v\rangle$ is continuously differentiable on the set where $\langle v\rangle > 0$ and $(\tau_1 - \tau)\,\partial_\tau\langle v(\tau)\rangle < 0$ by (12.156) and (12.157), we conclude that $\langle v\rangle$ has a local minimum at some point $\tau_2 \in \mathbb{M}_\varepsilon$ lying between τ and τ_1. But this contradicts (12.156). As $\tau \in \mathbb{D}_\varepsilon\backslash\mathbb{M}_\varepsilon$ was arbitrary (12.155) follows.

From Corollary 12.4 we know that for every $\tau_1, \tau_2 \in \mathbb{M}_\varepsilon^{(2)}$, either

$$|\tau_1 - \tau_2| \,\ge\, c_3 \qquad \text{or} \qquad |\tau_1 - \tau_2| \,\le\, C_3\varepsilon^4\,, \tag{12.158}$$

holds. Let $\{\gamma_k\}$ be a maximal subset of $\mathbb{M}_\varepsilon^{(2)}$ such that its elements are separated at least by a distance c_3. Then the set $\mathbb{M} := \partial\,\mathrm{supp}\, v \cup \{\gamma_k\}$ has the properties stated in the theorem. In particular,

$$\mathbb{D}_\varepsilon \,\subset \bigcup_{\tau_0 \in \partial\,\mathrm{supp}\, v} I_\varepsilon(\tau_0) \,\cup\, \bigcup_k \bigl[\gamma_k - C\varepsilon^3,\, \gamma_k + C\varepsilon^3\bigr]\,,$$

since $\mathbb{M}_\varepsilon^{(2)} + [-C\varepsilon^3, C\varepsilon^3] \,\subset\, \cup_k[\gamma_k - 2C\varepsilon^3, \gamma_k + 2C\varepsilon^3]$ for sufficiently small $\varepsilon \sim 1$. This completes the proof of Theorem 6.4. $\square$

Next we show that the support of a bounded generating density is a single interval provided the rows of S can not be split into two well separated subsets. We measure this separation using the following quantity

$$\xi_S(\kappa) := \sup\left\{ \inf_{\substack{x\in A\\ y\notin A}} \|S_x - S_y\|_1 \,:\, \kappa \le \pi(A) \le 1-\kappa,\, A \subset \mathfrak{X} \right\}, \qquad \kappa \ge 0\,. \tag{12.159}$$

LEMMA 12.18 (Generating density supported on single interval). *Assume S satisfies* **A1-5.** *and $\|\!|m|\!\|_{\mathbb{R}} \le \Phi$ for some $\Phi < \infty$. Then there exist $\xi_*, \kappa_* \sim 1$, such that under the assumption,*

$$\xi_S(\kappa_*) \,\le\, \xi_*\,, \tag{12.160}$$

the conclusions of Theorem 6.9 hold.

In Section 14 we present very simple examples of S which do not satisfy (12.160) and the associated generating density v is shown to have a non-connected support.

PROOF OF THEOREM 6.9. Let $\xi_*, \kappa_* \sim 1$ be from Lemma 12.18. Note that (6.25) is equivalent to $\xi_S(0) \le \xi_*$, and $\xi_S(\kappa') \le \xi_S(\kappa)$, whenever $\kappa' > \kappa$. Thus (6.25) implies $\xi_S(\kappa_*) \le \xi_S(0) \le \xi_*$, and hence the theorem follows from the lemma. □

PROOF OF LEMMA 12.18. Since $\|\!|m|\!\|_{\mathbb{R}} \le \Phi$ Theorem 6.4 and the symmetry (9.12) yield the expansion (6.26b) around the extreme edges $\pm\beta$ where $\beta := \sup\operatorname{supp} v$. In particular, there exists $\delta_1 \sim 1$ such that

$$v_x(-\beta+\omega) \,=\, v_x(\beta-\omega) \,\ge\, c_1|\omega|^{1/2}\,, \qquad |\omega| \le \delta_1\,. \tag{12.161}$$

Let us write

$$m_x(\tau) \,=\, p_x(\tau)u_x(\tau) + \mathrm{i}\,v_x(\tau)$$

where $p_x = \operatorname{sign}\operatorname{Re} m_x \in \{-1,+1\}$ and $u_x := |\operatorname{Re} m_x|, v_x = \operatorname{Im} m_x \ge 0$. By combining the uniform bound $\|\!|m|\!\|_{\mathbb{R}} \le \Phi$ with (8.8) we see that $|m_x| \sim 1$. In particular, there exists $\varepsilon_* \sim 1$ such that

$$\max\{u_x, v_x\} \,\ge\, 2\varepsilon_*\,. \tag{12.162}$$

Since $m_x(\tau)$ is continuous in τ, the constraint (12.162) means that $\operatorname{Re} m_x(\tau)$ can not be zero on the domain

$$\mathbb{K} := \Big\{\tau \in [-2,2] : \sup_x v_x(\tau) \le \varepsilon_*\Big\}\,.$$

If I is a connected component of $\mathbb{K}$, then there is $p^I_x \in \{-1,+1\}$, $x \in \mathfrak{X}$, such that

$$p(\tau) \,=\, p^I\,, \qquad \forall \tau \in I\,.$$

We choose $\varepsilon_* \sim 1$ to be so small that $v_x(\pm\beta \mp \delta_1) \ge \varepsilon_*$ by (12.161) and hence $\operatorname{supp} v$ is not contained in $\mathbb{K}$. Furthermore, we choose ε_* so small that Lemma 12.2 applies, i.e., $v_x > 0$ grows monotonically in $\mathbb{K}$ when $\Pi \ge \Pi_*$.

We will prove the lemma by showing that if some connected component I of $\mathbb{K}$ satisfies,

$$I = [\tau_1, \tau_2] \subset \mathbb{K}\,, \qquad \text{where} \quad -\beta + \delta_1 \le \tau_1 < \tau_2 \le \beta - \delta_1\,, \tag{12.163}$$

then the set

$$A = A^I := \big\{x \in \mathfrak{X} : p^I_x = +1\big\} \tag{12.164}$$

satisfies

$$\pi(A) \sim 1 \tag{12.165a}$$
$$\|S_x - S_y\|_1 \sim 1\,, \quad x \in A\,,\; y \notin A\,. \tag{12.165b}$$

The estimates (12.165) imply $\xi_S(\kappa_*) \geq \xi_*$, with $\kappa_* = \pi(A)$ and $\xi_* \sim 1$. In other words, under the assumption (12.160) each connected component of $\mathbb{K}$ contains either $-\beta$ or β. Together with (6.26b) this proves the remaining estimate (6.26a) of the lemma, and the $\operatorname{supp} v$ is a single interval.

In order to prove (12.165a) we will show below that there is a point $\tau_0 \in I$ such that

$$|\sigma(\tau_0)| \leq C_0 \varepsilon_*^2\,, \tag{12.166}$$

where $\sigma := \langle p f^3 \rangle$ was defined in (11.10). Let $f_- := \inf_x f_x$ and $f_+ := \sup_x f_x$. As m is uniformly bounded, Proposition 8.2 shows that $f_\pm \sim 1$. Hence, (12.166) yields bounds on the size of A,

$$\begin{aligned} \pi(A)\, f_+^3 - (1 - \pi(A))\, f_-^3 \;&\geq\; \sigma(\tau_0) \;\geq\; -C_0 \varepsilon_*^2 \\ \pi(A)\, f_-^3 - (1 - \pi(A))\, f_+^3 \;&\leq\; \sigma(\tau_0) \;\leq\; +C_0 \varepsilon_*^2\,. \end{aligned}$$

Solving for $\pi(A)$, we obtain

$$\frac{f_-^3 - C_0 \varepsilon_*^2}{f_+^3 + f_-^3} \;\leq\; \pi(A) \;\leq\; \frac{f_+^3 + C_0 \varepsilon_*^2}{f_+^3 + f_-^3}\,.$$

By making $\varepsilon_* \sim 1$ sufficiently small this yields (12.165a).

We now show that there exists $\tau_0 \in I$ satisfying (12.166). To this end we remark that at least one (actually exactly one) of the following three alternatives holds true:

(a) The interval I contains a non-zero local minimum τ_0 of $\langle v \rangle$.

(b) The interval I contains a left and right edge $\tau_- \in \partial \operatorname{supp} v$ and $\tau_+ \in \partial \operatorname{supp} v$.

(c) The average generating density $\langle v \rangle$ has a cusp at $\tau_0 \in I \cap (\operatorname{supp} v \backslash \partial \operatorname{supp} v)$ such that $v(\tau_0) = \sigma(\tau_0) = 0$.

In the case (a), since m is smooth on the set where $\langle v \rangle > 0$, Lemma 12.2 implies $\Pi(\tau_0) < \Pi_*$, and thus (12.166) holds for $C_0 \geq \Pi_*$. In the case (b) we know that $\pm\sigma(\tau_\pm) > 0$ by Proposition 12.8. Since $\sigma(\tau)$ is continuous (cf. Lemma 12.1) there hence exists $\tau_0 \in (\tau_-, \tau_+) \subset I$ such that $\sigma(\tau_0) = 0$. Finally, in the case (c) we have $\sigma(\tau_0) = 0$ by Proposition 12.8.

Now we prove (12.165b). Since $v_x \leq u_x \leq |m_x| \leq \Phi$ on I, and m solves the QVE we obtain for every $x \in A$, $y \notin A$ and $\tau \in I$

$$\begin{aligned} \frac{1}{\Phi} \leq \frac{1}{u_x} + \frac{1}{u_y} \leq 2\frac{u_x + u_y}{|m_x m_y|} \leq 2\frac{|(u_x + u_y) + \mathrm{i}\,(v_x - v_y)|}{|m_x||m_y|} \;=\; 2\left|\frac{1}{m_x} - \frac{1}{m_y}\right| \\ = 2\,|\langle S_x - S_y, m \rangle| \;\leq\; 2\,\Phi \|S_x - S_y\|_1\,. \end{aligned} \tag{12.167}$$

Here the definition (12.164) of A is used in the first bound while $u_x \geq v_x$ was used in the second estimate. The bound (12.167) is equivalent to (12.165b) as $\|S_x - S_y\|_1 \geq 1/(2\Phi^2) \sim 1$.

We have shown that $|\sigma| + \langle v \rangle \sim 1$. By using this in Corollary 10.2 we see that $v(\tau)$ is uniformly 1/2-Hölder continuous everywhere. □

13 Stability around small minima of generating density

The next result will imply the statement of (ii) in Theorem 6.10. Since it plays a direct role in Part III on random matrices, we state it here in the form that does not require any knowledge of the preceding expansions and the associated cubic analysis. In fact, together with our main results, Theorem 6.2 and Theorem 6.4, the next proposition is the only information we use in Part III of this work concerning the stability of the QVE.

PROPOSITION 13.1 (Cubic perturbation bound around critical points). *Assume S satisfies* **A1-A5.**, $\|\|m\|\|_{\mathbb{R}} \le \Phi$, *for some* $\Phi < \infty$, *and* $g, d \in \mathscr{B}$ *satisfy the perturbed QVE (6.27) at some fixed* $z \in \overline{\mathbb{H}}$. *There exists* $\varepsilon_* \sim 1$ *such that if*

$$\langle \operatorname{Im} m(z) \rangle \le \varepsilon_* \,, \qquad \text{and} \qquad \|g - m(z)\|_{\mathscr{B}} \le \varepsilon_*. \tag{13.1}$$

then there is a function $s : \overline{\mathbb{H}} \to \mathscr{B}$ *depending only on* S, *and satisfying*

$$\|s(z_1)\|_{\mathscr{B}} \lesssim 1 \,, \qquad \|s(z_1) - s(z_2)\|_{\mathscr{B}} \lesssim |z_1 - z_2|^{1/3} \,, \qquad \forall\, z_1, z_2 \in \overline{\mathbb{H}} \,, \tag{13.2}$$

such that the modulus of the complex variable,

$$\Theta \;:=\; \big\langle s(z) \,, g - m(z) \big\rangle \tag{13.3}$$

bounds the difference $g - m(z)$, *in the following senses:*

$$\|g - m(z)\|_{\mathscr{B}} \;\lesssim\; |\Theta| \,+\, \|d\|_{\mathscr{B}} \tag{13.4a}$$

$$|\langle w, g - m(z) \rangle| \;\lesssim\; \|w\|_{\mathscr{B}} |\Theta| \,+\, \|w\|_{\mathscr{B}} \|d\|_{\mathscr{B}}^2 \,+\, |\langle T(z) w, d \rangle| \,, \qquad \forall\, w \in \mathscr{B} \,. \tag{13.4b}$$

Here the linear operator $T(z) : \mathscr{B} \to \mathscr{B}$ *depends only on* S, *in addition to* z, *and satisfies* $\|T(z)\|_{\mathscr{B} \to \mathscr{B}} \lesssim 1$. *Moreover,* Θ *satisfies a cubic inequality*

$$\big|\Theta^3 + \pi_2 \Theta^2 + \pi_1 \Theta\big| \;\lesssim\; \|d\|_{\mathscr{B}}^2 + |\langle t^{(1)}(z), d \rangle| + |\langle t^{(2)}(z), d \rangle| \,, \tag{13.5}$$

where $\|t^{(k)}(z)\|_{\mathscr{B}} \lesssim 1$, $k = 1, 2$, *depend only on* S *in addition to* z. *The coefficients,* π_1 *and* π_2, *may depend on* S, g *and* z. *They satisfy the estimates,*

$$|\pi_1| \;\sim\; \langle \operatorname{Im} m(z) \rangle^2 \,+\, \widehat{\sigma}(z) \, \langle \operatorname{Im} m(z) \rangle \,+\, \frac{\operatorname{Im} z}{\langle \operatorname{Im} m(z) \rangle} \tag{13.6a}$$

$$|\pi_2| \;\sim\; \langle \operatorname{Im} m(z) \rangle \,+\, \widehat{\sigma}(z) \,, \tag{13.6b}$$

where the 1/3-Hölder continuous function $\widehat{\sigma} : \overline{\mathbb{H}} \to [0, \infty)$ *is determined by* S, *and has the following properties: Let* $\mathbb{M} = \{\alpha_i\} \cup \{\beta_j\} \cup \{\gamma_k\}$ *be the set (6.16) of minima from Theorem 6.4, and suppose* $\tau_0 \in \mathbb{M}$ *satisfies* $|z - \tau_0| = \operatorname{dist}(z, \mathbb{M})$. *If* $\tau_0 \in \partial \operatorname{supp} v = \{\alpha_i\} \cup \{\beta_j\}$, *then*

$$\widehat{\sigma}(\alpha_i) \;\sim\; \widehat{\sigma}(\beta_{i-1}) \;\sim\; (\alpha_i - \beta_{i-1})^{1/3} \tag{13.7a}$$

with the convention $\beta_0 := \alpha_1 - 1$ *and* $\alpha_{K'+1} := \beta_{K'} + 1$. *If* $\tau_0 \notin \partial \operatorname{supp} v = \{\gamma_k\}$, *then*

$$\widehat{\sigma}(\gamma_k) \;\lesssim\; \langle \operatorname{Im} m(\gamma_k) \rangle^2 \,, \tag{13.7b}$$

The comparison relations depend only on the parameters $(\rho, L, \|S\|_{\mathrm{L}^2 \to \mathscr{B}}, \Phi)$.

We remark here that the coefficients π_k do depend on g in addition to S as contrasted to the coefficients μ_k in Proposition 11.2. The important point is that the right hands sides of the comparison relations (13.6a) and (13.6b) are still independent of g. This result is geared towards problems where d and g are *random*. Such problems arise when the resolvent method (cf. Parrt III) is used to study the local spectral statistics of large random matrices (cf. discussion in Section 1 of Part I of this work). In that setup $m(z)$ represents the non-random part of a vector (function) $g(z)$ consisting of the diagonal elements of the resolvent of a random matrix whose matrix of variances is given by S. Random perturbations d are also the reason why we have not simply estimated $|\langle t^{(k)}(z), d\rangle| \leq \|t^{(k)}\|_2\|d\|_2$ in (13.5). For example, in Part III of this work the left hand side of this inequality is typically much smaller than the right hand side due to the cancellations in the weighted average of the random vector d. This effect is called the *fluctuation averaging* mechanism [26, 37] and it holds with very high probability. The continuity and regularity estimates (cf. (13.2)) will be needed to extend high probability bounds for each individual z to all z in a compact set of $\mathbb{H}$.

Proof of Proposition 13.1. Since z is fixed we write $m = m(z)$, etc. By choosing $\varepsilon_* \sim 1$ small enough we ensure that both Lemma 11.1 and Proposition 11.2 are applicable. We choose s such that Θ becomes the component of $u = (g-m)/|m|$ in the direction b exactly as in Proposition 11.2. Hence using the explicit formula (11.7) for the projector P we read off from $\Theta\, b = Pu$, that

$$s := \frac{1}{\langle b^2\rangle}\frac{\overline{b}}{|m|}\,. \tag{13.8}$$

It is clear from Lemma 11.1 and Proposition 10.1 that this function has the properties (13.2).

The first bound (13.4a) follows by using (11.28) and (11.29) in the definition (8.35) of u. Indeed, more precisely

$$\|g-m\|_{\mathscr{B}} \,\leq\, \|m\|_{\mathscr{B}}\|u\|_{\mathscr{B}} \,\leq\, \|m\|_{\mathscr{B}}\big(|\Theta|\|b\|_{\mathscr{B}} + \|r\|_{\mathscr{B}}\big) \,\lesssim\, |\Theta| + \|d\|_{\mathscr{B}}\,,$$

where $\|m\|_{\mathscr{B}} \sim 1$, $b = f + \mathcal{O}_{\mathscr{B}}(\alpha)$, $r = Rd + \mathcal{O}_{\mathscr{B}}(|\Theta|^2 + |d|^2)$, and $\|R\|_{\mathscr{B}\to\mathscr{B}}, \|f\|_{\mathscr{B}} \lesssim 1$, have been used.

In order to derive (13.4b) we first write:

$$\langle w, g-m\rangle \;=\; \langle |m|w, u\rangle \;=\; \langle |m|w, b\rangle\,\Theta + \langle |m|w, r\rangle\,. \tag{13.9}$$

Clearly, $|\langle |m|w, b\rangle| \lesssim \|w\|_{\mathscr{B}}$. Moreover, using (11.28) we obtain

$$\begin{aligned}\langle |m|w, r\rangle \;&=\; \Big\langle |m|w, Rd + \mathcal{O}_{\mathscr{B}}(\,|\Theta|^2 + \|d\|_{\mathscr{B}}^2)\Big\rangle \\ &=\; \langle R^*(|m|w), d\rangle \;+\; \mathcal{O}\Big(\|m\|_{\mathscr{B}}\|w\|_{\mathscr{B}}\big(|\Theta|^2 + \|d\|_{\mathscr{B}}^2\big)\Big)\,.\end{aligned}$$

Plugging this into (13.9), and setting $T := R^*(|m|\,\bullet\,)$, we recognise (13.4b). The bound (11.29) yields $\|T\|_{\mathscr{B}\to\mathscr{B}} \lesssim 1$.

As a next step we show that (13.5), (13.6a) and (13.6b) constitute just a simplified version of the cubic equation presented in Proposition 11.2. Combining (11.30) and (11.31) we see that

$$\big|\widetilde{\mu}_3\Theta^3 + \mu_2\Theta^2 + \mu_1\Theta\,\big| \;\lesssim\; |\langle |m|\overline{b}, d\rangle| + \|d\|_{\mathscr{B}}^2 + |\langle e, d\rangle|\,, \tag{13.10}$$

where the $\mathcal{O}(|\Theta|^4)$-sized part of the error $\kappa(u,d)$ in (11.30) has been included into $\widetilde{\mu}_3 = \mu_3 + \mathcal{O}(|\Theta|)$. Moreover, we have bounded the $\mathcal{O}(\,|\Theta|\,|\langle e, d\rangle|\,)$-sized part of κ by a larger $\mathcal{O}(|\langle e, d\rangle|)$ term. Recall that $|\Theta| \lesssim \varepsilon_*$ from (13.1). Hence taking $\varepsilon_* \sim 1$ small enough, the stability of the cubic (cf. (11.33)) implies that there is $c_0 \sim 1$ so that $|\mu_2| + |\widetilde{\mu}_3| = |\mu_2| + |\mu_3| + \mathcal{O}(|\Theta|) \geq 2c_0$ applies. Hence the coefficients

$$\begin{aligned} \pi_2 &:= \big(\mu_2 + (\widetilde{\mu}_3 - 1)\Theta\big)\, \mathbb{1}\big\{|\mu_2| \geq c_0\big\} + \frac{\mu_2}{\widetilde{\mu}_3}\, \mathbb{1}\big\{|\mu_2| < c_0\big\} \\ \pi_1 &:= \mu_1 \mathbb{1}\big\{|\mu_2| \geq c_0\big\} + \frac{\mu_1}{\widetilde{\mu}_3}\, \mathbb{1}\big\{|\mu_2| < c_0\big\}\,, \end{aligned} \tag{13.11}$$

scale just like μ_2 and μ_1 in size, i.e., $|\pi_2| \sim |\mu_2|$ and $|\pi_1| \sim |\mu_1|$, provided ε_* and thus $|\Theta|$ is sufficiently small. Moreover, by construction the bound (13.10) is equivalent to (13.5) once we set $t^{(1)} := |m|\bar{b}$ and $t^{(2)} := e$.

Let us first derive the scaling relation (13.6a) for π_1. Using $\sigma \in \mathbb{R}$, we obtain from (11.32c):

$$\begin{aligned} |\pi_1| \,\sim\, |\mu_1| \,&=\, \Big|-\langle f|m|\rangle\frac{\eta}{\alpha} \,+\, \mathrm{i}2\sigma\alpha \,-\, 2(\psi - \sigma^2)\,\alpha^2 \,+\, \mathcal{O}\big(\alpha^3 + \eta\big)\Big| \\ &\sim\, \Big|\frac{\langle f|m|\rangle}{2}\frac{\eta}{\alpha} + (\psi - \sigma^2)\,\alpha^2 \,+\, \mathcal{O}\big(\alpha^3 + \eta\big)\Big| \,+\, \Big|\sigma\alpha \,+\, \mathcal{O}\big(\alpha^3 + \eta\big)\Big|\,. \end{aligned} \tag{13.12}$$

The last comparison follows by expressing the size of μ_1 as the sum of sizes of its real and imaginary parts. We will now use the stability of the cubic, $\psi + \sigma^2 \gtrsim 1$ (cf. (11.33)). We treat two regimes separately.

First let us assume that $2\sigma^2 \leq \psi$. In that case $\psi \sim 1$, and we find

$$|\pi_1| \,\sim\, \frac{\eta}{\alpha} + \alpha^2 + |\sigma|\alpha + \mathcal{O}\big(\alpha^3 + \eta\big) \,\sim\, \frac{\eta}{\alpha} + \alpha^2 + |\sigma|\alpha\,. \tag{13.13}$$

In order to get the first comparison relation we have used the fact that $\psi - \sigma^2 \sim \psi \sim 1$ and $\langle f|m|\rangle \sim 1$ and hence the first two terms on the right hand side of the last line in (13.12) can not cancel each other. The second comparison in (13.13) holds provided $\varepsilon_* \sim 1$ is sufficiently small, recalling $\alpha \sim \langle v\rangle \leq \varepsilon_*$ (cf. (11.1), so that the error can be absorbed into the term $\eta/\alpha + \alpha^2$.

Now we treat the situation when $2\sigma^2 > \psi$. In this case $|\sigma| \sim 1$, and thus for small enough ε_*, we have

$$|\pi_1| \,\sim\, \Big|\frac{\eta}{\alpha} + \mathcal{O}(\alpha^2 + \eta)\Big| + \alpha \,=\, \frac{\eta}{\alpha} + \alpha + \mathcal{O}(\alpha^2 + \eta) \,\sim\, \frac{\eta}{\alpha} + \alpha \,\sim\, \frac{\eta}{\alpha} + |\sigma|\,\alpha + \alpha^2\,. \tag{13.14}$$

Here the first two terms in the last line of (13.12) may cancel each other but in that case both of the terms are $\mathcal{O}(\alpha^2)$ and hence the size of $|\pi_1|$ is given by the term $|\sigma|\alpha \sim \alpha$.

The scaling behaviour (13.6b) of π_2 follows from (11.32b) using $\|F\|_{\mathrm{L}^2\to\mathrm{L}^2} = 1 - \langle f|m|\rangle\,\eta/\alpha \sim 1$ (cf. (8.18) and (8.3)) and the stability of the cubic,

$$|\pi_2| \,\sim\, |\mu_2| \,\sim\, |\sigma| + |3\psi - \sigma^2|\alpha \,\sim\, |\sigma| + \alpha\,. \tag{13.15}$$

The formula (13.6a) now follows from (13.14) and (13.15) by using $\alpha \sim \langle \operatorname{Im} m\rangle$ and identifying

$$\widehat{\sigma}(z) \,:=\, |\sigma(z)|\,. \tag{13.16}$$

Since $\sigma : \overline{\mathbb{H}} \to [0,\infty)$ is 1/3-Hölder continuous, so is $\widehat{\sigma}$.

In order to obtain the relation (13.7a) we use (12.94) and Lemma 12.17 to get

$$\widehat{\sigma}(\tau_0) \,\sim\, \widehat{\Delta}(\tau_0)^{1/3} \,\sim\, \Delta(\tau_0)^{1/3},$$

for $\tau_0 \in \partial \operatorname{supp} v$ such that $|\sigma(\tau_0)| \leq \sigma_*$. On the other hand, if $|\sigma(\tau_0)| \geq \sigma_*$, i.e., $\widehat{\sigma}(\tau_0) \sim 1$, then also $\Delta(\tau_0) \sim 1$. This proves (13.7a).

In order to obtain (13.7b) we consider the cases $v(\gamma_k) = 0$ and $v(\gamma_k) > 0$ separately. If $v(\gamma_k) = 0$ then Lemma 12.12 shows that $\sigma(\gamma_k) = 0$. If $v(\gamma_k) > 0$ then $\partial_\tau \langle v(\gamma)\rangle|_{\tau=\gamma_k} = 0$. Lemma 12.2 thus yields $|\sigma(\gamma_k)| \leq \Pi_* \langle v(\gamma_k)\rangle^2$. Since $\Pi_* \sim 1$ this finishes the proof of (13.7b). □

Combining our two results concerning general perturbations, Lemma 8.10 and Proposition 13.1, with scaling behaviour of $m(z)$ as described by Theorem 6.4, we now prove Theorem 6.10 as well.

Proof of Theorem 6.10. Recall the definition (8.37) of the operator B. We will show below that

$$\|B(z)^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim \frac{1}{\varrho(z)^2+\varpi(z)^{2/3}}\,, \qquad |z|\le 4\,, \tag{13.17}$$

where $\varrho = \varrho(z)$ and $\varpi = \varpi(z)$ are defined in (6.31). Given (13.17) the statement (i) of the theorem follows by applying Lemma 8.10 with Φ introduced in the theorem and $\Psi := (\varrho+\varpi^{1/3})^{-2} \lesssim \varepsilon^{-2}$, where the constant $\varepsilon\in(0,1)$ is from (6.28). If $\varrho \ge \varepsilon_*$ or $\varpi \ge \varepsilon_*$ for some $\varepsilon_* \sim 1$, then (ii) follows similarly from Lemma 8.10 with $\Psi \sim 1$. Therefore, in order to prove the part (ii) it suffices to assume that $\varrho, \varpi \le \varepsilon_*$ for some sufficiently small threshold $\varepsilon_* \sim 1$.

We will take ε_* so small that Proposition 13.1 is applicable, and thus the cubic equation (13.5) can be written in the form

$$\big|\Theta^3 + \pi_2\Theta^2 + \pi_1\Theta\big| \;\lesssim\; \delta\,, \tag{13.18}$$

with $\delta = \delta(z,d) \le \|d\|_{\mathscr{B}}$ given in (6.31c) of Theorem 6.10. Combining the definition (13.3) of Θ with the a priori bound (6.32) for the difference $g-m$, we obtain

$$|\Theta| \le \|s\|_{\mathscr{B}}\|g-m\|_{\mathscr{B}} \lesssim \lambda\,(\varpi^{2/3}+\rho)\,. \tag{13.19}$$

For the last form we have used also (13.2). We will now show that if (13.19) holds for sufficiently small $\lambda\sim 1$, then the linear term of the cubic (13.18) dominates in the sense that

$$|\pi_1| \;\ge\; 3|\pi_2||\Theta|\,, \qquad \text{and} \qquad |\pi_1| \;\ge\; 3|\Theta|^2\,. \tag{13.20}$$

Let us first establish (13.20) when $\tau = \operatorname{Re} z \in \operatorname{supp} v$. From (13.19) and (13.6) we get

$$|\Theta| \lesssim \lambda\,(\varrho+\eta^{2/3}) \tag{13.21}$$
$$|\pi_1| \gtrsim (\widehat{\sigma}+\alpha)\alpha \tag{13.22}$$
$$|\pi_2| \sim \widehat{\sigma}+\alpha\,. \tag{13.23}$$

Here we have used the general property $v_x \sim \langle v\rangle \sim \alpha$ that always holds when $|\!|\!|m|\!|\!|_{\mathbb{R}} \lesssim \Phi$. Since $\tau\in\operatorname{supp} v$ we have $\varpi = \eta$ in (13.21). Let us show that

$$\varrho+\eta^{2/3} \lesssim \alpha\,. \tag{13.24}$$

To this end, let $\tau_0 = \tau_0(z) \in \mathbb{M}_{\varepsilon_*}$ be such that

$$|\tau-\tau_0| \;=\; \operatorname{dist}(\tau,\mathbb{M}_{\varepsilon_*}) \tag{13.25}$$

holds. If $\tau_0 \notin \partial\operatorname{supp} v$, then (d) of Corollary B.1 yields (13.24) immediately (take $\omega := \tau-\tau_0$ in the corollary). If on the other hand $\tau_0\in\partial\operatorname{supp} v$, then (a) of Corollary B.1 yields

$$\varrho+\eta^{2/3} \lesssim \frac{\omega^{1/2}}{(\Delta+\omega)^{1/6}} + \eta^{2/3} \lesssim \frac{(\omega+\eta)^{1/2}}{(\Delta+\omega+\eta)^{1/6}} \sim \alpha\,,$$

where $\Delta = \Delta(\tau_0)$ is the gap length (12.58) associated to the point $\tau_0\in\partial\operatorname{supp} v$ satisfying (13.25).

Combining (13.24) and (13.21) we get

$$|\Theta| \lesssim \lambda\,\alpha\,.$$

Using this bound together with (13.22) and (13.23) we obtain (13.20) for sufficiently small $\lambda \sim 1$.

Next we prove (13.20) when $\tau \notin \operatorname{supp} v$, i.e., $\varrho = 0$. In this case (13.19) and (13.6) yield

$$\begin{aligned} |\Theta| &\lesssim \lambda\,\varpi^{2/3} && (13.26)\\ |\pi_1| &\gtrsim \eta/\alpha && (13.27)\\ |\pi_2| &\lesssim 1\,. && (13.28) \end{aligned}$$

By combining the parts (b) and (c) of Corollary B.1 we get

$$\alpha \sim \frac{\eta}{(\Delta+\eta)^{1/6}\varpi^{1/2}} \lesssim \eta\,\varpi^{-2/3}\,, \tag{13.29}$$

where $\Delta = \Delta(\tau_0)$ is the gap length (12.58) associated to the point $\tau_0 \in \partial \operatorname{supp} v$. For the last bound in (13.29) we used $\varpi \sim \omega + \eta \le \Delta + \eta$. Plugging (13.29) into (13.27) we get

$$|\pi_1| \gtrsim \varpi^{2/3}\,. \tag{13.30}$$

Using this together with (13.26) and (13.28) we obtain (13.20) also when $\tau \notin \operatorname{supp} v$.

The estimates (13.20) imply

$$|\Theta|^3 \lesssim \bigl|\pi_1\Theta\bigr| \sim \bigl|\Theta^3 + \pi_2\,\Theta^2 + \pi_1\,\Theta\bigr|\,.$$

Using (13.18) we hence get

$$|\Theta|^3 \lesssim \bigl|\pi_1\Theta\bigr| \lesssim \delta\,,$$

from which it follows that

$$|\Theta| \lesssim \min\biggl\{\frac{\delta}{|\pi_1|}\,,\,\delta^{1/3}\biggr\}\,. \tag{13.31}$$

If $\tau \notin \operatorname{supp} v$ we have $\varrho = 0$ and thus (13.30) can be written as

$$|\pi_1| \gtrsim \varrho^2 + \varpi^{2/3}\,. \tag{13.32}$$

This estimate holds also when $\tau \in \operatorname{supp} v$. If the point $\tau_0 = \tau_0(E) \in \mathbb{M}_{\varepsilon_*}$ satisfying (13.25) is not an edge of $\operatorname{supp} v$, then (13.32) follows immediately from (d) of Corollary B.1 and from $|\pi_1| \gtrsim \alpha^2$ from (13.22). In order to get (13.32) when $\tau \in \operatorname{supp} v$ and $\tau_0 \in \partial \operatorname{supp} v$ we set $\omega = |\tau - \tau_0|$ and consider the cases $\omega + \eta > c_0\,\Delta$ and $\omega + \eta \le c_0\,\Delta$ for some small $c_0 \sim 1$ separately. If $\omega + \eta > c_0\,\Delta$ we get

$$\alpha^2 \sim (\omega+\eta)^{2/3} \sim \omega^{2/3} + \eta^{2/3} \sim \varrho^2 + \eta^{2/3}\,, \tag{13.33}$$

using (a) of Corollary B.1 in both the first and the last estimate. On the other hand, if $\omega + \eta \le c_0\,\Delta$ for sufficiently small $c_0 \sim 1$, then

$$\widehat{\sigma} = \widehat{\sigma}(z) \ge \widehat{\sigma}(\tau_0) - C|\tau_0 - z|^{1/3} \gtrsim \Delta^{1/3} - C(\omega+\eta)^{1/3} \ge \frac{1}{2}\Delta^{1/3}\,, \tag{13.34}$$

where we have used 1/3-Hölder continuity of $\widehat{\sigma}$ and the relation (13.7a) from Proposition 13.1. For the last bound we have used $|\tau_0 - z| \sim \omega + \eta$ as well. Therefore, we have

$$\widehat{\sigma}\alpha \sim \Delta^{1/6}(\omega+\eta)^{1/2} \gtrsim \omega^{2/3} + \eta^{2/3} \gtrsim \varrho^2 + \eta^{2/3}\,. \tag{13.35}$$

Here we have used (a) of Corollary B.1 twice. Combining (13.33) and (13.35) we get

$$\widehat{\sigma}\alpha + \alpha^2 \gtrsim \varrho^2 + \varpi^{2/3}, \qquad \tau \in \operatorname{supp} v. \tag{13.36}$$

Using this in (13.22) yields (13.32) when $\tau_0 \in \partial \operatorname{supp} v$.

By combining (13.31) and (13.32) we get

$$|\Theta| \lesssim \Upsilon, \tag{13.37}$$

with $\Upsilon = \Upsilon(z, d)$ defined in (6.34). The estimates (6.33) now follow from (13.4) using (13.37).

We still need to prove (13.17). If $\tau \in \operatorname{supp} v$ we know from (8.40) of Lemma 8.8 that

$$\|B^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim (\widehat{\sigma}\alpha + \alpha^2)^{-1}.$$

Using (13.36) we get (13.17) when $\tau \in \operatorname{supp} v$. In the remaining case $\tau \notin \operatorname{supp} v$ (13.17) reduces to

$$\|B^{-1}\|_{\mathscr{B}\to\mathscr{B}} \lesssim \varpi^{-2/3}. \tag{13.38}$$

In order to prove this we use (8.40) to get the first bound below:

$$\|B^{-1}\|_{\mathscr{B}\to\mathscr{B}} \le 1 + \|B^{-1}\|_{\mathrm{L}^2\to\mathrm{L}^2} \le 1 + \frac{1}{1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2}} \lesssim 1 + \frac{\alpha}{\eta}. \tag{13.39}$$

For the second estimate we have used the definition (8.37) of B and the identity (8.18). Finally, for the third inequality we used $\langle f\,|m|\rangle \sim 1$ to estimate $1 - \|F\|_{\mathrm{L}^2\to\mathrm{L}^2} \gtrsim \eta/\alpha$. Using (13.29) in (13.39) yields (13.38). This completes the proof of (13.17). $\square$

14 Examples

In this section we will present some examples that illustrate why the assumptions, made on S, are needed for our analysis. Moreover, we give specific choices of S, which lead to the different shapes of the generating density, described by our main result, Theorem 6.4. The assumptions **A1.** and **A2.** are structural in nature, while **A3** is simply a normalisation, which we will drop for the upcoming examples. The smoothing assumption **A4.** was made for technical reasons. We used it to deduce uniform bounds from spectral information. Here we focus on the remaining assumptions **A5.**, **B1.** and **B2.** Recall that the later two conditions are defined in the beginning of Section 9, and that they were introduced to prove uniform bounds for the solution m away and around $z = 0$, respectively (cf. Theorem 9.1). The key character of **B1-2.** is that these uniform bounds depend only on the parameters appearing in their definitions. The simpler result Theorem 6.8 is a qualitative version of Theorem 9.1 in the sense that the uniform bounds can not be expressed only in terms of the model parameters.

Most of the examples here are represented in the special setting where $\mathfrak{X}$ in the unit interval and π is the restriction of the Lebesgue measure on this interval w.r.t. standard Borel σ-algebra:

$$(\mathfrak{X}, \mathcal{B}, \pi) := \big([0,1], \mathcal{B}([0,1]), \mathrm{d}x\big). \tag{14.1}$$

14.1 The band operator and lack of self-averaging

The uniform primitivity assumption, **A5.**, was made to exclude choices of S that lead to an essentially decoupled system. Without sufficient coupling of the components m_x in the QVE the components of the imaginary part of the solution are not necessarily comparable in size, i.e., $v_x \sim v_y$, may not hold. No universal growth behaviour at the edge of the support of the generating density, as described by Theorem 6.4, can be expected in this case, since the support of v_x may not even be independent of x.

The simplest such situation is if there exists a subset $I \subsetneq [0,1]$ such that S leaves the functions invariant that are supported on I and also the ones that are supported on the complement of I. In this case the QVE will decouple and we may apply the developed theory to each of the resulting equations independently. Assumption **A5.** also excludes a situation, where the functions supported on I are mapped to the function supported on the complement of I, and vice versa. This case has an instability at the origin $\tau = 0$ and requires a special treatment of the lowest lying eigenvalue of S.

Another example, illustrating why **A5.** is needed, is the following S-operator with a small band along the diagonal:

$$S_{xy} \,=\, \varepsilon^{-1}\xi(x+y)\,\mathbb{1}\big\{|x-y| \le \varepsilon\big\}\,.$$

Here, $\xi : \mathbb{R} \to (0,\infty)$ is some smooth function and ε a small positive constant. For any fixed ε this operator satisfies all our assumptions **A1.**, **A2**, **A4-5.** and **B1-2.**. As ε approaches zero, however, the constant L from assumption **A5.**, as well as $\|S\|_{\mathrm{L}^2\to\mathscr{B}}$ from assumption **A4.** diverge. In the limit, S becomes a multiplication operator and QVE decouples completely,

$$-\frac{1}{m_x(z)} \,=\, z + \xi(x)\,m_x(z)\,.$$

The solution becomes trivial. Each component m_x is the Stieltjes transform of Wigner's semicircle law (1.4), scaled by the corresponding value $\xi(x)$. In particular, the support of the components v_x of the generating density depend on x.

14.2 Divergencies for special x-values: Outlier rows

Theorem 6.8 shows that away from $z = 0$ the solution of the QVE stays bounded if, in addition to **A1-5.**, assumption **B1.** is satisfied. We present two examples in which the condition **B1.** is violated. In both cases a few exceptional row functions, $S_x(y) = S_{xy}$, cause divergencies in the corresponding components, m_x, of the solution. The first example is so simple that the QVE can be solved explicitly and thus the divergence can be read off from the solution formula. The second example is a bit more involved. It illustrates how divergencies may arise from smoothing out discontinuities in the kernel of S on small scales.

We start with the simple 2×2 - block operator:

$$S^{(0)}_{xy} \,=\, \lambda\,\mathbb{1}\{\,x \le \delta,\, y > \delta\,\} + \lambda\,\mathbb{1}\{\,y \le \delta,\, x > \delta\,\} + \mathbb{1}\{\,x > \delta,\, y > \delta\,\}\,, \tag{14.2}$$

with two positive parameters λ and δ. For any fixed values of $\lambda > 0$ and $\delta \in (0, 1/2)$ this operator satisfies **A1.**, **A2**, **A4-5.** and **B1-2.**. In fact, the solution has the structure

$$m_x(z) \,=\, \mu(z)\,\mathbb{1}\{\,x \le \delta\,\} + \nu(z)\,\mathbb{1}\{\,x > \delta\,\}\,, \tag{14.3}$$

where the two functions $\mu, \nu : \overline{\mathbb{H}} \to \overline{\mathbb{H}}$ satisfy the coupled equations

$$-\frac{1}{\mu(z)} \,=\, z + (1-\delta)\,\lambda\,\nu(z)\,, \qquad -\frac{1}{\nu(z)} \,=\, z + \lambda\delta\,\mu(z) + (1-\delta)\,\nu(z)\,. \tag{14.4}$$

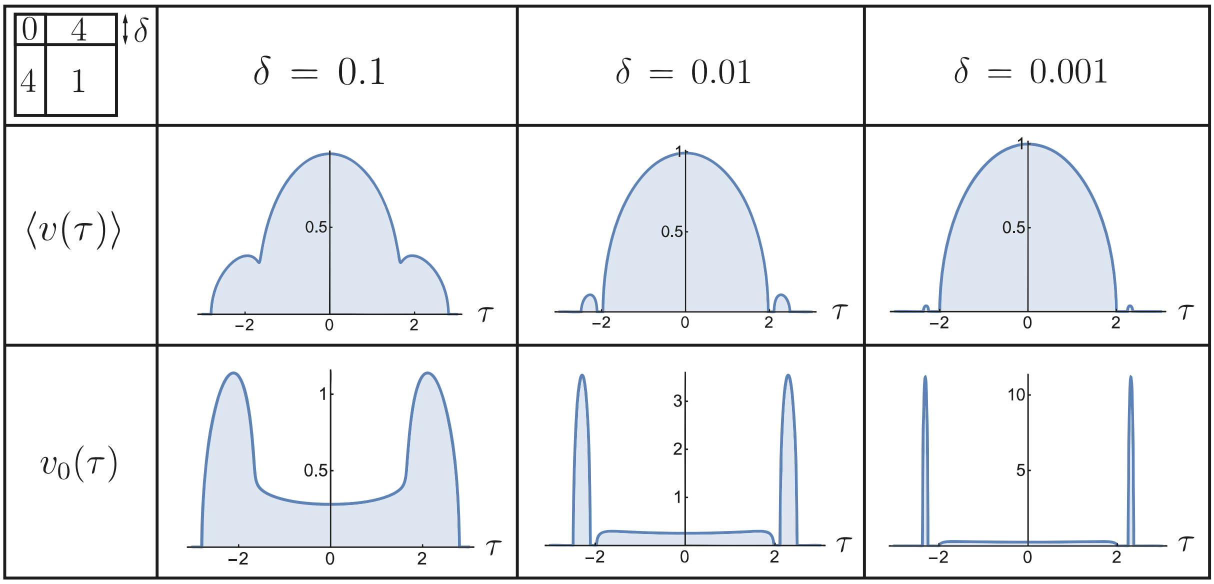

Figure 14.1: As δ decreases the average generating density remains bounded, but the 0-th component of the generating density blows up at $\pm\tau_0$.

Let us consider a fixed $\lambda > 2$. Then, as we take the limit $\delta \downarrow 0$, the condition **B1** becomes ineffective, while all the other assumptions still hold uniformly in δ. Indeed, the row functions $(S_x^{(0)})_{x\in[0,\delta]}$ differ from the row functions $(S_x^{(0)})_{x\in(\delta,1]}$. This leads to a blow-up in the components $m_x(z)$ with $x \in [0, \delta]$ at a specific value of z. More precisely, we find $|\mu(\pm\tau_0)| \sim \delta^{-1/2}$ with

$$\tau_0 \;:=\; \frac{2\,\lambda}{\sqrt{\lambda^2 - (\lambda-2)^2}}\,.$$

While the $\mathscr{B}$-norm of m diverges as δ approaches zero, the $\mathrm{L}^2[0,1]$-norm stays finite, because the divergent components contribute less and less. The situation is illustrated in Figure 14.1.

The operator-kernel $S^{(0)}$ even makes sense for $\delta = 0$. In this case we get for the generating measure the formulas,

$$\begin{aligned} v_0(\mathrm{d}\tau) \;&=\; \frac{\lambda\sqrt{4-\tau^2}}{2\,\lambda^2 - 2\,\tau^2(\lambda-1)}\,\mathbb{1}\{\,\tau\in[-2,2]\,\}\,\mathrm{d}\tau \;+\; \frac{\pi(\lambda-2)}{2(\lambda-1)}\big(\delta_{-\tau_0}(\mathrm{d}\tau)+\delta_{\tau_0}(\mathrm{d}\tau)\big)\,,\\ v_x(\mathrm{d}\tau) \;&=\; \frac{1}{2}\sqrt{4-\tau^2}\,\mathbb{1}\{\,\tau\in[-2,2]\,\}\,\mathrm{d}\tau\,, \qquad x\in(0,1]\,. \end{aligned}$$

The non-zero value that v_0 assigns to τ_0 and $-\tau_0$ reflects the divergence of m in the uniform norm at these points.

In the context of random matrix theory the operator $S^{(0)}$ with small values of the parameter δ corresponds to the variance matrix of a perturbation of a Wigner matrix. The part of the generating density, which is supported around τ_0 corresponds to a small collection of eigenvalues away from the bulk of the spectrum of the random matrix. These outliers will induce a divergence in some elements of the resolvent of this matrix. This divergence is what we see as the divergence of μ here.

We present a second example of a different nature, which also violates assumption **B1.**. The smoothing of discontinuities in S may cause blow-ups in the solution of the QVE (cf. Figure 14.2). This is somewhat surprising, since by conventional wisdom, smoother data implies smoother solutions. The key point here is that the smoothing procedure creates a few row functions that are far away from all the other row functions. The following choice of operator demonstrates this mechanism:

$$S^{(\varepsilon)}_{xy} \;=\; \frac{1}{2}(r_x s_y + r_y s_x)\,.$$

Here the two continuous functions $r, s : [0,1] \to (0,1]$ are given by

$$\begin{aligned} r_x &= (1 + \varepsilon^{-1}(x-\delta))\,\mathbb{1}\{\,\delta - \varepsilon < x \le \delta\,\} + \mathbb{1}\{\,x > \delta\,\}\,, \\ s_x &= 2\lambda\,\mathbb{1}\{\,x \le \delta\,\} + (2\lambda - \varepsilon^{-1}(2\lambda - 1)(x-\delta))\,\mathbb{1}\{\,\delta < x \le \delta + \varepsilon\,\} + \mathbb{1}\{\,x > \delta + \varepsilon\,\}\,, \end{aligned}$$

and $\lambda > 0$, $\delta \in (0,1)$ are considered fixed positive model parameters, while $\varepsilon \in (0,\delta)$ is varied. The continuous kernel $S^{(\varepsilon)}$ represents a smoothed out version of the 2×2-block operator $S^{(0)}$ from (14.2). Assumption **B1.** is satisfied for the operator $S^{(\varepsilon)}$ for any fixed $\varepsilon > 0$ and for the limiting operator $S^{(0)}$ as well. Nevertheless, as ε approaches zero, condition **B1.** becomes ineffective, since the constants in this assumption depend on ε. This is due to the distance that some row functions in $(S^{(\varepsilon)}_x)_{x \in [\delta-\varepsilon, \delta+\varepsilon]}$ have from all the other row functions.

Let $m = m^{(\varepsilon)}$ denote the solution of the QVE corresponding to $S^{(\varepsilon)}$. We will now show that, even though $m^{(0)}$ is uniformly bounded, the $\mathscr{B}$-norm of $m^{(\varepsilon)}$ diverges as ε approaches zero for certain parameters λ and δ.

The solution $m = m^{(\varepsilon)}$ has the form

$$m_x(z) \;=\; -\,\frac{1}{z + \varphi(z)\,r_x + \psi(z)\,s_x}\,.$$

Here, the two functions $\varphi^{(\varepsilon)} = \varphi = \langle s, m\rangle, \psi^{(\varepsilon)} = \psi = \langle r, m\rangle : \mathbb{H} \to \mathbb{H}$ satisfy the coupled equations

$$\varphi(z) \;=\; -\int_{\mathfrak{X}} \frac{s_x\,\pi(\mathrm{d}x)}{z + \varphi(z)\,r_x + \psi(z)\,s_x}\,, \qquad \psi(z) \;=\; -\int_{\mathfrak{X}} \frac{r_x\,\mathrm{d}x}{z + \varphi(z)\,r_x + \psi(z)\,s_x}\,. \tag{14.5}$$

In the parameter regime $\lambda \ge 10$ and $\delta \le 1/10$ the support of the generating density of $m^{(0)}$ consists of three disjoint intervals,

$$\operatorname{supp} v^{(0)} \;=\; \operatorname{supp} \varphi^{(0)} \;=\; \operatorname{supp} \psi^{(0)} \;=\; [-\beta_1, -\alpha_1] \cup [-\alpha_0, \alpha_0] \cup [\alpha_1, \beta_1]\,.$$

Inside the gap (α_0, α_1) the norm $\|m^{(\varepsilon)}\|_{\mathscr{B}}$ diverges as $\varepsilon \downarrow 0$. This can be seen indirectly, by utilising Theorem 6.4. We will now sketch an argument, which shows that assuming a uniform bound on m leads to a contradiction. Suppose there were an ε-independent bound on the uniform norm. Then a local version of Theorem 6.4 would be applicable and the generating density $v^{(\varepsilon)}$ of $m^{(\varepsilon)}$ could approach zero only in the specific ways described in that theorem. Instead, the average generating density $\langle v^{(\varepsilon)}\rangle$ takes small non-zero values along the whole interval (α_0, α_1), as we explain below. This contradicts the assertion of the theorem.

In fact, a stability analysis of the two equations for $\varphi^{(\varepsilon)}$ and $\psi^{(\varepsilon)}$ shows that they are uniformly Lipshitz-continuous in ε. In particular, for τ well inside the interval (α_0, α_1) we have

$$\operatorname{Im} \varphi^{(\varepsilon)}(\tau) + \operatorname{Im} \psi^{(\varepsilon)}(\tau) \;\le\; C\,\varepsilon\,.$$

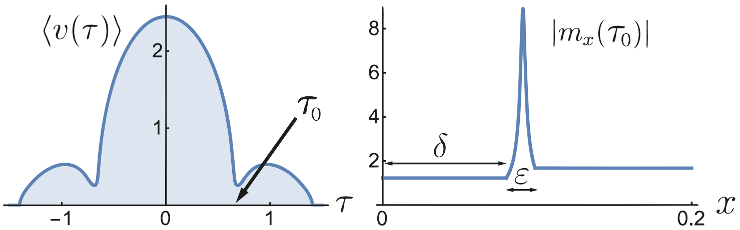

Figure 14.2: As ε decreases the average generating density remains bounded. The absolute value of the solution as a function of x at a fixed value E_0 inside the gap of the limiting generating density has a blow up.

Thus, the average generating density takes small values here as well, $\langle v^{(\varepsilon)}(\tau)\rangle \le C\varepsilon$. On the other hand, $\operatorname{Im}\varphi$ and $\operatorname{Im}\psi$ do not vanish on (α_0, α_1). Their supports coincide with the support of the generating density, $v^{(\varepsilon)}$. By Theorem 6.9 this support is a single interval for all $\varepsilon > 0$ and by the continuity of φ and ψ in ε, every point

$\tau \in (-\beta_1, -\alpha_1) \cup (\alpha_1, \beta_1)$ is contained in this interval in the limit $\varepsilon \downarrow 0$.

This example demonstrates that certain features of the solution of the QVE cannot be expected to be stable under smoothing of the corresponding operator S. Among these features are gaps in the support of the generating density, as well as the universal shapes described by Theorem 6.4.

14.3 Discretisation of the QVE

By choosing $\mathfrak{X} := \{1, \ldots, N\}$ and $\pi(i) := N^{-1}$ for some $N \in \mathbb{N}$ the QVE (6.5) takes the form

$$-\frac{1}{m_i} \;=\; z + \frac{1}{N}\sum_{j=1}^{N} S_{ij}\, m_j\,, \qquad i = 1, \ldots, N\,, \tag{14.6}$$

and hence this matrix equation is covered by our analysis. Alternatively, we may treat the discrete equation (14.6) in the setting $\mathfrak{X} = [0,1]$ and $\pi(\mathrm{d}x) = \mathrm{d}x$. Namely, we interpret the matrix $\mathbf{S} = (S_{ij})$ as $N \times N$-block operator S with square blocks of equal size that takes the constant value S_{ij} on the (i,j)-th block,

$$S_{xy} \;:=\; S_{ij}\, \mathbb{1}\big\{Nx \in [i-1, i)\,,\; Ny \in [i-1, i)\big\}\,. \tag{14.7}$$

The value m_i is the value of the corresponding continuum solution, $(m_x)_{x \in [0,1]}$, on the i-th block, i.e., $m_i = m_x$ for $Nx \in [i-1, i)$. This follows from the uniqueness of the solution m and the fact that both sides of the QVE conserve the block structure.

This translation of discrete QVEs into the continuos setting $(\mathfrak{X}, \pi(\mathrm{d}x)) = ([0,1], \mathrm{d}x)$ is convenient when comparing different discrete QVEs of non-matching dimensions N. For example, the convergence of a sequence of QVEs generated by a smooth symmetric function $f : [0,1]^2 \to [0, \infty)$ through the discretisation

$$S_{ij} \;:=\; f\Big(\frac{i}{N}, \frac{j}{N}\Big)\,, \qquad i, j = 1, \ldots, N\,,$$

can be handled this way. Indeed, with $(m_1, \ldots, m_N)$ the solution of the discrete problem, the step function m with value m_i on the i-th out of N equally sized blocks will converge to the solution of the continuous QVE with $S_{xy} := f(x,y)$ as $N \to \infty$. If the continuum operator satisfies **A5.** and **B2.** (all other assumptions are automatic in this case), then the convergence of the generating densities is uniform and the support of the generating density is a single interval for large enough N. This is a consequence of the stability result, Theorem 6.10, more precisely of Remark 6.11 following it and of the fact that $(S_{ij})_{i,j=1}^{N}$ is block fully indecomposable, and the knowledge about the shape of the generating density from Theorem 6.4 and Theorem 6.9.

14.4 The DAD-problem and divergencies at $z = 0$

Assumption **B2.** is designed to prevent divergencies in the solution at the origin of the complex plane. These divergencies are caused by the structure of small values of the kernel S_{xy}. At $z = 0$ the QVE reduces to

$$v_x \int_{\mathfrak{X}} \pi(\mathrm{d}y)\, S_{xy} v_y \;=\; 1\,, \qquad x \in \mathfrak{X}\,. \tag{14.8}$$

This equation for $v_x = \operatorname{Im} m_x(0)$ is known as the DAD- or scaling-problem and there is an extensive literature on its solvability that dates back at least to [57], mostly in the discrete case, i.e., when S is a matrix with positive entries. If equation (14.8) does not have a bounded solution, $\|m(z)\|_{\mathscr{B}}$ will diverge as z approaches zero.

To formulate simple clear cut statements about the solvability of the DAD-problem and its connection to the boundedness of the solution of the QVE at the origin of the complex plane, we consider its discrete analog,

$$v_i \,\frac{1}{N}\sum_{j=1}^{N} S_{ij} v_j \;=\; 1\,, \qquad i = 1,\dots,N\,, \tag{14.9}$$

for a solution vector, $(v_1,\dots,v_N)$, with positive entries. This equation is equivalent to the discrete QVE, (14.6), at $z = 0$. If the DAD-problem (14.9) is not solvable, the solution of the discrete QVE has a divergence at the origin of the complex plane. Therefore, divergencies in the solution of the QVE at $z = 0$ can be understood in terms of solvability of the corresponding DAD-problem. It is a well-known fact [13] that the DAD-problem has a unique solution for a symmetric irreducible matrix $(S_{ij})_{i,j=1}^N$ with non-negative entries if and only if S is fully indecomposable. This fact is reflected in the assumption **B2.**.

Let us go back to the continuum setting. If assumption **B2.** is violated the generating measure may have a singularity at $z = 0$. In fact, there are two types of divergencies that may occur. Either the generating density exists in a neighborhood of $\tau = 0$ and has a singularity at the origin, or the generating measure has delta-component at the origin. Both cases can be illustrated using the 2×2-block operator $S^{(0)}$ from (14.2).

The latter case occurs if the kernel S_{xy} contains a rectangular zero-block whose circumference is larger than 2. For $S^{(0)}$ this means that $\delta > 1/2$. Expanding the corresponding QVE for small values of z reveals

$$v_x(\mathrm{d}\tau) \;=\; \frac{\pi\,(2\,\delta-1)}{\delta}\,\mathbb{1}\{x\le\delta\}\,\delta_0(\mathrm{d}\tau) \;+\; \mathcal{O}(1)\mathrm{d}\tau\,.$$

The components of the generating measure with $x\in[0,\delta]$ assign a non-zero value to the origin.

The case of a singular, but existing generating density can be seen from the same example, (14.2), with the choice $\delta = 1/2$. From an expansion of QVE at small values of z we find for the generating density:

$$v_x(\tau) \;=\; (2\,\lambda)^{-2/3}\sqrt{3}\,|\tau|^{-1/3}\,\mathbb{1}\{x\le 1/2\} + \mathcal{O}(1)\,.$$

The blow-up at $z = 0$ has a simple interpretation in the context of random matrix theory. It corresponds to an accumulation of eigenvalues at zero. If the generating density assigns a non-zero value to the origin, a random matrix with the corresponding S as its variance matrix will have a kernel, whose dimension is a finite fraction of the size N of the matrix.

Assumption **B2.** excludes the above examples by ensuring that the DAD-problem has a unique solution and is sufficiently stable.

14.5 Shapes of the generating density

We will now discuss how all possible shapes of the generating density from Theorem 6.4 can be seen in the simple example of the 2×2-block operator $S^{(0)}$ from (14.2) by choosing the parameters λ and δ appropriately. For the choice of parameters $\lambda > 2$ and $\delta = \delta_c(\lambda)$ with

$$\delta_c(\lambda) \;:=\; \frac{(\lambda-2)^3}{2\,\lambda^3 - 3\,\lambda^2 + 15\,\lambda - 7}\,,$$

the generating density exists everywhere and its support is a single interval. In the interior of this interval the generating density has exactly two zeros at some values τ_c and $-\tau_c$. The shape of the generating density at these zeros in the interior of its own support is a cubic cusp, represented by the shape function $\Psi^{(0)}_{\mathrm{min}}$. If we increase δ above $\delta_c(\lambda)$, then the zeros of the generating density disappear. The support is a single interval with local minima close to τ_c and $-\tau_c$ and the shape around these minima is described by $\Psi^{(\rho)}_{\mathrm{min}}$ for some small positive ρ. Finally, if we decrease δ slightly below $\delta_c(\lambda)$ a gap opens in the support. Now the support of the generating density consists of three disjoint intervals and the shape of the generating density at the two neighbouring edges is described by Ψ_{edge}. The different choices of δ are illustrated in Figure 14.3.

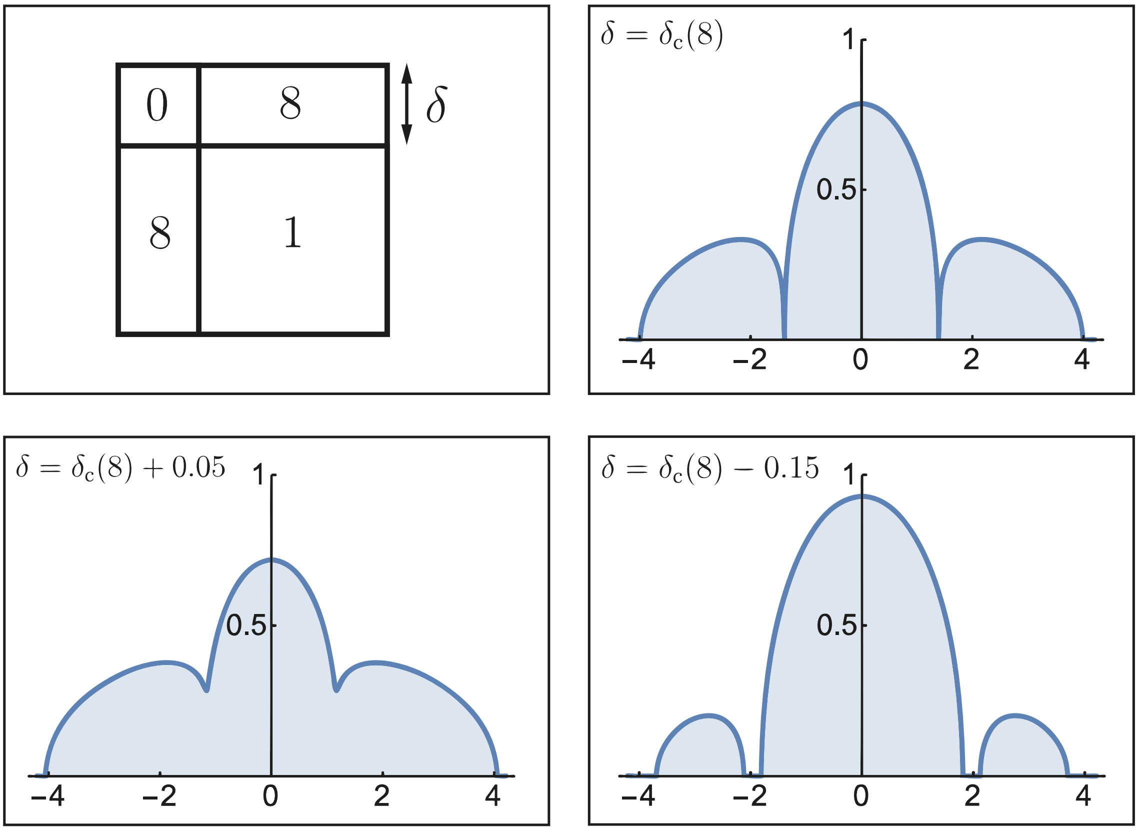

Figure 14.3: Decreasing δ from its critical value δ_c opens a gap in the support of the average generating density. Increasing delta lifts the cubic cusp singularity.

Part III

15 Introduction to Part III

In this part we will prove the local law and local spectral universality in the bulk for general random matrices, $\mathbf{H}$, with independent entries. We show that the diagonal elements of the resolvent, $\mathbf{G}(z) := (\mathbf{H} - z)^{-1}$, satisfy the approximate self-consistent equation

$$-\frac{1}{G_{ii}(z)} \approx z + \sum_{j=1}^{N} s_{ij} G_{jj}(z)\,, \qquad i = 1, \dots, N\,.$$

Here, $\mathbf{S} = (s_{ij})_{i,j=1}^{N}$ is the matrix with the variances of the matrix elements of $\mathbf{H}$ as its entries. The stability analysis for this equation, carried out in Part II of this work, is used heavily in the proof. Nevertheless, this part is completely self-contained and all results from Part II are repeated when they are needed.

15.1 Set-up and main results

Let $\mathbf{H}^{(N)} \in \mathbb{C}^{N \times N}$ be a sequence of self-adjoint random matrices. In particular, if the entries of $\mathbf{H}$ are real then $\mathbf{H}^{(N)}$ is symmetric. The matrix ensemble $\mathbf{H} = \mathbf{H}^{(N)}$ is said to be of **Wigner type** if its entries h_{ij} are independent for $i \le j$ and centered, i.e.,

$$\mathbb{E}\, h_{ij} \;=\; 0 \quad \text{ for all } \quad i, j \;=\; 1, \dots, N\,. \tag{15.1}$$

The dependence of $\mathbf{H}$ and other quantities on the dimension N will be suppressed in our notation. The matrix of variances, $\mathbf{S} = (s_{ij})_{i,j=1}^{N}$, is defined through

$$s_{ij} := \mathbb{E}\,|h_{ij}|^2\,. \tag{15.2}$$

It is symmetric with non-negative entries. In Part II it was shown that for every such matrix $\mathbf{S}$ the **quadratic vector equation** (QVE),

$$-\frac{1}{m_i(z)} \;=\; z + \sum_{j=1}^{N} s_{ij} m_j(z)\,, \quad \text{for all } \; i = 1, \dots, N \text{ and } z \in \mathbb{H}\,, \tag{15.3}$$

for a function $\mathbf{m} = (m_1, \dots, m_N) : \mathbb{H} \to \mathbb{H}^N$ on the complex upper half plane, $\mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$, has a unique solution. The main result of this paper is to establish the local law for Wigner-type matrices, i.e. that for large N the resolvent, $\mathbf{G}(z) = (\mathbf{H} - z)^{-1}$, with spectral parameter $z = \tau + \mathrm{i}\eta \in \mathbb{H}$, is close to the diagonal matrix, $\operatorname{diag}(\mathbf{m}(z))$, as long as $\eta \gg N^{-1}$. As a consequence, we obtain rigidity estimates on the eigenvalues and complete delocalisation for the eigenvectors. Combining this information with the Dyson Brownian motion, we obtain universality of the eigenvalue gap statistics in the bulk.

We now list the assumptions on the variance matrices $\mathbf{S} = \mathbf{S}^{(N)}$. The restrictions on $\mathbf{S}$ are controlled by three **model parameters**, $p, P > 0$ and $L \in \mathbb{N}$, which do not depend on N. These parameters will remain fixed throughout this paper. In the following we will always assume that $\mathbf{S}$ satisfies the following conditions:

(A) For all N the matrix $\mathbf{S}$ is **flat**, i.e.,

$$s_{ij} \;\le\; \frac{1}{N}\,, \qquad i, j = 1, \dots, N\,. \tag{15.4}$$

(B) For all N the matrix $\mathbf{S}$ is **uniformly primitive**, i.e.,

$$\left(\mathbf{S}^L\right)_{ij} \,\geq\, \frac{p}{N}\,, \qquad i,j = 1,\dots,N\,. \tag{15.5}$$

(C) For all N the matrix $\mathbf{S}$ **induces a bounded solution of the QVE**, i.e., the unique solution $\mathbf{m}$ of (15.3) corresponding to $\mathbf{S}$ is bounded,

$$|m_i(z)| \,\leq\, P\,, \qquad i = 1,\dots,N\,,\; z \in \mathbb{H}\,. \tag{15.6}$$

Remark 15.1. *The assumption on the boundedness of* $\mathbf{m}$ *is an implicit condition in the sense that it can be checked only after solving* (15.3)*. In Part II we list sufficient, explicitly checkable conditions on* $\mathbf{S}$*, which ensure* (15.6)*. We also remark that the assumption* (15.4) *can be replaced by* $s_{ij} \leq C/N$ *for some positive constant* C*. This will lead to a rescaling of* $\mathbf{m}$*. We pick the normalisation* $C = 1$ *just for convenience.*

In addition to the assumptions on the variances of h_{ij}, we also require uniform boundedness of higher moments. This leads to another basic **model parameter**, $\mu = (\mu_1, \mu_2, \dots)$, which is a sequence of non-negative real numbers.

(D) For all N the entries h_{ij} of the random matrix $\mathbf{H}$ have **bounded moments**,

$$\mathbb{E}\,|h_{ij}|^k \,\leq\, \mu_k\, s_{ij}^{k/2}\,, \qquad k \in \mathbb{N}\,,\; i,j = 1,\dots,N\,. \tag{15.7}$$

In order to state our main result, in the next Corollary we collect a few facts about the solution of the QVE that are proven in Part II. Although these properties are sufficient for the formulation of our results, for their proofs we will need much more precise information about the solution of the QVE. Theorems 18.1 and 18.2 summarise everything that is needed from Part II besides the existence and uniqueness of the solution of the QVE. In particular, the statement of Corollary 15.2 follows easily from Theorem 18.1.

Corollary 15.2 (Solution of QVE). *Assume* $\mathbf{S}$ *satisfies assumptions (A), (B) and (C). Let* $\mathbf{m} : \mathbb{H} \to \mathbb{H}^N$ *be the solution the QVE* (15.3) *corresponding to* $\mathbf{S}$*. Then* $\mathbf{m}$ *is analytic and has a continuous extension (denoted again by* $\mathbf{m}$*) to the closed upper half plane,* $\mathbf{m} : \overline{\mathbb{H}} \to \overline{\mathbb{H}}^N$*, with* $\overline{\mathbb{H}} := \mathbb{H} \cup \mathbb{R}$*. The function* $\rho : \mathbb{R} \to [0,\infty)$*, defined by*

$$\rho(\tau) \,:=\, \frac{1}{\pi N} \sum_{i=1}^{N} \operatorname{Im} m_i(\tau)\,, \tag{15.8}$$

is a probability density. Its support is contained in $[-2, 2]$ *and is a union of closed disjoint intervals*

$$\operatorname{supp} \rho \,=\, \bigcup_{k=1}^{K} [\alpha_k, \beta_k]\,, \qquad \textit{where} \qquad \alpha_k < \beta_k < \alpha_{k+1}\,. \tag{15.9}$$

There exists a positive constant δ_**, depending only on the model parameters* p*,* P *and* L*, such that the sizes of these intervals are bounded from below by*

$$\beta_k - \alpha_k \,\geq\, 2\,\delta_*\,. \tag{15.10}$$

Note that (15.10) provides a lower bound on the length of the intervals that constitute $\operatorname{supp}\rho$, while the length of the gaps, $\alpha_{k+1} - \beta_k$, between neighbouring intervals can be arbitrarily small. Figure 15.1 shows a shape that the density of states typically might have. In particular, ρ may have gaps in its support and may have additional zeros (cusps) in the interior of $\operatorname{supp}\rho$. For more details on the possible edge shapes see Theorem 6.4 of Part II.

Definition 15.3 (Density of states). *The function ρ defined in* (15.8) *is called the* **density of states**. *Its harmonic extension to the upper half plane*

$$\rho(\tau+\mathrm{i}\eta) := \int_{\mathbb{R}} \pi_\eta(\tau-\sigma)\,\rho(\sigma)\,\mathrm{d}\sigma\,, \qquad \pi_\eta(\tau) := \frac{1}{\pi}\frac{\eta}{\tau^2+\eta^2}\,; \qquad \tau\in\mathbb{R}\,,\ \eta>0\,, \tag{15.11}$$

is again denoted by ρ. With a slight abuse of notation we still write $\operatorname{supp}\rho$, *as in* (15.9), *for the support of the density of states as a function on the real line.*

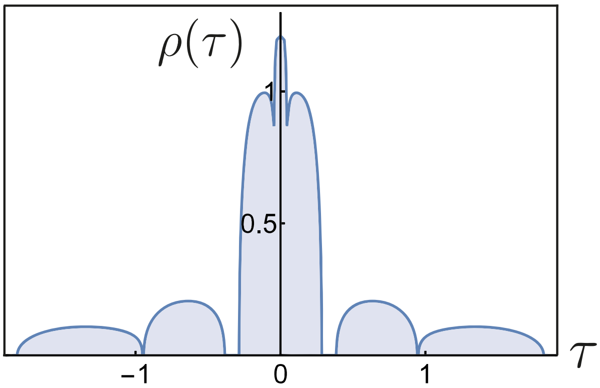

Figure 15.1: The density of states may have gaps, cusps and local minima.

The density of states will be shown to be the eigenvalue distribution of $\mathbf{H}$ in the large N limit on the macroscopic scale. For any fixed values $\tau_1, \tau_2 \in \mathbb{R}$ with $\tau_1 < \tau_2$ it satisfies

$$\lim_{N\to\infty} \frac{\left|\operatorname{Spec}(\mathbf{H}^{(N)})\cap[\tau_1,\tau_2]\right|}{N\int_{\tau_1}^{\tau_2}\rho^{(N)}(\tau)\,\mathrm{d}\tau} = 1\,, \tag{15.12}$$

provided the denominator does not vanish in the limit. The identity (15.12) motivates the terminology of density of states for the function ρ. The harmonic extension of ρ to $\mathbb{H}$ is a version of the density of states, that is smoothed out on the scale η. It satisfies the identity $\rho(z) = \frac{1}{\pi N}\sum_{i=1}^N \operatorname{Im} m_i(z)$ not just for $z\in\mathbb{R}$ (cf. (15.8)) but for all $z\in\overline{\mathbb{H}}$ and it will be used in the statement of our main result, Theorem 15.6.

In fact, Theorem 15.6, implies a local version of (15.12), where the length of the interval, $[\tau_1,\tau_2]$, may converge to zero as N tends to infinity. Our estimates and thus the proven speed of convergence depend on the distance of the interval to the edges of $\operatorname{supp}\rho$ and even on the length of the closest gap in this case. We introduce a function $\Delta:\mathbb{R}\to[0,\infty)$, which encodes this relation.

Definition 15.4 (Local gap size). *Let α_k and β_k be the* **edges** *of the support of the density of states (cf.* (15.9)*) and δ_* the constant introduced in Corollary 15.2. Then for any $\delta\in[0,\delta_*)$ we set*

$$\Delta_\delta(\tau) := \begin{cases} \alpha_{k+1}-\beta_k & \text{if } \beta_k-\delta\le\tau\le\alpha_{k+1}+\delta \text{ for some } k=1,\dots,K-1,\\ 1 & \text{if } \tau\le\alpha_1+\delta \text{ or } \tau\ge\beta_K-\delta,\\ 0 & \text{otherwise.}\end{cases} \tag{15.13}$$

Finally, we fix an arbitrarily small **tolerance exponent** $\gamma\in(0,1)$. This number will stay fixed throughout this paper in the same fashion as the model parameters P, p, L and μ. Our main result is stated for spectral parameters $z=\tau+\mathrm{i}\eta$ whose imaginary parts satisfy

$$\eta \ge N^{\gamma-1}\,. \tag{15.14}$$

For a compact statement of the main theorem we define the notion of stochastic domination, introduced in [23] and [26]. This notion is designed to compare sequences of random variables in the large N limit up to small powers of N on high probability sets.

Definition 15.5 (Stochastic domination). *Suppose $N_0 : (0,\infty)^2 \to \mathbb{N}$ is a given function, depending only on the model parameters p, P, L and μ, as well as on the tolerance exponent*

γ. For two sequences, $\varphi = (\varphi^{(N)})_N$ and $\psi = (\psi^{(N)})_N$, of non-negative random variables we say that φ is **stochastically dominated** *by ψ if for all $\varepsilon > 0$ and $D > 0$,*

$$\mathbb{P}\big[\varphi^{(N)} > N^{\varepsilon}\psi^{(N)}\big] \,\le\, N^{-D}, \qquad N \,\ge\, N_0(\varepsilon, D)\,. \tag{15.15}$$

In this case we write $\varphi \prec \psi$.

The threshold $N_0(\varepsilon, D) = N_0(\varepsilon, D, P, p, L, \mu, \gamma)$ will always be an explicit function whose value will be increased throughout the paper, though we will not follow its form. This will happen only finitely many times, ensuring that N_0 stays finite. The threshold is uniform in all other parameters, e.g. $z, i, j, \dots,$ that the sequences φ and ψ may depend on. Typically, we will not mention the existence of N_0, it is implicit in the notation $\varphi \prec \psi$. As an example, we see that the bounded moment condition, (D), implies

$$|h_{ij}| \,\prec\, N^{-1/2}.$$

Actually, the function N_0 depends only on finitely many moment parameters $(\mu_1, \dots, \mu_M)$ instead of the entire sequence μ, where now the number of required moments, $M = M(\varepsilon, D, P, p, L, \gamma)$, is an explicit function.

Now we are ready to state our main result on the local law. Suppose $\mathbf{H} = \mathbf{H}^{(N)}$ is a sequence of self-adjoint random matrices with the corresponding sequence $\mathbf{S} = \mathbf{S}^{(N)}$ of variance matrices and $\rho = \rho^{(N)}$ the induced sequence of densities of state. Recall that δ_* is the positive constant, depending only on p, P and L, introduced in Corollary 15.2 and Δ_δ is defined as in Definition 15.4.

THEOREM 15.6 (Local law). *Suppose that assumptions (A)-(D) are satisfied and fix an arbitrary $\gamma \in (0, 1)$. There is a deterministic function $\kappa = \kappa^{(N)} : \mathbb{H} \to (0, \infty]$ such that uniformly for all $z = \tau + \mathrm{i}\eta \in \mathbb{H}$ with $\eta \ge N^{\gamma-1}$ the resolvents of the random matrices $\mathbf{H} = \mathbf{H}^{(N)}$ satisfy*

$$\max_{i,j} |G_{ij}(z) - m_i(z)\delta_{ij}| \,\prec\, \sqrt{\frac{\rho(z)}{N\eta}} + \frac{1}{N\eta} + \min\Big\{\frac{1}{\sqrt{N\eta}}\,,\, \frac{\kappa(z)}{N\eta}\Big\} \tag{15.16}$$

Furthermore, for any sequence of deterministic vectors $\mathbf{w} = \mathbf{w}^{(N)} \in \mathbb{C}^N$ with $\max_i |w_i| \le 1$ the averaged resolvent diagonal has an improved convergence rate,

$$\Big|\frac{1}{N}\sum_{i=1}^{N} w_i\big(G_{ii}(z) - m_i(z)\big)\Big| \,\prec\, \min\Big\{\frac{1}{\sqrt{N\eta}}\,,\, \frac{\kappa(z)}{N\eta}\Big\}\,. \tag{15.17}$$

In particular, for $w_i = 1$ this implies

$$\Big|\frac{1}{N}\,\mathrm{Im}\,\mathrm{Tr}\,\mathbf{G}(z) - \pi\,\rho(z)\Big| \,\prec\, \min\Big\{\frac{1}{\sqrt{N\eta}}\,,\, \frac{\kappa(z)}{N\eta}\Big\}\,. \tag{15.18}$$

The function κ satisfies the following bounds. There is a constant $\delta \in (0, \delta_*)$, depending only on the model parameters p, P and L, such that with $\Delta = \Delta_\delta$,

$$\kappa(z) \,\le\, \frac{1}{\Delta(\tau)^{1/3} + \rho(z)}\,. \tag{15.19}$$

If z is not too close to the support of the density of states in the sense that

$$(\,\Delta(\tau)^{1/3} + \rho(z))\,\mathrm{dist}(z, \mathrm{supp}\,\rho) \,\ge\, \frac{N^{\gamma}}{(N\eta)^2}\,, \tag{15.20}$$

then κ satisfies the improved bound

$$\kappa(z) \,\le\, \frac{\eta}{\mathrm{dist}(z, \mathrm{supp}\,\rho)\,(\Delta(\tau)^{1/3} + \rho(z))} + \frac{1}{N\eta\,\mathrm{dist}(z, \mathrm{supp}\,\rho)^{1/2}\,(\Delta(\tau)^{1/3} + \rho(z))^{1/2}}\,. \tag{15.21}$$

This local law generalises the previous local laws for stochastic variance matrices $\mathbf{S}$ (see [26] and references therein). It is valid for densities ρ with an edge behaviour different from the square root growth that is known from Wigner's semicircular law. In particular, singularities that interpolate between a square root and a cubic root are possible. In the bulk of the support of the density of states, i.e., where ρ is bounded away from zero, the function κ is bounded. The same is true near the edges, unless the nearby gap is small. The bound deteriorates near small gaps in the support of ρ.

In applications, the sequence $\mathbf{S} = \mathbf{S}^{(N)}$ may be constructed by discretising a continuous limit function. As a simple example, suppose f is a smooth, non-negative, symmetric, $f(x,y) = f(y,x)$, function on $[0,1]^2$ with a positive diagonal, $f(x,x) > 0$. Then the sequence of variance matrices,

$$s_{ij}^{(N)} := \frac{1}{N} f\Big(\frac{i}{N}, \frac{j}{N}\Big)\,, \qquad i,j = 1,\dots,N\,,$$

satisfies conditions (A)-(C). The validity of (C) can be verified by using the criteria proven in Part II. In this case the solution, $\mathbf{m} = (m_1,\dots,m_N)$, of the QVE converges to a limit in the sense that

$$\max_i \sup_z \big| m_i(z) - m(i/N;z) \big| \;\to\; 0\,,$$

where $m : [0,1] \times \overline{\mathbb{H}} \to \overline{\mathbb{H}}$ is the solution of the continuous QVE,

$$-\frac{1}{m(x;z)} \;=\; z + \int_0^1 f(x,y) m(y;z) \,\mathrm{d}y\,, \qquad x \in [0,1]\,,\; z \in \overline{\mathbb{H}}\,.$$

These continuum versions of the QVE and their stability are analysed in Part II of this work as well. In particular, the density of states converges to a limit,

$$\rho^{(N)}(\tau) \;\to\; \frac{1}{\pi} \int_0^1 \operatorname{Im} m(x;\tau) \mathrm{d}x\,.$$

We introduce a notion for expressing that events hold with high probability in the limit as N tends to infinity.

Definition 15.7 (Overwhelming probability). *Suppose $N_0 : (0,\infty) \to \mathbb{N}$ is a given function, depending only on the model parameters p, P, L and μ, as well as on the tolerance exponent γ. For a sequence $A = (A^{(N)})_N$ of random events we say that A hold* **asymptotically with overwhelming probability** *(a.w.o.p.), if for all $D > 0$:*

$$\mathbb{P}[A^{(N)}] \;\geq\; 1 - N^{-D}, \qquad N \;\geq\; N_0(D)\,. \tag{15.22}$$

There is a simple connection between the notions of stochastic domination and asymptotically overwhelming probability. For two sequences $A = A^{(N)}$ and $B = B^{(N)}$ the statement 'A implies B a.w.o.p.' is equivalent to $\mathbb{1}_A \prec \mathbb{1}_B$, where the threshold N_0, implicit in the stochastic domination, does not depend on ε, i.e., $N_0(\varepsilon, D) = N_0(D)$.

We denote by $\lambda_1 \leq \dots \leq \lambda_N$ the eigenvalues of the random matrix $\mathbf{H}$. The following corollary shows that the eigenvalue distribution converges to the density of states as N tends to infinity.

Corollary 15.8 (Convergence of cumulative eigenvalue distribution). *Uniformly for all $\tau \in \mathbb{R}$ the cumulative empirical eigenvalue distribution approaches the integrated density of states,*

$$\bigg| \#\{ i : \lambda_i \leq \tau \} - N \int_{-\infty}^{\tau} \rho(\omega)\, \mathrm{d}\omega \bigg| \;\prec\; \min\bigg\{ \frac{1}{\Delta(\tau)^{1/3} + \rho(\tau)},\, N^{1/5} \bigg\}\,. \tag{15.23}$$

Furthermore, for an arbitrary tolerance exponent $\gamma \in (0,1)$ *there are no eigenvalues away from the support of the density of states,*

$$\max_{k=0}^{K} \#\{\, i : \beta_k + \delta_k \,<\, \lambda_i \,<\, \alpha_{k+1} - \delta_k \} \,=\, 0 \qquad a.w.o.p.\,, \tag{15.24}$$

where we interpret $\beta_0 := -\infty$, $\alpha_{K+1} := +\infty$ *and* δ_k *is defined as* $\delta_0 := \delta_K := N^{\gamma-2/3}$, *as well as*

$$\delta_k \,:=\, \frac{N^{\gamma}}{(\alpha_{k+1} - \beta_k)^{1/3} N^{2/3}}\,, \qquad k = 1, \dots, K-1\,. \tag{15.25}$$

We define the index, $i(\tau)$, of an eigenvalue that we expect close to the spectral parameter τ by

$$i(\tau) \,:=\, \left\lceil N \int_{-\infty}^{\tau} \rho(\omega)\, \mathrm{d}\omega \right\rceil. \tag{15.26}$$

Here, $\lceil \omega \rceil$ denotes the smallest integer that is bigger or equal to ω for any $\omega \in \mathbb{R}$.

Corollary 15.9 (Rigidity of eigenvalues). *Let* $\gamma \in (0,1)$ *be an arbitrary tolerance exponent. Uniformly for all*

$$\tau \,\in\, \bigcup_{k=1}^{K} [\alpha_k + \varepsilon_{k-1}, \beta_k - \varepsilon_k]\,, \tag{15.27}$$

where $\varepsilon_0 := \varepsilon_K := N^{\gamma-2/3}$ *and the other* ε_k *are defined as*

$$\varepsilon_k \,:=\, N^{\gamma} \min\left\{ \frac{1}{N^{3/5}}, \frac{1}{(\alpha_{k+1} - \beta_k)^{1/9} N^{2/3}} \right\}, \qquad k \,=\, 1, \dots, K-1\,, \tag{15.28}$$

the eigenvalues satisfy the rigidity

$$|\lambda_{i(\tau)} - \tau| \,\prec\, \min\left\{ \frac{1}{\rho(\tau)(\Delta(\tau)^{1/3} + \rho(\tau))N}, \frac{1}{N^{3/5}} \right\}. \tag{15.29}$$

For $\tau \in (\alpha_1, \alpha_1 + \varepsilon_0)$ *the eigenvalues at the leftmost edge satisfy*

$$|\lambda_{i(\tau)} - \tau| \,\prec\, N^{-2/3}\,, \tag{15.30}$$

and similarly, for $\tau \in (\beta_K - \varepsilon_K, \beta_K]$ *the eigenvalues at the rightmost edge satisfy*

$$|\lambda_{i(\tau)} - \tau| \,\prec\, N^{-2/3}\,. \tag{15.31}$$

Uniformly for $\tau \in (\beta_k - \varepsilon_k, \alpha_{k+1} + \varepsilon_k)$ *with some* $k = 1, \dots, K-1$, *the eigenvalues close to the internal edge satisfy*

$$\lambda_{i(\tau)} \,\in\, [\beta_k - 2\varepsilon_k, \beta_k + \delta_k] \cup [\alpha_{k+1} - \delta_k, \alpha_{k+1} + 2\varepsilon_k] \qquad a.w.o.p.\,, \tag{15.32}$$

where δ_k *is defined in* (15.25).

Remark 15.10. *The statements* (15.30), (15.31) *and* (15.32) *are an immediate consequence of* (15.29) *and* (15.24). *They simply express the fact that the small number of* $\mathcal{O}(N^{\varepsilon})$ *eigenvalues, very close to the edges, are found in the space that is left for them by the other eigenvalues for which the rigidity statement* (15.29) *applies. For an illustration see Figure 15.2.*

Corollary 15.11 (Delocalisation of eigenvectors). *Let* $\mathbf{u}_i$ *be the normalised eigenvector of* $\mathbf{H}$ *corresponding to the eigenvalue* λ_i. *All eigenvectors are delocalised in the sense that*

$$\max_i \|\mathbf{u}_i\|_\infty \prec \frac{1}{\sqrt{N}}\,.$$

Theorem 15.12 (Isotropic law). *Let* $\mathbf{w},\mathbf{v} \in \mathbb{C}^N$ *be deterministic unit vectors and* $\gamma \in (0,1)$. *Then uniformly for all* $z = \tau + \mathrm{i}\eta \in \mathbb{H}$ *with* $\eta \geq N^{\gamma-1}$,

$$\left|\sum_{i,j=1}^{N} w_i G_{ij}(z) v_j - \sum_{i=1}^{N} m_i(z) w_i v_i\right| \prec \sqrt{\frac{\rho(z)}{N\eta}} + \frac{1}{N\eta} + \min\left\{\frac{1}{\sqrt{N\eta}}, \frac{\kappa(z)}{N\eta}\right\}, \tag{15.33}$$

where κ *is the function from Theorem 15.6.*

Definition 15.13 (q-full random matrix). *We say that* $\mathbf{H}$ *is* q**-full** *for some* $q > 0$ *if either of the following applies:*

- $\mathbf{H}$ *is real symmetric and* $\mathbb{E}\, h_{ij}^2 \geq q/N$ *for all* $i,j = 1,\dots,N$;
- $\mathbf{H}$ *is complex hermitian and for all* $i,j = 1,\dots,N$ *the real symmetric* 2×2*-matrix,*

$$\sigma_{ij} := \begin{pmatrix} \mathbb{E}\,(\mathrm{Re}\, h_{ij})^2 & \mathbb{E}\,(\mathrm{Re}\, h_{ij})(\mathrm{Im}\, h_{ij}) \\ \mathbb{E}\,(\mathrm{Re}\, h_{ij})(\mathrm{Im}\, h_{ij}) & \mathbb{E}\,(\mathrm{Im}\, h_{ij})^2 \end{pmatrix},$$

  satisfies $\sigma_{ij} \geq q/N$.

Theorem 15.14 (Universality). *Suppose that in addition to (A)-(D) being satisfied, the matrix* $\mathbf{H}$ *is* q*-full. Then for all* $\varepsilon > 0$, $n \in \mathbb{N}$ *and all smooth compactly supported observables* $F : \mathbb{R}^n \to \mathbb{R}$, *there are two positive constants* C *and* c *such that for any* $\tau \in \mathbb{R}$ *with* $\rho(\tau) \geq \varepsilon$ *the local eigenvalue distribution is universal,*

$$\left|\mathbb{E}\, F\Big(\big(N\rho(\lambda_{i(\tau)})(\lambda_{i(\tau)} - \lambda_{i(\tau)+j})\big)_{j=1}^n\Big) - \mathbb{E}_{\mathrm{G}}\, F\Big(\big(N\rho_{\mathrm{sc}}(0)(\lambda_{\lceil N/2\rceil} - \lambda_{\lceil N/2\rceil+j})\big)_{j=1}^n\Big)\right| \leq C N^{-c}.$$

Here, $\mathbb{E}_{\mathrm{G}}$ *denotes the expectation with respect to the standard Gaussian ensemble, i.e., with respect to GUE and GOE in the cases of complex Hermitian and real symmetric* $\mathbf{H}$, *respectively, and* $\rho_{\mathrm{sc}}(0) = 1/(2\pi)$ *is the value of Wigner's semicircle law at the origin.*

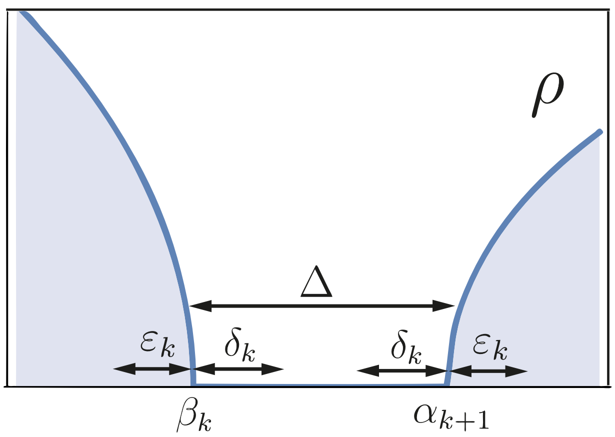

Figure 15.2: At the edges of a gap of length Δ in $\mathrm{supp}\,\rho$ the bound on the eigenvalue fluctuation is δ_k inside the gap and ε_k inside the support

This paper focusses on random matrices in the real symmetric and the complex hermitian symmetry class. The third universality class, the quaternion self dual case can be treated in a similar fashion.

We introduce a few conventions and notations used throughout this paper.

Convention 15.15 (Constants and comparison relation). *We use the convention that every positive constant with a lower star index, such as* δ_*, c_* *and* λ_*, *explicitly depends only on the model parameters* P, p *and* L. *These dependencies can be reconstructed from the proofs, but we will not follow them. Constants* $c, c_1, c_2, \dots, C, C_1, C_2, \dots$ *also depend only on*

P, p and L. They will have a local meaning within a specific proof.

For two non-negative functions φ and ψ depending on a set of parameters $u \in U$, we use the **comparison relation**

$$\varphi \gtrsim \psi, \tag{15.34}$$

if there exists a positive constant c, depending explicitly on P, p and L such that $\varphi(u) \geq c\psi(u)$ for all $u \in U$. The notation $\psi \sim \varphi$ means that both $\psi \lesssim \varphi$ and $\psi \gtrsim \varphi$ hold true. In this case we say that ψ and φ are **comparable**. *We also write $\psi = \varphi + \mathcal{O}(\vartheta)$, if $|\psi - \varphi| \lesssim \vartheta$.*

We denote the normalised scalar product between two vectors $\mathbf{w}, \mathbf{u} \in \mathbb{C}^N$ and the average of a vector by

$$\langle \mathbf{w}, \mathbf{u} \rangle := \frac{1}{N} \sum_{i=1}^{N} w_i u_i, \qquad \langle \mathbf{w} \rangle := \frac{1}{N} \sum_{i=1}^{N} w_i. \tag{15.35}$$

15.2 Gaussian random matrices with correlated entries

The results of the Subsection 15.1 can be applied to Gaussian random matrices having dependent entries with a translation invariant correlation structure. Let $N \in \mathbb{N}$ and consider the complex self-adjoint random matrix

$$\mathbf{H}^{(N)} = (h_{ij}^{(N)})_{i,j \in \mathbb{T}^{(N)}}, \tag{15.36}$$

indexed by a discrete torus

$$\mathbb{T} = \mathbb{T}^{(N)} := \mathbb{Z}/N\mathbb{Z}. \tag{15.37}$$

We assume that the matrix is centred, i.e.,

$$\mathbb{E}\, h_{ij} = 0, \quad \forall\, i, j \in \mathbb{T}. \tag{15.38a}$$

Moreover, we assume that $\mathbf{H}$ is Gaussian, i.e., the elements h_{ij} are jointly Gaussian. The covariances of the elements of $\mathbf{H}$ are specified by two self-adjoint matrices $\mathbf{A} = \mathbf{A}^{(N)}$ and $\mathbf{B} = \mathbf{B}^{(N)}$ with elements $a_{ij} \in \mathbb{C}$ and $b_{ij} \in \mathbb{C}$, $i, j \in \mathbb{T}$, through

$$\mathbb{E}\, h_{ij}\overline{h_{kl}} = \frac{1}{N}(a_{i-k,j-l} + b_{i-l,j-k}), \qquad \forall\, i, j, k, l \in \mathbb{T}. \tag{15.38b}$$

Here the subtractions in $i - k$ and $j - l$, etc., are done in $\mathbb{T}$.

We will say that **A decays exponentially** if there is a constant $\nu > 0$, such that

$$|a_{xy}| \leq \mathrm{e}^{-\nu(|x|+|y|)}, \qquad \forall\, x, y \in \mathbb{T}, \tag{15.39}$$

where $|x|$ is the distance of x from 0 on $\mathbb{T}$, i.e., $|x| := \min_{k \in \mathbb{Z}} |x + kN|$.

We say that **A** is **non-resonant** if

$$\sum_{x \in \mathbb{T}} \mathrm{e}^{\mathrm{i}2\pi\phi x} a_{x0} \geq \xi, \qquad \forall\, \phi \in [0, 1], \tag{15.40}$$

for some $\xi > 0$.

Let $\mathbb{S} = \mathbb{S}^{(N)} := N^{-1}\mathbb{T}^{(N)}$ be the discrete dual torus of $\mathbb{T} = \mathbb{T}^{(N)}$.

THEOREM 15.16 (Local law for Gaussian matrices with correlated entries). *Suppose* $\mathbf{A}$ *decays exponentially and is non-resonant, i.e.,* (15.39) *and* (15.40) *hold with some constants* $\nu, \xi > 0$. *Let* $\gamma > 0$ *be any tolerance exponent. Then the resolvent* $\mathbf{G}(z) = (\mathbf{H} - z)^{-1}$ *satisfies the optimal local law everywhere, i.e., uniformly for all* $z = \tau + \mathrm{i}\eta \in \mathbb{H}$ *with* $\eta \geq N^{\gamma - 1}$,

$$\Big|\frac{1}{N}\mathrm{Tr}\,\mathbf{G}(z) - \frac{1}{N}\sum_{\phi \in \mathbb{S}} m_\phi(z)\Big| \prec \frac{1}{N\eta}, \tag{15.41}$$

where $\mathbf{m}(z) = (m_\phi(z))_{\phi \in \mathbb{S}} \in \mathbb{H}^{\mathbb{S}}$ *is the unique solution of the QVE*

$$-\frac{1}{m_\phi(z)} = z + \sum_{\theta \in \mathbb{S}} \widehat{a}_{\phi\theta}\, m_\theta(z), \tag{15.42}$$

and

$$\widehat{a}_{\phi\theta} := \frac{1}{N}\sum_{x,y \in \mathbb{T}} \mathrm{e}^{\mathrm{i}2\pi(x\phi - y\theta)} a_{xy}\,. \tag{15.43}$$

The solution $\mathbf{m}$ *is uniformly bounded in* N. *The density of states is a symmetric function* $\rho = \rho^{(N)} : \mathbb{R} \to [0, \infty)$, *that is uniquely determined by*

$$\frac{1}{N}\sum_{\phi \in \mathbb{S}} m_\phi(z) := \int_{-\infty}^{\infty} \frac{\rho(\omega)\mathrm{d}\omega}{\omega - z}, \qquad z \in \mathbb{H}, \tag{15.44}$$

and it has the following properties:

(i) *Single interval support:* $\mathrm{supp}\,\rho = [-\sigma_*, \sigma_*]$, *for some* $\sigma_* \sim 1$;

(ii) *Uniform boundedness:* $\|\rho\|_\infty \lesssim 1$;

(iii) 1/2-*Hölder continuity:* $|\rho(\tau_1) - \rho(\tau_2)| \lesssim |\tau_1 - \tau_2|^{1/2}$, *for all* $\tau_1, \tau_2 \in \mathbb{R}$;

(iv) *Real analyticity away from the edges:* $|\partial_\tau^k \rho(\tau)| \leq k!(C_0/\rho(\tau)^3)^k$, $\tau \in (-\sigma_*, \sigma_*)$;

(v) *Square-root shape around the edges of the support: There is* $C_1 \sim 1$ *such that*

$$\rho(-\sigma_* + \omega) = \rho(\sigma_* - \omega) = C_1\omega^{1/2} + \mathcal{O}(\omega), \qquad \omega \geq 0, \tag{15.45}$$

while for any $\delta > 0$ *we have* $\rho(\tau) \gtrsim \delta^{1/2}$ *and* $-\sigma_* + \delta \leq \tau \leq \sigma_* - \delta$.

The off-diagonal resolvent elements inherit the decay from the matrix $\mathbf{A}$, *i.e.,*

$$\big|G_{xy}(z) - q_{x-y}(z)\big| \prec \sqrt{\frac{\rho(z)}{N\eta}} + \frac{1}{N\eta}, \tag{15.46}$$

where

$$q_x(z) := \frac{1}{N}\sum_{\phi \in \mathbb{S}} \mathrm{e}^{-\mathrm{i}2\pi x\phi} m_\phi(z), \qquad x \in \mathbb{T} \tag{15.47}$$

decays exponentially

$$|q_x(z)| \lesssim \mathrm{e}^{-\nu'|x|} + \mathrm{dist}(z, \{\sigma_*, -\sigma_*\})^{-1/2} N^{-1/2}, \qquad x \in \mathbb{T}, \tag{15.48}$$

for some $\nu' \leq \nu$, *satisfying* $\nu' \sim 1$. *Here the comparison relations* $\lesssim$ *depend only on the two model parameters* ξ *and* ν.

Note that there are no explicit conditions on the correlation matrix $\mathbf{B}$. However, $\mathbf{A}$ and $\mathbf{B}$ are related. For example, if $\mathbf{H}$ is real valued then $\mathbf{A} = \mathbf{B}$.

We will prove these results in Section 21. There we will also consider correlation matrices that do not decay exponentially (cf. Proposition 21.1). We finish with a remark that optimal local law implies bulk universality:

REMARK 15.17. *Suppose* $\mathbf{H}$ *satisfies* (15.38) *with a non-resonant and exponentially decaying correlation matrix* $\mathbf{A}$. *We also assume that* $\mathbf{H}$ *contains a small standard Gaussian component, i.e.,* $\mathbf{H}$ *can be represented as a sum of two independent Gaussian random matrices,*

$$\mathbf{H} \,=\, \mathbf{H}_0 + \varepsilon \mathbf{W}\,,$$

where $\varepsilon > 0$ *and* $\mathbf{W}$ *is a standard GUE/GOE matrix. Since* $\mathbf{H}$ *itself is Gaussian, this condition can be translated into a certain non-degeneracy condition on the covariance matrices* $\mathbf{A}$ *and* $\mathbf{B}$, *similar to the q-fullness condition. Then the eigenvalues of* $\mathbf{H}$ *satisfy bulk universality, i.e., the conclusion of Theorem 15.14 holds. The proof is similar but even simpler than that of Theorem 15.14 given in Section 20.2; local law and small Gaussian component imply universality. Since the original matrix* $\mathbf{H}$ *itself has a Gaussian component, no comparison argument and moment matching are necessary.*

16 Bound on the random perturbation of the QVE

We introduce the notation $\mathbf{G}^{(V)}$ for the resolvent of the matrix $\mathbf{H}^{(V)}$, which is identical to $\mathbf{H}$ except for the removal of the rows and columns corresponding to the indices $V \subseteq \{1, \dots, N\}$. The enumeration of the indices is kept, even though $\mathbf{G}^{(V)}$ has a lower dimension.

The diagonal elements of the resolvent, $\mathbf{g} := (G_{11}, \dots, G_{NN})$, satisfy the perturbed quadratic vector equation

$$-\frac{1}{g_i(z)} = z + \sum_{j=1}^{N} s_{ij}\, g_j(z) + d_i(z)\,, \tag{16.1}$$

for all $z \in \mathbb{H}$ and $i = 1, \dots, N$. The random perturbation $\mathbf{d} = (d_1, \dots, d_N)$ is given by

$$d_k \,:=\, \sum_{i \neq j}^{(k)} h_{ki} G_{ij}^{(k)} h_{jk} \,+\, \sum_{i}^{(k)} (|h_{ki}|^2 - s_{ki})\, G_{ii}^{(k)} \,-\, \sum_{i}^{(k)} s_{ki} \frac{G_{ik} G_{ki}}{g_k} \,-\, h_{kk} \,-\, s_{kk}\, g_k\,. \tag{16.2}$$

Here and in the following, the upper indices on the sums indicate which indices are not summed over. For the proof of this simple identity as well as (16.3) below via the Schur complement formula we refer to [26]. As in (16.2) we will often omit the dependence on the spectral parameter z in our notation, i.e., $G_{ij} = G_{ij}(z)$, $d_k = d_k(z)$, etc..

We will now derive an upper bound on $\|\mathbf{d}\|_\infty = \max_i |d_i|$, provided $|g_i - m_i|$ is bounded by a small constant. At the same time we will control the off-diagonal elements G_{kl} of the resolvent. These satisfy the identity

$$G_{kl} \,=\, G_{kk} G_{ll}^{(k)} \sum_{i,j}^{(kl)} h_{ki} G_{ij}^{(kl)} h_{jl} \,-\, G_{kk} G_{ll}^{(k)} h_{kl}\,, \tag{16.3}$$

for $k \neq l$. The strategy in what follows below is that (16.2) and (16.3) are used to improve a rough bound on the entries of the resolvent $\mathbf{G}$ to get the correct bounds on the random perturbation and the off-diagonal resolvent elements. Later, in Section 17, the stability of the QVE under the small perturbation, $\mathbf{d}$, will provide the improved bound on the diagonal elements, $G_{ii} - m_i = g_i - m_i$.

We introduce a short notation for the difference between $\mathbf{g}$ and the solution $\mathbf{m}$ of the unperturbed equation (15.3),

$$\begin{aligned} \Lambda_{\mathrm{d}}(z) &:= \max_i |G_{ii}(z) - m_i(z)| \,, \\ \Lambda_{\mathrm{o}}(z) &:= \max_{i \neq j} |G_{ij}(z)| \,, \\ \Lambda(z) &:= \max\big(\Lambda_{\mathrm{d}}(z), \Lambda_{\mathrm{o}}(z)\big) \,. \end{aligned} \tag{16.4}$$

Lemma 16.1 (Bound on perturbation). *There is a small positive constant λ_*, such that uniformly for all spectral parameters $z = \tau + \mathrm{i}\eta \in \mathbb{H}$ with $\eta \geq N^{\gamma-1}$:*

$$\begin{aligned} |d_k(z)|\, \mathbb{1}\Big(\Lambda(z) \leq \frac{\lambda_*}{1+|z|}\Big) &\prec \frac{1}{1+|z|^2}\sqrt{\frac{\mathrm{Im}\langle \mathbf{g}(z)\rangle}{N\eta}} + \frac{1}{\sqrt{N}} \,, \\ \Lambda_{\mathrm{o}}(z)\, \mathbb{1}\Big(\Lambda(z) \leq \frac{\lambda_*}{1+|z|}\Big) &\prec \frac{1}{1+|z|^2}\bigg(\sqrt{\frac{\mathrm{Im}\langle \mathbf{g}(z)\rangle}{N\eta}} + \frac{1}{\sqrt{N}}\bigg) \,. \end{aligned} \tag{16.5}$$

This lemma is analogous to Lemma 5.2 in [26] with minor modifications. For the completeness of this paper, we repeat these arguments. One modifications is that our estimates also deal with the regime where $|z|$ is large. For the proof of this lemma we will need an additional property of the solution of the QVE that is a corollary of Theorem 18.1, where all properties of $\mathbf{m}$ taken from Part II are summarised.

Corollary 16.2. *The absolute value of the solution of the QVE satisfies*

$$|m_i(z)| \sim \frac{1}{1+|z|} \,, \qquad z \in \mathbb{H} \,,\; i = 1, \dots, N \,. \tag{16.6}$$

Proof of Lemma 16.1. Here we use the three large deviation estimates,

$$\Big|\sum_{i \neq j}^{(k)} h_{ki} G^{(k)}_{ij} h_{jk}\Big| \prec \Big(\sum_{i \neq j}^{(k)} s_{ki} s_{jk} \big|G^{(k)}_{ij}\big|^2\Big)^{1/2} , \tag{16.7a}$$

$$\Big|\sum_{i,j}^{(kl)} h_{ki} G^{(kl)}_{ij} h_{jl}\Big| \prec \Big(\sum_{i,j}^{(kl)} s_{ki} s_{jl} \big|G^{(kl)}_{ij}\big|^2\Big)^{1/2} , \tag{16.7b}$$

$$\Big|\sum_{i}^{(k)} (|h_{ki}^2| - s_{ki})\, G^{(k)}_{ii}\Big| \prec \Big(\sum_{i}^{(k)} s_{ki}^2 \big|G^{(k)}_{ii}\big|^2\Big)^{1/2} . \tag{16.7c}$$

Since $\mathbf{G}^{(V)}$ is independent of the rows and columns of $\mathbf{H}$ with indices in V, these estimates follow directly from the large deviation bounds in Appendix C of [26]. Furthermore, we use

$$h_{ij} \prec N^{-1/2} \,, \qquad s_{ij} \leq N^{-1} \,. \tag{16.8}$$

The latter inequality is just assumption (15.4) and the bound on h_{ij} follows from (15.7).

We will now show that the removal of a few rows and columns in $\mathbf{H}$ will only have a small effect on the entries of the resolvent. The general resolvent identity,

$$G_{ij} = G^{(k)}_{ij} + \frac{G_{ik} G_{kj}}{G_{kk}} \,, \qquad k \notin \{i, j\} \,, \tag{16.9}$$

leads to the bound

$$\big|G_{ij}^{(k)} - G_{ij}\big|\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;=\; \frac{|G_{ik}G_{kj}|}{|g_k|}\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\lesssim\; (1+|z|)\,\Lambda_{\rm o}^2\,. \tag{16.10}$$

In the inequality we used that $|m_k(z)| \sim (1+|z|)^{-1}$ (cf. Corollary 16.2), $|g_k| = |m_k| + \mathcal{O}(\Lambda)$ and that λ_* is chosen to be small enough. We use (16.10) in a similar calculation for $G_{ij}^{(l)}$ and find that on the event where $\Lambda \le (1+|z|)^{-1}\lambda_*$,

$$\big|G_{ij}^{(kl)} - G_{ij}^{(l)}\big| \;=\; \frac{\big|G_{ik}^{(l)}G_{kj}^{(l)}\big|}{\big|G_{kk}^{(l)}\big|} \;\lesssim\; \frac{\big(|G_{ik}| + \mathcal{O}\big((1+|z|)\Lambda_{\rm o}^2\big)\big)\big(|G_{kj}| + \mathcal{O}\big((1+|z|)\Lambda_{\rm o}^2\big)\big)}{|g_k| + \mathcal{O}\big((1+|z|)\Lambda_{\rm o}^2\big)}\,. \tag{16.11}$$

Again using (16.10) and that the denominator of the last expression is comparable to $(1+|z|)^{-1}$, we conclude

$$|G_{ij}^{(kl)} - G_{ij}|\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\lesssim\; (1+|z|)\,\Lambda_{\rm o}^2\,, \tag{16.12}$$

provided λ_* is small. Therefore, we see that it is possible to remove one or two upper indices from G_{ij} for the price of a term, whose size is at most $(1+|z|)\,\Lambda_{\rm o}^2$.

We have now collected all necessary ingredients and use them to estimate all the terms in (16.2) one by one. We start with the first summand. By (16.7a) we find

$$\Big|\sum_{i\neq j}^{(k)} h_{ki}G_{ij}^{(k)}h_{jk}\Big|^2 \;\prec\; \sum_{i\neq j}^{(k)} s_{ki}s_{jk}|G_{ij}^{(k)}|^2 \;\le\; \frac{1}{N^2}\sum_{i\neq j}^{(k)}|G_{ij}^{(k)}|^2\,. \tag{16.13}$$

With the help of (16.10) we remove the upper index from $G_{ij}^{(k)}$ and get

$$\Big|\sum_{i\neq j}^{(k)} h_{ki}G_{ij}^{(k)}h_{jk}\Big|^2\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\prec\; \big(\Lambda_{\rm o}^2 + (1+|z|)^2\Lambda_{\rm o}^4\big)\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\lesssim\; \Lambda_{\rm o}^2\,. \tag{16.14}$$

For the second summand in (16.2) we use the large deviation bound for the diagonal, (16.7c), and find that

$$\Big|\sum_{i}^{(k)}(|h_{ki}|^2 - s_{ki})\,G_{ii}^{(k)}\Big|^2 \;\prec\; \sum_{i}^{(k)} s_{ki}^2\,\big|G_{ii}^{(k)}\big|^2 \;\le\; \frac{1}{N^2}\sum_{i}^{(k)}\big|G_{ii}^{(k)}\big|^2\,. \tag{16.15}$$

By removing the upper index again we estimate

$$\big|G_{ii}^{(k)}\big|\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\lesssim\; |m_i| + \Lambda_{\rm d} + (1+|z|)\,\Lambda_{\rm o}^2\,. \tag{16.16}$$

We use this in (16.15) and for sufficiently small λ_* we arrive at

$$\Big|\sum_{i}^{(k)}(|h_{ki}|^2 - s_{ki})\,G_{ii}^{(k)}\Big|^2\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\prec\; \frac{1}{(1+|z|)^2N}\,. \tag{16.17}$$

The third summand in (16.2) is estimated directly by

$$\Big|\sum_{i}^{(k)} s_{ki}\frac{G_{ik}G_{ki}}{g_k}\Big|\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\le\; \frac{\Lambda_{\rm o}^2}{|g_k|}\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \;\lesssim\; \Lambda_{\rm o}\,. \tag{16.18}$$

We combine the estimates for the individual terms (16.14), (16.17), (16.18) and (16.8). Altogether we conclude that

$$|d_k|\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \prec \Lambda_{\mathrm{o}}(z) + \frac{1}{\sqrt{N}}\,. \tag{16.19}$$

We will now derive in a similar fashion a stochastic domination bound for the off-diagonal error term Λ_{o}. Afterwards, we will combine the two bounds and infer the claim of the lemma. For the off-diagonal error term we proceed along the same lines as for $|d_k|$, using (16.3) instead of (16.2). For $k \ne l$ we find

$$|G_{kl}|^2 \prec |g_k|^2 \big|G^{(k)}_{ll}\big|^2 \Big(\frac{1}{N^2}\sum_{i,j}^{(kl)} \big|G^{(kl)}_{ij}\big|^2 + \frac{1}{N}\Big)\,. \tag{16.20}$$

Here, we applied the large deviation bound (16.7b). With the Ward identity for the resolvent $\mathbf{G}^{(kl)}$,

$$\sum_j^{(kl)} \big|G^{(kl)}_{ij}\big|^2 = \frac{\operatorname{Im} G^{(kl)}_{ii}}{\eta}\,. \tag{16.21}$$

and (16.10) for removing the upper index of $G^{(k)}_{ll}$, we get

$$|G_{kl}|^2\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \prec \frac{1}{(1+|z|)^4}\Big(\frac{1}{N^2\eta}\sum_i^{(kl)} \operatorname{Im} G^{(kl)}_{ii} + \frac{1}{N}\Big)\,. \tag{16.22}$$

We remove the upper indices from $G^{(kl)}_{ii}$ and end up with

$$\Lambda_{\mathrm{o}}\,\mathbb{1}\Big(\Lambda \le \frac{\lambda_*}{1+|z|}\Big) \prec \frac{1}{(1+|z|)^2}\bigg(\sqrt{\frac{\operatorname{Im}\langle \mathbf{g}\rangle}{N\eta}} + \sqrt{\frac{1+|z|}{N\eta}}\,\Lambda_{\mathrm{o}} + \frac{1}{\sqrt{N}}\bigg)\,. \tag{16.23}$$

The bound remains true without the summand containing Λ_{o} on the right hand side, since this term can be absorbed into the left hand side, as its coefficient is bounded by $N^{-\gamma/2}$, while on the left Λ_{o} is not multiplied by a small coefficient. Putting (16.19) and (16.23) together yields the desired result (16.5). □

17 Local law away from local minima

In this section we will use the stability of the QVE to establish the main result away from the local minima of the density of states inside its own support, i.e. away from the set

$$\mathbb{M} := \big\{\tau \in \operatorname{supp}\rho :\ \tau \text{ is the location of a local minimum of } \rho\big\}\,. \tag{17.1}$$

The case where z is close to $\mathbb{M}$ requires a more detailed analysis. This is given is Section 18. At the end of this section we will have proven the following proposition.

Proposition 17.1 (Local law away from local minima). *Let δ_* be any positive constant, depending only on the model parameters p, P and L. Then, uniformly for all $z = \tau + \mathrm{i}\eta$ with $\eta \ge N^{\gamma-1}$ and $\operatorname{dist}(z,\mathbb{M}) \ge \delta_*$, we have*

$$\begin{aligned}
(1+|z|)^2\Lambda_{\mathrm{d}}(z) + \|\mathbf{d}(z)\|_\infty &\prec (1+|z|)^{-2}\sqrt{\frac{\rho(z)}{N\eta}} + (1+|z|)^{-6}\frac{1}{N\eta} + \frac{1}{\sqrt{N}}\,,\\
\Lambda_{\mathrm{o}}(z) &\prec (1+|z|)^{-2}\sqrt{\frac{\rho(z)}{N\eta}} + (1+|z|)^{-4}\frac{1}{N\eta} + (1+|z|)^{-2}\frac{1}{\sqrt{N}}\,.
\end{aligned} \tag{17.2}$$

Furthermore, on the same domain, for any sequence of deterministic vectors $\mathbf{w} = \mathbf{w}^{(N)} \in \mathbb{C}^N$ *with the uniform bound,* $\|\mathbf{w}\|_\infty \le 1$*, we have*

$$|\langle \mathbf{w}, \mathbf{g}(z) - \mathbf{m}(z)\rangle| \prec (1+|z|)^{-3}\frac{\rho(z)}{N\eta} + (1+|z|)^{-7}\frac{1}{(N\eta)^2} + (1+|z|)^{-2}\frac{1}{N} . \tag{17.3}$$

This proposition shows the local law (Theorem 15.6) away from the set $\mathbb{M}$. Its proof uses a continuity argument in z. In particular, continuity of the solution of the QVE is needed. The statement of the following corollary is part of the properties of $\mathbf{m}$ listed in Theorem 18.1.

Corollary 17.2. *For every* $i = 1, \dots, N$ *there is a probability density* $v_i : \mathbb{R} \to [0, \infty)$ *such that* m_i *is the Stieltjes transform of this density, i.e.,*

$$m_i(z) = \int_{\mathbb{R}} \frac{v_i(\tau)\,\mathrm{d}\tau}{\tau - z}, \qquad z \in \mathbb{H}. \tag{17.4}$$

The solution of the QVE is uniformly Hölder-continuous,

$$\|\mathbf{m}(z_1) - \mathbf{m}(z_2)\|_\infty \lesssim |z_1 - z_2|^{1/3}, \qquad z_1, z_2 \in \overline{\mathbb{H}}. \tag{17.5}$$

Since the solution can be extended to the real line, it is the harmonic extension to the complex upper half plane of its own restriction to the real line. Therefore, $\operatorname{Im} m_i(\tau) = \pi\, v_i(\tau)$ for $\tau \in \mathbb{R}$. The density of states is the average of the probability densities v_i, i.e., $\rho = \langle \mathbf{v}\rangle$.

Since we will estimate the difference, $\mathbf{g} - \mathbf{m}$, we start by deriving an equation for this quantity. Using the QVE for $\mathbf{m}$ and the perturbed equation (16.1) for $\mathbf{g}$ we find

$$\begin{aligned} g_i - m_i &= -\frac{1}{z + (\mathbf{S}\mathbf{g})_i + d_i} + \frac{1}{z + (\mathbf{S}\mathbf{m})_i} \\ &= \frac{(\mathbf{S}(\mathbf{g}-\mathbf{m}))_i + d_i}{(z + (\mathbf{S}\mathbf{g})_i + d_i)(z + (\mathbf{S}\mathbf{m})_i)} \\ &= m_i^2 (\mathbf{S}(\mathbf{g}-\mathbf{m}))_i + m_i(g_i - m_i)(\mathbf{S}(\mathbf{g}-\mathbf{m}))_i + m_i\, g_i\, d_i . \end{aligned} \tag{17.6}$$

Rearranging the terms leads to

$$\big((\mathbf{1} - \operatorname{diag}(\mathbf{m})^2 \mathbf{S})(\mathbf{g}-\mathbf{m})\big)_i = m_i(g_i - m_i)(\mathbf{S}(\mathbf{g}-\mathbf{m}))_i + m_i^2\, d_i + m_i\,(g_i - m_i)\, d_i . \tag{17.7}$$

In the proof of Proposition 17.1 we will view (17.7) as a quadratic equation for $\mathbf{g} - \mathbf{m}$ and we use its stability to bound Λ_{d} in terms of $\|\mathbf{d}\|_\infty$. We will now demonstrate this effect in the case when z is far away from the support of the density of states.

Lemma 17.3 (QVE stability away from the support). *For* $z \in \mathbb{H}$ *with* $|z| \ge 10$*, we have*

$$\Lambda_{\mathrm{d}}(z)\, \mathbb{1}(\Lambda_{\mathrm{d}}(z) \le 4|z|^{-1}) \lesssim |z|^{-2} \|\mathbf{d}(z)\|_\infty . \tag{17.8}$$

Furthermore, there is a matrix valued function $\mathbf{T} : \mathbb{H} \to \mathbb{C}^{N\times N}$*, depending only on* S *and satisfying the uniform bound* $\|\mathbf{T}(z)\|_{\infty\to\infty} \lesssim 1$*, such that for all* $\mathbf{w} \in \mathbb{C}^N$ *and* $|z| \ge 10$ *the averaged difference between* $\mathbf{g}$ *and* $\mathbf{m}$ *satisfies the improved bound*

$$\big|\big\langle \mathbf{w}, \mathbf{g}(z) - \mathbf{m}(z)\big\rangle\big|\, \mathbb{1}\big(\Lambda_{\mathrm{d}}(z) \le 4|z|^{-1}\big) \lesssim |z|^{-2}\big(\|\mathbf{w}\|_\infty \|\mathbf{d}(z)\|_\infty^2 + |\langle \mathbf{T}(z)\mathbf{w}, \mathbf{d}(z)\rangle|\big) . \tag{17.9}$$

Proof. Since the matrix $\mathbf{S}$ is flat (cf. (15.4)), it satisfies the norm bound $\|\mathbf{S}\|_{\infty\to\infty} \le 1$. We also have the trivial bound $|m_i(z)| \le 1/\operatorname{dist}(z, \operatorname{supp}\rho) \le 2|z|^{-1} \le 1/5$ at our disposal. This follows directly from the Stieltjes transform representation (17.4). In particular,

$$\big\|(\mathbf{1} - \operatorname{diag}(\mathbf{m})^2 \mathbf{S})^{-1}\big\|_{\infty\to\infty} \le 2, \tag{17.10}$$

from the geometric series. By inverting the matrix $\mathbf{1}-\operatorname{diag}(\mathbf{m})^2\mathbf{S}$ and using the trivial bound on $\mathbf{m}$ in (17.7) we find

$$\Lambda_{\mathrm{d}}(z) \,\le\, 4\Big(\,|z|^{-1}\Lambda_{\mathrm{d}}(z)^2+|z|^{-1}\Lambda_{\mathrm{d}}(z)\|\mathbf{d}(z)\|_\infty+2\,|z|^{-2}\|\mathbf{d}(z)\|_\infty\Big)\,. \tag{17.11}$$

Using the bound inside the indicator function from (17.8) and $|z|\ge 10$ the assertion (17.8) of the lemma follows.

The bound for the average, (17.9), follows by taking the inverse of $\mathbf{1}-\operatorname{diag}(\mathbf{m})^2\mathbf{S}$ on both sides of (17.7) and using (17.8) and $|m_i|\sim|z|^{-1}$. □

For the proof of Proposition 17.1 we use the stability of (17.7) also close to $\operatorname{supp}\rho$. This requires more care and is carried out in detail in Part II of this work. The result of that analysis is Theorem 18.2. Here we will only need the following consequence of that theorem.

Corollary 17.4 (Rough stability). *Suppose δ_* is an arbitrary positive constant, depending only on the model parameters p, P and L. Let $\mathbf{d}:\mathbb{H}\to\mathbb{C}^N$, $\mathbf{g}:\mathbb{H}\to(\mathbb{C}\setminus\{0\})^N$ be arbitrary vector valued functions on the complex upper half plane that satisfy*

$$-\frac{1}{g_i(z)} \,=\, z+\sum_{j=1}^N s_{ij}\,g_j(z)+d_i(z)\,,\quad z\in\mathbb{H}\,. \tag{17.12}$$

There exist a positive constant λ_*, such that the QVE is stable away from $\mathbb{M}$,

$$\|\mathbf{g}(z)-\mathbf{m}(z)\|_\infty\,\mathbb{1}\big(\|\mathbf{g}(z)-\mathbf{m}(z)\|_\infty\le\lambda_*\big)\,\lesssim\,\|\mathbf{d}(z)\|_\infty\,,\quad z\in\mathbb{H}\,,\ \operatorname{dist}(z,\mathbb{M})\ge\delta_*\,. \tag{17.13}$$

Furthermore, there is a matrix valued function $\mathbf{T}:\mathbb{H}\to\mathbb{C}^{N\times N}$, depending only on S and satisfying the uniform bound $\|\mathbf{T}(z)\|_{\infty\to\infty}\lesssim 1$, such that for all $\mathbf{w}\in\mathbb{C}^N$,

$$|\langle\mathbf{w},\mathbf{g}(z)-\mathbf{m}(z)\rangle|\,\mathbb{1}\big(\|\mathbf{g}(z)-\mathbf{m}(z)\|_\infty\le\lambda_*\big)\,\lesssim\,\|\mathbf{w}\|_\infty\|\mathbf{d}(z)\|_\infty^2+|\langle\mathbf{T}(z)\mathbf{w},\mathbf{d}(z)\rangle|\,, \tag{17.14}$$

for $z\in\mathbb{H}$ with $\operatorname{dist}(z,\mathbb{M})\ge\delta_*$.

Furthermore, the following fluctuation averaging result is needed. It was first established for generalised Wigner matrices with Bernoulli distributed entries in [37].

Theorem 17.5 (Fluctuation Averaging). *For any $z\in\mathbb{D}$ and any sequence of deterministic vectors $\mathbf{w}=\mathbf{w}^{(N)}\in\mathbb{C}^N$ with the uniform bound, $\|\mathbf{w}\|_\infty\le 1$ the following holds true: If $\Lambda_{\mathrm{o}}(z)\prec\Phi/(1+|z|)^2$ for some deterministic (N-dependent) $\Phi\le N^{-\gamma/3}$ and $\Lambda(z)\prec N^{-\gamma/3}/(1+|z|)$ a.w.o.p., then*

$$\langle\mathbf{w},\mathbf{d}(z)\rangle\,\prec\,(1+|z|)^{-1}\,\Phi^2+\frac{1}{N}\,. \tag{17.15}$$

Proof. The proof directly follows the one given in [26]. We only mention some minor necessary modifications. Let $Q_kX:=X-\mathbb{E}[X|\mathbf{H}^{(k)}]$ be the complementary projection to the conditional expectation of a random variable X given the matrix $\mathbf{H}^{(k)}$, in which the k-th row and column are removed. From the definition of $\mathbf{d}$ in (16.2) and Schur's complement formula in the form,

$$\frac{1}{G_{kk}} \,=\, h_{kk}-z-\sum_{i,j}^{(k)}h_{ki}G_{ij}^{(k)}h_{jk}\,, \tag{17.16}$$

we infer the identity

$$d_k \,=\, -Q_k\frac{1}{G_{kk}}-s_{kk}G_{kk}-\sum_i^{(k)}s_{ki}\frac{G_{ik}G_{ki}}{G_{kk}}\,.$$

In particular, we have that a.w.o.p.

$$\Big|d_k + Q_k \frac{1}{G_{kk}}\Big| \lesssim (1+|z|)^{-1}\frac{1}{N} + (1+|z|)\Lambda_{\mathrm{o}}^2 \,.$$

Thus, proving (17.15) reduces to showing

$$\Big|\frac{1}{N}\sum_{k=1}^{N} w_k\, Q_k \frac{1}{G_{kk}}\Big| \prec (1+|z|)^{-1}\,\Phi^2 + \frac{1}{N}\,.$$

In the setting where $\mathbf{H}$ is a generalised Wigner matrix and $|z| \le 10$ this bound is precisely the content of Theorem 4.7 from [26].

The a priori bound used in the proof of that theorem is replaced by

$$\Big|Q_k \frac{1}{G_{kk}^{(V)}}\Big| \prec \Lambda_{\mathrm{o}} + \frac{1}{\sqrt{N}}\,, \tag{17.17}$$

for any $V \subseteq \{1,\dots,N\}$ with N-independent size. This bound is proven in the same way as (16.19). Here, the N_0 hidden in the stochastic domination depends on the size $|V|$ of the index set. Following the proof of Theorem 4.7 given in [26] with (17.17) and tracking the z-dependence,

$$\frac{1}{|G_{kk}^{(V)}(z)|} \prec 1 + |z|\,,$$

yields the fluctuation averaging, Theorem 17.5.

□

PROOF OF PROPOSITION 17.1. Let us show first that (17.3) follows directly from (17.2) by applying the fluctuation averaging, Theorem 17.5. Indeed, (17.2) provides a deterministic bound on the off-diagonal error, Λ_{o}, which is needed to apply the fluctuation averaging to the right hand side of (17.14). It also shows that the indicator functions on the left hand side of (17.14) and (17.9) are a.w.o.p. nonzero. Thus, (17.3) is proven, provided (17.2) is true.

The proof of (17.2) is split into the consideration of two different regimes. In the first regime the absolute value of z is large, $|z| \ge N^5$. In this case we make use only of weak a priori bounds on the resolvent elements and the entries of $\mathbf{d}$. Together with Lemma 17.3 they will suffice to prove (17.2) in this case. In the second regime, $|z| \le N^5$, we use a continuity argument. We will establish a gap in the possible values that the continuous function, $z \mapsto (1+|z|)\Lambda(z)$, might have. Here, the stability result Corollary 17.4 is used. We use this gap to propagate the bound with the help of Lemma C.1 in the appendix from $|z| = N^5$ to the whole domain where $|z| \le N^5$, $\eta \ge N^{\gamma-1}$ and we stay away from $\mathbb{M}$.

REGIME 1: Let $|z| \ge N^5$. We show that the indicator functions in the statement of Lemma 16.1 are a.w.o.p. not vanishing. We start by showing that the diagonal contribution, Λ_{d}, to Λ is sufficiently small. The reduced resolvent elements for an arbitrary $V \subseteq \{1,\dots,N\}$ satisfy

$$|G_{ij}^{(V)}(z)| \le \eta^{-1} \le N^{1-\gamma}\,. \tag{17.18}$$

From this and the definition of $\mathbf{d}$ in (16.2) we read off the a priori bound,

$$\|\mathbf{d}(z)\|_\infty \prec N^{2-\gamma}. \tag{17.19}$$

Here, we used the general resolvent identity (16.9) in the form $G_{ik}G_{ki} = g_k(g_i - G_{ii}^{(k)})$. Since $\mathbf{g}$ satisfies the perturbed QVE (16.1) we conclude that uniformly for $|z| \ge N^2$ we have

$$|g_k(z)| \le 2\,|z|^{-1}, \qquad \text{a.w.o.p.}. \tag{17.20}$$

With the trivial bound $|m_i(z)| \leq 1/\mathrm{dist}(z, \operatorname{supp}\rho)$ on the solution of the QVE we infer that on this domain the indicator function in (17.8) is a.w.o.p. non-zero and therefore uniformly for $|z| \geq N^2$,

$$\Lambda_{\mathrm{d}}(z) \lesssim |z|^{-2}\,\|\mathbf{d}(z)\|_\infty \leq N^{-\gamma/2}\,|z|^{-1}, \qquad \text{a.w.o.p.}\,. \tag{17.21}$$

In the last inequality we used the bound on $\mathbf{d}$ from (17.19). Thus, for $|z| \geq N^2$ the diagonal contribution to Λ does not play a role in the indicator function in the statement of Lemma 16.1.

Now we derive a similar bound for the off-diagonal contribution Λ_{o}. Using the resolvent identity (16.9) for $i = j$ again and the a priori bound on the reduced resolvent elements, (17.18), in the expansion formula (16.3) yields

$$|G_{kl}(z)| \prec \big(|g_k(z)g_l(z)| + |G_{kl}(z)G_{lk}(z)|\big)\,N^{2-\gamma}\,, \qquad |G_{kl}(z)| \prec |g_k(z)|\,N^{3-\gamma}\,, \tag{17.22}$$

for $k \neq l$. With the bound (17.20) we conclude that

$$\Lambda_{\mathrm{o}}(z) \prec |z|^{-2}\,N^{2-\gamma} + |z|^{-1}\,N^{5-2\gamma}\Lambda_{\mathrm{o}}(z)\,, \qquad |z| \geq N^2\,. \tag{17.23}$$

Thus, $\Lambda_{\mathrm{o}} \prec N^{-3}|z|^{-1}$ on the domain where $|z| \geq N^5$. We conclude that Lemma 16.1 applies in this regime even without the indicator functions in the formulas (16.5). We use the bound from this lemma for the norm of $\mathbf{d}$ and the off-diagonal contribution, Λ_{o}, to Λ, while we use the first inequality in (17.21) for the diagonal contribution, Λ_{d}. In this way we get

$$\begin{aligned} |z|^2\Lambda_{\mathrm{d}} + \|\mathbf{d}\|_\infty &\prec |z|^{-2}\sqrt{\frac{\rho}{N\eta}} + |z|^{-2}\sqrt{\frac{\Lambda_{\mathrm{d}}}{N\eta}} + \frac{1}{\sqrt{N}}\,,\\ |z|^2\Lambda_{\mathrm{o}} &\prec \sqrt{\frac{\rho}{N\eta}} + \sqrt{\frac{\Lambda_{\mathrm{d}}}{N\eta}} + \frac{1}{\sqrt{N}}\,, \end{aligned} \tag{17.24}$$

where we also used $g_k = m_k + \mathcal{O}(\Lambda_{\mathrm{d}})$. We find for any $\varepsilon \in (0,\gamma)$ that the right hand side of the first inequality can be estimated further by

$$|z|^2\Lambda_{\mathrm{d}} + \|\mathbf{d}\|_\infty \prec |z|^{-2}\sqrt{\frac{\rho}{N\eta}} + N^{-\varepsilon}|z|^2\Lambda_{\mathrm{d}} + |z|^{-6}\frac{N^\varepsilon}{N\eta} + \frac{1}{\sqrt{N}}\,.$$

The term $N^{-\varepsilon}|z|^2\Lambda_{\mathrm{d}}$ can be absorbed into the left hand side and by the definition of the stochastic domination and since ε is arbitrarily small the remaining N^ε on the right hand side can be replaced by 1 without changing the correctness of this bound. In this way we arrive at

$$|z|^2\Lambda_{\mathrm{d}} + \|\mathbf{d}\|_\infty \prec |z|^{-2}\sqrt{\frac{\rho}{N\eta}} + |z|^{-6}\frac{1}{N\eta} + \frac{1}{\sqrt{N}}\,.$$

For the bound on the off-diagonal error term we plug this result into (17.24) and get

$$\Lambda_{\mathrm{o}} \prec |z|^{-2}\sqrt{\frac{\rho}{N\eta}} + |z|^{-6}\frac{1}{N\eta} + |z|^{-3}\frac{1}{N^{1/4}}\sqrt{\frac{1}{N\eta}} + |z|^{-2}\frac{1}{\sqrt{N}}\,.$$

Regime 2: Now let $|z| \leq N^5$ and suppose that δ_* is a positive constant, depending only on the model parameters p, P and L. We start by establishing a gap in the possible values of Λ. The diagonal contribution, Λ_{d}, satisfies

$$\Lambda_{\mathrm{d}}(z)\,\mathbb{1}\Big(\Lambda_{\mathrm{d}}(z) \leq \frac{\lambda_*}{1+|z|}\Big) \lesssim \frac{\|\mathbf{d}(z)\|_\infty}{1+|z|^2}\,, \tag{17.25}$$

according to (17.8) in Lemma 17.3 (for $|z| \geq 10$) and (17.13) from Corollary 17.4 (for $|z| \leq 10$), where λ_* is a sufficiently small positive constant.

We estimate the norm of $\mathbf{d}$ by Lemma 16.1 and also use the the bound on the off-diagonal contribution, Λ_{o}, from the same lemma,

$$
\begin{aligned}
\big((1+|z|^2)\Lambda_{\mathrm{d}} + \|\mathbf{d}\|_\infty\big)\,\mathbb{1}\Big(\Lambda \leq \frac{\lambda_*}{1+|z|}\Big) &\prec (1+|z|)^{-2}\sqrt{\frac{\operatorname{Im}\langle \mathbf{g}\rangle}{N\eta}} + \frac{1}{\sqrt{N}}\,,\\
(1+|z|^2)\Lambda_{\mathrm{o}}\,\mathbb{1}\Big(\Lambda \leq \frac{\lambda_*}{1+|z|}\Big) &\prec \sqrt{\frac{\operatorname{Im}\langle \mathbf{g}\rangle}{N\eta}} + \frac{1}{\sqrt{N}}\,.
\end{aligned}
\tag{17.26}
$$

Now we use $g_k = m_k + \mathcal{O}(\Lambda_{\mathrm{d}})$ and that $\operatorname{Im}\langle \mathbf{m}(z)\rangle = \pi\rho(z)$. With these two ingredients we find for any $\varepsilon \in (0,\gamma)$ that

$$
\begin{aligned}
&\big((1+|z|)^2\Lambda_{\mathrm{d}} + \|\mathbf{d}\|_\infty\big)\,\mathbb{1}\Big(\Lambda \leq \frac{\lambda_*}{1+|z|}\Big)\\
&\prec (1+|z|)^{-2}\sqrt{\frac{\rho}{N\eta}} + (1+|z|)^{-6}\frac{N^\varepsilon}{N\eta} + \frac{1}{\sqrt{N}} + N^{-\varepsilon}(1+|z|)^2\Lambda_{\mathrm{d}}\,.
\end{aligned}
\tag{17.27}
$$

The term $N^{-\varepsilon}(1+|z|)^2\Lambda_{\mathrm{d}}$ can be absorbed into the left hand side and we arrive at

$$
\big((1+|z|)^2\Lambda_{\mathrm{d}} + \|\mathbf{d}\|_\infty\big)\,\mathbb{1}\Big(\Lambda \leq \frac{\lambda_*}{1+|z|}\Big) \prec (1+|z|)^{-2}\sqrt{\frac{\rho}{N\eta}} + (1+|z|)^{-6}\frac{1}{N\eta} + \frac{1}{\sqrt{N}}\,. \tag{17.28}
$$

For the off-diagonal error terms we plug this into the second bound of (17.26) after using $\operatorname{Im}\langle \mathbf{g}\rangle \lesssim \rho + \Lambda_{\mathrm{d}}$ and get

$$
\Lambda_{\mathrm{o}} \prec (1+|z|)^{-2}\sqrt{\frac{\rho}{N\eta}} + (1+|z|)^{-6}\frac{1}{N\eta} + (1+|z|)^{-3}\frac{1}{N^{1/4}}\sqrt{\frac{1}{N\eta}} + (1+|z|)^{-2}\frac{1}{\sqrt{N}}\,. \tag{17.29}
$$

In particular, we combine (17.28) and (17.29) to establish a gap in the values that Λ can take,

$$
\Lambda\,\mathbb{1}\Big(\Lambda \leq \frac{\lambda_*}{1+|z|}\Big) \prec \frac{N^{-\gamma/2}}{1+|z|}\,. \tag{17.30}
$$

Here we used $\eta \geq N^{\gamma-1}$.

Now we apply Lemma C.1 on the connected domain

$$
\{z \in \mathbb{H} : \operatorname{Im} z \geq N^{\gamma-1},\, \operatorname{dist}(z,\mathbb{M}) \geq \delta_*,\, |z| \leq N^5\}\,,
$$

with the choices

$$
\varphi(z) := (1+|z|)\Lambda(z)\,, \qquad \Phi(z) := N^{-\gamma/3}\,, \qquad z_0 := \mathrm{i}N^5\,. \tag{17.31}
$$

The continuity condition (C.1) of the lemma for these two functions follows from the Hölder-continuity, (17.5), of the solution of the QVE and the weak continuity of the resolvent elements,

$$
|G_{ij}(z_1) - G_{ij}(z_2)| \leq \frac{|z_1 - z_2|}{(\operatorname{Im} z_1)(\operatorname{Im} z_2)} \leq N^2|z_1 - z_2|\,. \tag{17.32}
$$

The condition (C.3) holds since by (17.2) on the first regime we have a.w.o.p. $\varphi(z_0) \leq \Phi(z_0)$. Finally, (17.30) implies a.w.o.p. $\varphi\,\mathbb{1}(\varphi \in [\Phi - N^{-1}, \Phi]) < \Phi - N^{-1}$ and thus (C.2). We infer that a.w.o.p. $\varphi \leq \Phi$. In particular, the indicator function in (17.28) and (17.29) is non-zero a.w.o.p.. Thus, (17.28) and (17.29) imply (17.2) in the second regime. □

18 Local law close to local minima

18.1 The solution of the QVE

In this section we state a few facts about the solution $\mathbf{m}$ of the QVE (15.3) and about the stability of this equation against perturbations. These facts are summarized in two theorems that are taken from Part II of this work. The first theorem contains regularity properties of $\mathbf{m}$. Furthermore, it provides lower and upper bounds on the imaginary part, $\mathrm{Im}\langle\mathbf{m}\rangle = \pi\,\rho$, by explicit functions. It is a combination of the statements from Theorem 1.1, Theorem 6.1, Theorem 6.2 and Corollary B.1 of Part II.

THEOREM 18.1 (Solution of the QVE). *Let the sequence $\mathbf{S} = \mathbf{S}^{(N)}$ satisfy the assumptions (A)-(C). Then for every component, $m_i : \mathbb{H} \to \mathbb{H}$, of the unique solution, $\mathbf{m} = (m_1, \dots, m_N)$, of the QVE there is a probability density $v_i : \mathbb{R} \to [0,\infty)$ with support in the interval $[-2,2]$, such that*

$$m_i(z) \,=\, \int_{\mathbb{R}} \frac{v_i(\tau)\mathrm{d}\tau}{\tau - z}\,, \qquad z \in \mathbb{H}\,,\; i = 1,\dots,N\,. \tag{18.1}$$

The probability densities are comparable,

$$v_i(\tau) \,\sim\, v_j(\tau)\,, \qquad \tau \in \mathbb{R}\,,\; i,j = 1,\dots,N\,. \tag{18.2}$$

The solution $\mathbf{m}$ has a uniformly Hölder-continuous extension (denoted again by $\mathbf{m}$) to the closed complex upper half plane $\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{R}$,

$$\|\mathbf{m}(z_1) - \mathbf{m}(z_2)\|_\infty \,\lesssim\, |z_1 - z_2|^{1/3}\,, \qquad z_1, z_2 \in \overline{\mathbb{H}}\,. \tag{18.3}$$

Its absolute value satisfies

$$|m_i(z)| \,\sim\, \frac{1}{1+|z|}\,, \qquad z \in \overline{\mathbb{H}}\,,\; i = 1,\dots,N\,.$$

Let $\rho : \mathbb{R} \to [0,\infty), \tau \mapsto \langle\mathbf{v}(\tau)\rangle$ be the density of states, defined in (15.8). *Then there exists a positive constant δ_*, depending only on the model parameters p, P and L, such that the following holds true. The support of the density consists of $K \sim 1$ disjoint intervals of lengths at least $2\delta_*$, i.e.,*

$$\operatorname{supp}\rho \,=\, \bigcup_{i=1}^{K} [\alpha_i, \beta_i]\,, \quad \textit{where} \quad \beta_i - \alpha_i \geq 2\delta_*\,, \quad \textit{and} \quad \alpha_i < \beta_i < \alpha_{i+1}\,. \tag{18.4}$$

The size of the harmonic extension, ρ, up to constant factors, is given by explicit functions as follows. Let $\eta \in [0, \delta_*]$.

- **Bulk:** *Close to the support of the density of states but away from the local minima in $\mathbb{M}$ (cf.* (17.1)*) the function ρ is comparable to* 1, *i.e.,*

$$\rho(\tau + \mathrm{i}\eta) \,\sim\, 1\,, \qquad \tau \in \operatorname{supp}\rho\,,\; \operatorname{dist}(\tau, \mathbb{M}) \geq \delta_*\,. \tag{18.5a}$$

- **At an internal edge:** *At the edges α_i, β_{i-1} with $i = 2,\dots,K$ in the direction where the support of the density of states continues the size of ρ is*

$$\rho(\alpha_i + \omega + \mathrm{i}\eta) \,\sim\, \rho(\beta_{i-1} - \omega + \mathrm{i}\eta) \,\sim\, \frac{(\omega+\eta)^{1/2}}{(\alpha_i - \beta_{i-1} + \omega + \eta)^{1/6}}\,, \qquad \omega \in [0, \delta_*]\,. \tag{18.5b}$$

- **Inside a gap:** *Between two neighbouring edges β_{i-1} and α_i with $i = 2, \dots, K$, the function ρ satisfies*

$$\rho(\beta_{i-1} + \omega + \mathrm{i}\eta) \sim \rho(\alpha_i - \omega + \mathrm{i}\eta) \sim \frac{\eta}{(\alpha_i - \beta_{i-1} + \eta)^{1/6}(\omega + \eta)^{1/2}}, \tag{18.5c}$$

for all $\omega \in [0, (\alpha_i - \beta_{i-1})/2]$.

- **Around an extreme edge:** *At the extreme points α_1 and β_K of* $\operatorname{supp}\rho$ *the density of states grows like a square root ,*

$$\rho(\alpha_1 + \omega + \mathrm{i}\eta) \sim \rho(\beta_K - \omega + \mathrm{i}\eta) \sim \begin{cases} (\omega + \eta)^{1/2}, & \omega \in [0, \delta_*], \\ \dfrac{\eta}{(|\omega| + \eta)^{1/2}}, & \omega \in [-\delta_*, 0]. \end{cases} \tag{18.5d}$$

- **Close to a local minimum:** *In a neighbourhood of a local minimum in the interior of the support of the density of states, i.e., for $\tau_0 \in \mathbb{M} \cap \operatorname{int}\operatorname{supp}\rho$, we have*

$$\rho(\tau_0 + \omega + \mathrm{i}\eta) \sim \rho(\tau_0) + (|\omega| + \eta)^{1/3}, \qquad \omega \in [-\delta_*, \delta_*]. \tag{18.5e}$$

- **Away from the support:** *Away from the interval in which* $\operatorname{supp}\rho$ *is contained*

$$\rho(z) \sim \frac{\operatorname{Im} z}{|z|^2}, \qquad z \in \overline{\mathbb{H}}, \ \operatorname{dist}(z, [\alpha_1, \beta_K]) \geq \delta_*. \tag{18.5f}$$

The next theorem shows that the QVE is stable under small perturbations, $\mathbf{d}$, in the sense that once a solution of the perturbed QVE (18.6) is sufficiently close to $\mathbf{m}$, then the difference between the two can be estimated in terms of $\|\mathbf{d}\|_\infty$. In Part II it is stated as Proposition 13.1.

THEOREM 18.2 (Refined stability). *There are a scalar function $\sigma : \overline{\mathbb{H}} \to [0, \infty)$, three vector valued functions $\mathbf{s}, \mathbf{t}^{(1)}, \mathbf{t}^{(2)} : \overline{\mathbb{H}} \to \mathbb{C}^N$, a matrix valued function $\mathbf{T} : \overline{\mathbb{H}} \to \mathbb{C}^{N \times N}$, all depending only on $\mathbf{S}$, and a positive constant λ_*, depending only on the model parameters p, P and L, such that for two arbitrary vector valued functions $\mathbf{d} : \mathbb{H} \to \mathbb{C}^N$ and $\mathbf{g} : \mathbb{H} \to (\mathbb{C} \setminus \{0\})^N$ that satisfy*

$$-\frac{1}{g_i(z)} = z + \sum_{j=1}^N s_{ij} g_j(z) + d_i(z), \quad z \in \mathbb{H}, \tag{18.6}$$

the difference between $\mathbf{g}$ and $\mathbf{m}$ is bounded in terms of

$$\Theta(z) := |\langle \mathbf{s}(z), \mathbf{g}(z) - \mathbf{m}(z) \rangle|, \qquad z \in \mathbb{H}, \tag{18.7}$$

in the following two ways. For all $\mathbf{w} \in \mathbb{C}^N$ on the whole complex upper half plane:

$$\|\mathbf{g} - \mathbf{m}\|_\infty \mathbb{1}\big(\|\mathbf{g} - \mathbf{m}\|_\infty \leq \lambda_*\big) \lesssim \Theta + \|\mathbf{d}\|_\infty, \tag{18.8a}$$

$$|\langle \mathbf{w}, \mathbf{g} - \mathbf{m} \rangle| \mathbb{1}\big(\|\mathbf{g} - \mathbf{m}\|_\infty \leq \lambda_*\big) \lesssim \|\mathbf{w}\|_\infty \Theta + \|\mathbf{w}\|_\infty \|\mathbf{d}\|_\infty^2 + |\langle \mathbf{T}\mathbf{w}, \mathbf{d} \rangle|. \tag{18.8b}$$

The scalar function Θ satisfies a cubic equation of the form

$$\big|\Theta^3 + \pi_2 \Theta^2 + \pi_1 \Theta\big| \mathbb{1}\big(\|\mathbf{g} - \mathbf{m}\|_\infty \leq \lambda_*\big) \lesssim \|\mathbf{d}\|_\infty^2 + |\langle \mathbf{t}^{(1)}, \mathbf{d} \rangle| + |\langle \mathbf{t}^{(2)}, \mathbf{d} \rangle|. \tag{18.9}$$

The coefficients $\pi_1, \pi_2 : \mathbb{H} \to \mathbb{C}$ may depend on S and $\mathbf{g}$, but they are bounded in terms of $\mathbf{g}$-independent functions,

$$\begin{aligned} |\pi_1(z)| &\sim \frac{\operatorname{Im} z}{\rho(z)} + \rho(z)\,(\rho(z) + \sigma(z))\,, \\ |\pi_2(z)| &\sim \rho(z) + \sigma(z)\,, \end{aligned} \tag{18.10}$$

for all $z \in \mathbb{H}$. Moreover, the functions σ, $\mathbf{s}$, $\mathbf{t}^{(1)}$, $\mathbf{t}^{(2)}$ and $\mathbf{T}$ have the following additional properties:

$$|\sigma(z_1) - \sigma(z_2)| + \|\mathbf{s}(z_1) - \mathbf{s}(z_2)\| \lesssim |z_1 - z_2|^{1/3}\,, \qquad z_1, z_2 \in \overline{\mathbb{H}}\,, \tag{18.11a}$$

$$|\sigma(z)| + \|\mathbf{s}(z)\|_\infty + \|\mathbf{t}^{(1)}(z)\|_\infty + \|\mathbf{t}^{(2)}(z)\|_\infty + \|\mathbf{T}(z)\|_{\infty\to\infty} \lesssim 1\,, \qquad z \in \overline{\mathbb{H}}\,. \tag{18.11b}$$

At the local minima of the density of states inside its own support, i.e. at the points in $\mathbb{M}$, the function σ satisfies

$$\begin{aligned} \sigma(\alpha_i) &\sim \sigma(\beta_{i-1}) \sim (\alpha_i - \beta_{i-1})^{1/3}\,, \qquad i = 2, \dots, K\,, \\ \sigma(\alpha_1) &\sim \sigma(\beta_K) \sim 1\,, \\ \sigma(\tau_0) &\lesssim \rho(\tau_0)^2\,, \qquad \tau_0 \in \mathbb{M} \setminus \{\alpha_i, \beta_i\}\,. \end{aligned} \tag{18.11c}$$

The function σ appears naturally in the analysis of the QVE. Analogous to the more explicitly constructed function Δ from Definition 15.4, at an edge the value of σ^3 encodes the size of the corresponding gap in $\operatorname{supp}\rho$. At the local minima in $\mathbb{M} \setminus \{\alpha_i, \beta_i\}$ the value of σ^3 is small, provided the density of states has a small value at the minimum. In this sense it is again analogous to Δ, which vanishes at these internal minima.

18.2 Coefficients of the cubic equation

The stability of QVE near the points in $\mathbb{M}$ requires a careful analysis of the cubic equation (18.9) for Θ from Theorem 18.2. For this, we will provide a more explicit description of the upper and lower bounds from (18.10) on the coefficients, π_1 and π_2, of the cubic equation.

Proposition 18.3 (Behaviour of the coefficients). *There are positive constants δ_* and c_* such that for all $\eta \in [0, \delta_*]$ the coefficients, π_1 and π_2, of the cubic equation* (18.9) *satisfy the following bounds.*

- **Around an internal edge:** *At the edges α_i, β_{i-1} of the gap with length $\Delta := \alpha_i - \beta_{i-1}$ for $i = 2, \dots, K$, we have*

$$\begin{aligned} &|\pi_1(\alpha_i + \omega + \mathrm{i}\eta)| \sim |\pi_1(\beta_{i-1} - \omega + \mathrm{i}\eta)| \sim (|\omega| + \eta)^{1/2}(|\omega| + \eta + \Delta)^{1/6}, \\ &|\pi_2(\alpha_i + \omega + \mathrm{i}\eta)| \sim |\pi_2(\beta_{i-1} - \omega + \mathrm{i}\eta)| \sim (|\omega| + \eta + \Delta)^{1/3}, \qquad \omega \in [-c_*\Delta, \delta_*]\,. \end{aligned} \tag{18.12a}$$

- **Well inside a gap:** *Between two neighbouring edges β_{i-1} and α_i of the gap with length $\Delta := \alpha_i - \beta_{i-1}$ for $i = 2, \dots, K$, the first coefficient, π_1, satisfies*

$$|\pi_1(\alpha_i - \omega + \mathrm{i}\eta)| \sim |\pi_1(\beta_{i-1} + \omega + \mathrm{i}\eta)| \sim (\eta + \Delta)^{2/3}\,, \qquad \omega \in \left[c_*\Delta, \frac{\Delta}{2}\right] \tag{18.12b}$$

The second coefficient, π_2, satisfies the upper bounds,

$$\begin{aligned} |\pi_2(\alpha_i - \omega + \mathrm{i}\eta)| &\lesssim (\eta + \Delta)^{1/3}\,, \\ |\pi_2(\beta_{i-1} + \omega + \mathrm{i}\eta)| &\lesssim (\eta + \Delta)^{1/3}\,, \end{aligned} \qquad \omega \in \left[c_*\Delta, \frac{\Delta}{2}\right]. \tag{18.12c}$$

- **Around an extreme edge:** *Around the extreme points α_1 and β_K of the support of the density of states*

$$\begin{aligned} |\pi_1(\alpha_1+\omega+\mathrm{i}\eta)| \,&\sim\, |\pi_1(\beta_K-\omega+\mathrm{i}\eta)| \,\sim\, (\omega+\eta)^{1/2} \\ |\pi_2(\alpha_1+\omega+\mathrm{i}\eta)| \,&\sim\, |\pi_2(\beta_K-\omega+\mathrm{i}\eta)| \,\sim\, 1\,, \end{aligned} \qquad \omega\in[-\delta_*,\delta_*]\,. \tag{18.12d}$$

- **Close to a local minimum:** *In a neighbourhood of the local minimum in the interior of the support of the density of states, i.e. for $\tau_0\in\mathbb{M}\cap\operatorname{int}\operatorname{supp}\rho$, we have*

$$\begin{aligned} |\pi_1(\tau_0+\omega+\mathrm{i}\eta)| \,&\sim\, \rho(\tau_0)^2+(|\omega|+\eta)^{2/3}\,, \\ |\pi_2(\tau_0+\omega+\mathrm{i}\eta)| \,&\sim\, \rho(\tau_0)+(|\omega|+\eta)^{1/3}\,, \end{aligned} \qquad \omega\in[-\delta_*,\,\delta_*]\,. \tag{18.12e}$$

Proof. The proof is split according to the cases above. In each case we combine the general formulas (18.10) with the knowledge about the harmonic extension, ρ, of the density of states from Theorem 18.1 and about the behaviour of the positive Hölder-continuous function, σ, at the minima in $\mathbb{M}$ from (18.11c). The positive constant δ_* is chosen to have at most the same value as in Theorem 18.1. We start with the simplest case.

Around an extreme edge: By the Hölder-continuity of σ (cf. (18.11a)) and because σ is comparable to 1 at the points α_1 and β_K (cf. (18.11c)), this function is comparable to 1 in the whole δ_*-neighbourhood of the extreme edges. Thus, using (18.10) inside this neighbourhood, we find

$$|\pi_1(z)|\,\sim\,\frac{\operatorname{Im}z}{\rho(z)}+\rho(z)\,,\qquad |\pi_2(z)|\,\sim\,1\,.$$

The claim now follows from the behaviour of ρ, given in Theorem 18.1, inside this domain.

Close to a local minimum: In this case $\rho+\sigma$ is comparable to ρ. In fact, using the 1/3-Hölder-continuity of σ (cf. (18.11a)) and its bound at the minimum, $\tau_0\in\mathbb{M}$, (cf. (18.11c)) we find

$$\rho(z)\,\le\,\rho(z)+\sigma(z)\,\lesssim\,\rho(z)+\rho(\tau_0)^2+|z-\tau_0|^{1/3}\,\sim\,\rho(z)\,,\qquad |z-\tau_0|\le\delta_*\,. \tag{18.13}$$

In the last relation we used the behaviour (18.5e) of ρ from Theorem 18.1. By (18.10) we conclude that inside the δ_*-neighbourhood of τ_0,

$$|\pi_1(z)|\,\sim\,\frac{\operatorname{Im}z}{\rho(z)}+\rho(z)^2\,,\qquad |\pi_2(z)|\,\sim\,\rho(z)\,. \tag{18.14}$$

Using the upper and lower bounds on $\rho(z)$ again, gives the desired result, (18.12e).

Around an internal edge: First we prove the bounds on $|\pi_2|$, starting from (18.10). The upper bound simply uses the 1/3-Hölder-continuity and the behaviour at the edge points of σ,

$$|\pi_2(z)|\,\sim\,\rho(z)+\sigma(z)\,\lesssim\,\rho(z)+\Delta^{1/3}+|z-\tau_0|^{1/3}, \tag{18.15}$$

where τ_0 is one of the edge points α_i or β_{i-1}. The claim follows from plugging in the size of ρ from the two corresponding domains in Theorem 18.1, i.e., the domain close to an edge, (18.5b), and the domain inside a gap, (18.5c).

For the lower bound we consider two different regimes. In the first case z is close to the edge point, $|z-\tau_0|\le c\Delta$, for some small positive constant c, depending only on the model parameters p, P and L. We find

$$|\pi_2(z)|\,\sim\,\rho(z)+\sigma(z)\,\gtrsim\,\rho(z)+\Delta^{1/3}-C|z-\tau_0|^{1/3}\,\sim\,\rho(z)+\Delta^{1/3}\,,$$

provided c is small enough. This bound coincides with the lower bound on π_2 in (18.12a), once the size of ρ from (18.5b) is used.

In the second regime, $|z-\tau_0| \geq c\Delta$, we simply use $|\pi_2(z)| \gtrsim \rho(z)$ from (18.10). If $\mathrm{Re}\, z \in \operatorname{supp}\rho$, then the size of ρ from (18.5b) yields the desired lower bound. If, on the other hand, $\mathrm{Re}\, z$ lies inside a gap of $\operatorname{supp}\rho$, then we use the freedom of choosing the constant c_* in Proposition 18.3. Suppose $c_* \leq c/2$. Then $|z-\tau_0| \geq c\Delta$ and $|\mathrm{Re}\, z - \tau_0| \leq c_*\Delta$ imply $\mathrm{Im}\, z \gtrsim \Delta$ and

$$\rho(z) \sim (\mathrm{Im}\, z)^{1/3} \gtrsim \Delta^{1/3} + |z-\tau_0|^{1/3}.$$

This finishes the proof of the upper and lower bound on $|\pi_2|$ on this domain. For the claim about $|\pi_1|$ we plug the result about $|\pi_2|$ and the size of ρ into

$$|\pi_1| \sim \frac{\mathrm{Im}\, z}{\rho(z)} + \rho(z)|\pi_2(z)|\,. \tag{18.16}$$

Well inside a gap: For the upper bound on $|\pi_2|$ we simply use (18.15) again, which follows from (18.11a) and (18.11c). The comparison relation for $|\pi_1|$ now follows from (18.16) again. For the lower bound, $|\pi_1| \gtrsim \mathrm{Im}\, z/\rho$ and (18.5c) from Theorem 18.1 are sufficient. This finishes the proof of the proposition. $\square$

18.3 Rough bound on Λ close to local minima

In the following lemma we will see that a.w.o.p. $\Lambda \leq c$ for some arbitrarily small constant $c>0$. Since the local law away from $\mathbb{M}$ is already shown in Proposition 17.1 we may restrict to bounded z in the following. From here on until the end of Section 18 we assume $|z| \leq 10$.

Lemma 18.4 (Rough bound). *Let λ_* be a positive constant. Then, uniformly for all $z=\tau+\mathrm{i}\eta \in \mathbb{H}$ with $\eta \geq N^{\gamma-1}$, the function Λ is uniformly small,*

$$\Lambda(z) \leq \lambda_* \qquad a.w.o.p.\,. \tag{18.17}$$

Proof. Away from the local minima in $\mathbb{M}$ the claim follows from (17.2) in Proposition 17.1. We will therefore prove that Λ is smaller than any fixed positive constant in some δ-neighbourhood of $\mathbb{M}$. We will use the freedom to choose the size δ of these neighbourhoods as small as we like.

Let us sketch the upcoming argument. Close to the points in $\mathbb{M}$ we make use of Theorem 18.2. Using Lemma 16.1, we will see that the right hand side of the cubic equation in Θ, (18.9), is smaller than a small negative power, $N^{-\varepsilon}$, of N, provided Λ is bounded by a small constant, $\Lambda \leq \lambda_*$. This will imply that Θ itself is small and through (18.8a) that the bound on Λ can be improved to $\Lambda \leq \lambda_*/2$. In this way we establish a gap in the possible values that the continuous function Λ can take. Lemma C.1 in the appendix is then used to propagate the bound on Λ from Proposition 17.1 into the δ-neighbourhoods of the points in $\mathbb{M}$.

Now we start the detailed proof from the fact that Θ satisfies the cubic equation (18.9), whose right hand side is bounded by $C\|\mathbf{d}\|_\infty$ for some constant C, depending only on the model parameters. Note that $\|\mathbf{d}\|_\infty \lesssim 1$ as long as $\Lambda \leq \lambda_*$ because in this case $|m_i| \sim 1$, $|g_i| \sim 1$ and $\mathbf{g}$ satisfies the perturbed QVE with perturbation $\mathbf{d}$. From the definition of Θ in (18.7) and the uniform bound on $\mathbf{s}$ from (18.11b), we get $\Theta \lesssim \Lambda$. Since the coefficient $|\pi_2|$ is uniformly bounded (cf. (18.10)), the cubic equation for Θ implies the three bounds

$$\Theta\,\mathbb{1}(\Lambda \leq \varepsilon_1, |\pi_1| \geq C_1\varepsilon_1) \lesssim \frac{\|\mathbf{d}\|_\infty}{|\pi_1|}\,, \tag{18.18a}$$

$$\Theta\,\mathbb{1}(\Lambda \leq \varepsilon_2, |\pi_2| \geq C_2\varepsilon_2) \lesssim \frac{|\pi_1|}{|\pi_2|} + \frac{\|\mathbf{d}\|_\infty^{1/2}}{|\pi_2|^{1/2}}\,, \tag{18.18b}$$

$$\Theta\,\mathbb{1}(\Lambda \le \lambda_*) \;\lesssim\; |\pi_2| + \sqrt{|\pi_1|} + \|\mathbf{d}\|_\infty^{1/3}. \tag{18.18c}$$

Here, $\varepsilon_1, \varepsilon_2 \in (0,1)$ are arbitrary constants and $C_1, C_2 > 0$ depend only on the model parameters.

Let $\delta \in (0,1)$ be another constant, which we will later fix later to only depend on the model parameters p, P, and L. We split $\mathbb{M}$ into four subsets, which are treated separately,

$$\mathbb{M}_1(\delta) \;:=\; \{\tau_0 \in \mathbb{M}\setminus \partial\operatorname{supp}\rho :\; \rho(\tau_0) > \delta^{1/3}\}\,, \quad \mathbb{M}_2(\delta) \;:=\; \{\tau_0 \in \partial\operatorname{supp}\rho :\; \Delta(\tau_0) > \delta^{1/2}\}\,,$$
$$\mathbb{M}_3(\delta) \;:=\; \{\tau_0 \in \mathbb{M}\setminus \partial\operatorname{supp}\rho :\; \rho(\tau_0) \le \delta^{1/3}\}\,, \quad \mathbb{M}_4(\delta) \;:=\; \{\tau_0 \in \partial\operatorname{supp}\rho :\; \Delta(\tau_0) \le \delta^{1/2}\}\,.$$

The function Δ here is taken from Definition 15.4 and its value is simply the length of the gap at the point $\tau_0 \in \partial\operatorname{supp}\rho$ where it is evaluated. We also define the δ-neighbourhoods of these subsets,

$$\mathbb{D}_k(\delta) \;:=\; \{z \in \mathbb{H} :\; \operatorname{dist}(z, \mathbb{M}_k(\delta)) \le \delta\}\,, \qquad k = 1,2,3,4\,.$$

As an immediate consequence of the upper and lower bounds on the coefficients, π_1 and π_2, presented in Proposition 18.3, we see that

$$|\pi_1(z)| \gtrsim \delta^{2/3}\,, \qquad\qquad z \in \mathbb{D}_1(\delta)\,, \tag{18.19a}$$
$$|\pi_1(z)| \lesssim \delta^{1/2}\,, \qquad |\pi_2(z)| \gtrsim \delta^{1/6}\,, \qquad z \in \mathbb{D}_2(\delta)\,, \tag{18.19b}$$
$$|\pi_1(z)| \lesssim \delta^{1/2}\,, \qquad |\pi_2(z)| \lesssim \delta^{1/6}\,, \qquad z \in \mathbb{D}_3(\delta)\cup\mathbb{D}_4(\delta)\,. \tag{18.19c}$$

On $\mathbb{D}_2(\delta)$ only the regimes around an internal edge, (18.12a), and around an extreme edge, (18.12d), are relevant. The case well inside the gap, (18.12b) and (18.12c), does not apply for small enough δ, since $\Delta(\tau_0) > \delta^{1/2}$ but $|z-\tau_0| \le \delta$.

Now we make a choice for the two constants ε_1 and ε_2. We express them in terms of δ as

$$\varepsilon_1 \;:=\; \delta\,, \qquad \varepsilon_2 \;:=\; \delta^{1/5}\,.$$

We pair the bounds on Θ from (18.18) with the corresponding bounds from (18.19) on the coefficients of the cubic equation. For small enough δ the conditions on π_1 in (18.18a) and π_2 in (18.18b) are automatically satisfied by the choice of ε_1 and ε_2, as well as the upper and lower bounds from (18.19a) and (18.19b). Thus, for small enough δ we end up with

$$\Theta(z)\,\mathbb{1}(\Lambda(z) \le \delta) \;\lesssim\; \delta^{-2/3}\|\mathbf{d}(z)\|_\infty\,, \qquad z \in \mathbb{D}_1(\delta)\,,$$
$$\Theta(z)\,\mathbb{1}(\Lambda(z) \le \delta^{1/5}) \;\lesssim\; \delta^{1/3} + \delta^{-1/12}\|\mathbf{d}(z)\|_\infty^{1/2}\,, \qquad z \in \mathbb{D}_2(\delta)\,,$$
$$\Theta(z)\,\mathbb{1}(\Lambda(z) \le \lambda_*) \;\lesssim\; \delta^{1/6} + \|\mathbf{d}(z)\|_\infty^{1/3}\,, \qquad z \in \mathbb{D}_3(\delta)\cup\mathbb{D}_4(\delta)\,.$$

At this stage we use Lemma 16.1 in the form of $\|\mathbf{d}\|_\infty \prec N^{-\gamma/2}$ on the set where $\Lambda \le \lambda_*/10$, say, and (18.8a) from Theorem 18.2. We may choose λ_* to be sufficiently small compared to the constants with the same name from these two statements. Furthermore, we choose δ so small that $\delta^{1/5} \le \lambda_*$. On the different regimes $\mathbb{D}_k(\delta)$, we find that

$$\Lambda(z)\,\mathbb{1}(\Lambda(z) \le \delta) \;\lesssim\; \delta^{-2/3}N^{-\gamma/3}\,, \qquad z \in \mathbb{D}_1(\delta)\,, \tag{18.21a}$$
$$\text{a.w.o.p.} \qquad \Lambda(z)\,\mathbb{1}(\Lambda(z) \le \delta^{1/5}) \;\lesssim\; \delta^{1/3} + \delta^{-1/12}N^{-\gamma/5}\,, \qquad z \in \mathbb{D}_2(\delta)\,, \tag{18.21b}$$
$$\Lambda(z)\,\mathbb{1}(\Lambda(z) \le \lambda_*) \;\lesssim\; \delta^{1/6} + N^{-\gamma/7}\,, \qquad z \in \mathbb{D}_3(\delta)\cup\mathbb{D}_4(\delta)\,. \tag{18.21c}$$

The right hand sides, including the constants from the comparison relation, can be made smaller than any given constant λ_* by choosing $\delta = \delta_*$, depending only on the model parameters, small enough and N sufficiently large. Furthermore, (18.21) establish a gap in the possible values that Λ can take on the δ_*-neighbourhood of any point in $\mathbb{M}$. By Proposition 17.1 we have the bound $\Lambda \prec N^{-\gamma/2}$ outside these δ_*-neighbourhoods and thus also for at least one point in the boundary of each neighbourhood. Now we apply Lemma C.1 to each neighbourhood and in this way we propagate the bound $\Lambda \le \lambda_*$ to every point z in the δ_*-neighbourhood of $\mathbb{M}$ with $\operatorname{Im} z \ge N^{\gamma-1}$. $\square$

18.4 Proof of Theorem 15.6

According to Proposition 17.1 the local law, Theorem 15.6, holds outside the δ_*-neighbourhoods of the points in $\mathbb{M}$. It remains to show that it is true inside these neighbourhoods as well. From here on we assume that $z \in \mathbb{H}$ satisfies $\mathrm{dist}(z, \mathbb{M}) \leq \delta_*$ and $\mathrm{Im}\, z \geq N^{\gamma-1}$. Let $\tau_0 \in \mathbb{M}$ be one of the closest points to z in $\mathbb{M}$, i.e.,

$$|z - \tau_0| \,=\, \mathrm{dist}(z, \mathbb{M})\,.$$

When $\tau_0 \in \partial\, \mathrm{supp}\, \rho$ we denote by $\theta = \theta(\tau_0) \in \{\pm 1\}$ the direction that points towards the gap in $\mathrm{supp}\, \rho$ at τ_0. In case $\tau_0 \in \partial\, \mathrm{supp}\, \rho$ we make the arbitrary choice $\theta := +1$, i.e.,

$$\theta := \begin{cases} -1 & \text{if } \ \tau_0 \in \{\alpha_i\}\,, \\ +1 & \text{if } \ \tau_0 \in \{\beta_i\}\,, \\ +1 & \text{if } \ \tau_0 \in \mathbb{M} \setminus \partial\, \mathrm{supp}\, \rho\,. \end{cases}$$

For $z = \tau_0 + \theta\, \omega + \mathrm{i}\eta$ with $\eta \in (0, \delta_*]$ and $\omega \in [-\delta_*, \delta_*]$ we will then prove the local law in the form

$$\Lambda(z) \,\prec\, \sqrt{\frac{\rho(z)}{N\eta}} + \frac{1}{N\eta} + \mathcal{E}(\omega, \eta)\,, \tag{18.22a}$$

$$\big|\big\langle \mathbf{w}, \mathbf{g}(z) - \mathbf{m}(z) \big\rangle\big| \,\prec\, \mathcal{E}(\omega, \eta)\,, \tag{18.22b}$$

where the positive error function $\mathcal{E} : [-\delta_*, \delta_*] \times (0, \delta_*] \to (0, \infty)$ is given as the unique solution of an explicit cubic equation in (18.26) below.

To define $\mathcal{E}$ we introduce explicit auxiliary functions $\widetilde{\pi}_1$, $\widetilde{\pi}_2$ and $\widetilde{\rho}$ that are comparable in size to the corresponding functions π_1, π_2 and ρ. The reason for using these auxiliary quantities for the definition of $\mathcal{E}$ instead of the original ones is twofold. Firstly, in this way $\mathcal{E}$ will be an explicit function instead of one that is implicitly defined through the solution of the QVE. The function $\mathcal{E}$ is explicit in the sense that there is a formula for the solution of the cubic equation that defines it and the coefficients are given by the explicit functions $\widetilde{\pi}_1$, $\widetilde{\pi}_2$ and $\widetilde{\rho}$. Secondly, $\mathcal{E}$ will be monotonic of its second variable, η. This property will be used later. The definition of the three auxiliary functions will be different, depending on whether τ_0 is in the boundary of the support of the density of states or not.

- **Edge:** If $\tau_0 \in \partial\, \mathrm{supp}\, \rho$, i.e. τ_0 is an edge of a gap of size $\Delta := \Delta_0(\tau_0)$ in the support of the density of states or an extreme edge. Then we define the three explicit functions

$$\widetilde{\rho}(\omega, \eta) \quad := \quad \begin{cases} \dfrac{(|\omega| + \eta)^{1/2}}{(\Delta + |\omega| + \eta)^{1/6}}\,, & \omega \in \big[-\delta_*, 0\big]\,, \\ \dfrac{\eta}{(\Delta + \eta)^{1/6}(\omega + \eta)^{1/2}}\,, & \omega \in \big[0, c_*\Delta\big]\,, \\ \dfrac{\eta}{(\Delta + \eta)^{2/3}}\,, & \omega \in \Big[c_*\Delta, \dfrac{\Delta}{2}\Big]\,. \end{cases} \tag{18.23a}$$

$$\widetilde{\pi}_1(\omega, \eta) \quad := \quad \begin{cases} (|\omega| + \eta)^{1/2}(|\omega| + \eta + \Delta)^{1/6}\,, & \omega \in \big[-\delta_*, 0\big]\,, \\ (\omega + \eta)^{1/2}(\Delta + \eta)^{1/6}\,, & \omega \in \big[0, c_*\Delta\big]\,, \\ (\Delta + \eta)^{2/3}\,, & \omega \in \Big[c_*\Delta, \tfrac{\Delta}{2}\Big] \end{cases} \tag{18.23b}$$

$$\widetilde{\pi}_2(\omega, \eta) \quad := \quad \begin{cases} (|\omega| + \eta + \Delta)^{1/3}\,, & \omega \in \big[-\delta_*, 0\big]\,, \\ (\Delta + \eta)^{1/3}\,, & \omega \in \big[0, c_*\Delta\big]\,, \\ (\Delta + \eta)^{1/3}\,, & \omega \in \Big[c_*\Delta, \dfrac{\Delta}{2}\Big] \end{cases} \tag{18.23c}$$

Here, c_* is the constant from Proposition 18.3.

- **Internal minimum:** If $\tau_0 \in \mathbb{M} \setminus \partial \operatorname{supp} \rho$, then we define for $\omega \in [-\delta_*, \delta_*]$ the three functions

$$\begin{aligned} \widetilde{\rho}(\omega,\eta) &:= \rho(\tau_0) + (|\omega|+\eta)^{1/3}, && (18.24\text{a}) \\ \widetilde{\pi}_1(\omega,\eta) &:= \rho(\tau_0)^2 + (|\omega|+\eta)^{2/3}, && (18.24\text{b}) \\ \widetilde{\pi}_2(\omega,\eta) &:= \rho(\tau_0) + (|\omega|+\eta)^{1/3}, && (18.24\text{c}) \end{aligned}$$

By design (cf. Proposition 18.3 and Theorem 18.1) these functions satisfy

$$\rho(\tau_0 + \theta\,\omega + \mathrm{i}\eta) \sim \widetilde{\rho}(\omega,\eta), \quad |\pi_k(\tau_0 + \theta\,\omega + \mathrm{i}\eta)| \sim \widetilde{\pi}_k(\omega,\eta), \tag{18.25}$$

except in one special case where the second bound does not hold, namely when $k = 2$, $\tau_0 \in \partial \operatorname{supp} \rho$ and $\omega \in [c_*\Delta, \Delta/2]$. In this case only the direction $|\pi_2| \lesssim \widetilde{\pi}_2$ is true (cf. (18.12c)).

We fix a positive constant $\widetilde{\varepsilon} \in (0, \gamma/16)$. The value of the function $\mathcal{E}$ at (ω, η) is then defined to be the unique positive solution of the cubic equation

$$\mathcal{E}(\omega,\eta)^3 + \widetilde{\pi}_2(\omega,\eta)\,\mathcal{E}(\omega,\eta)^2 + \widetilde{\pi}_1(\omega,\eta)\,\mathcal{E}(\omega,\eta) \;=\; N^{8\widetilde{\varepsilon}}\frac{\mathcal{E}(\omega,\eta)}{N\eta} + \frac{\widetilde{\rho}(\omega,\eta)}{N\eta} + \frac{1}{(N\eta)^2}\,. \tag{18.26}$$

We explain shortly what the relationship is between the $\mathcal{E}$ defined by (18.26) and the function κ that appears in the statement of Theorem 15.6. By the definition of $\widetilde{\pi}_1$, $\widetilde{\pi}_2$ and $\widetilde{\rho}$ we find a function κ, satisfying the bounds (15.19) and (15.21) from the statement of Theorem 15.6, such that

$$\mathcal{E} \;\leq\; N^{9\widetilde{\varepsilon}} \min\biggl\{\frac{1}{\sqrt{N\eta}}\,, \frac{\kappa}{N\eta}\biggr\}\,, \tag{18.27}$$

for any $N \geq N_0$, where the threshold N_0 here depends on $\widetilde{\varepsilon}$ in addition to p, P, L, $\underline{\mu}$ and γ. Indeed, we may simply define κ by the right hand side of (15.19) whenever (15.20) is not satisfied and by the right hand side of (15.21) whenever (15.20) is satisfied. The inequality (18.27) is verified by plugging its right hand side into (18.26) in place of $\mathcal{E}$ and checking that on each regime the resulting expression on the right hand side of (18.26) is smaller than the resulting expression on the left hand side of (18.26). The factor of $N^{9\widetilde{\varepsilon}}$ in (18.27) can be absorbed in the stochastic domination in (18.22). Thus (18.22) becomes (15.16) and (15.17) of Theorem 15.6.

Before we start the proof of the local law (18.22), let us motivate the definition of $\mathcal{E}$. As a consequence of Lemma 18.4 in the statement of Lemma 16.1 the indicator function is a.w.o.p. non-zero. Thus, uniformly in the δ_*-neighbourhood of τ_0 we have

$$\|\mathbf{d}\|_\infty + \Lambda_{\mathrm{o}} \;\prec\; \sqrt{\frac{\rho + |\langle \mathbf{g} - \mathbf{m}\rangle|}{N\eta}} + \frac{1}{N\eta}\,. \tag{18.28}$$

Here we used $\operatorname{Im}\langle \mathbf{g}\rangle \lesssim \rho + |\langle \mathbf{g} - \mathbf{m}\rangle|$ and $\rho \gtrsim \eta$. Since at the end the local law implies $|\langle \mathbf{g} - \mathbf{m}\rangle| \prec \mathcal{E}$, heuristically we may replace $|\langle \mathbf{g} - \mathbf{m}\rangle|$ in (18.28) by $\mathcal{E}$. In this case, from the fluctuation averaging, Theorem 17.5, we would be able to conclude that for any deterministic vector $\mathbf{w}$ with bounded entries,

$$\|\mathbf{d}\|_\infty^2 + |\langle \mathbf{w}, \mathbf{d}\rangle| \;\prec\; \frac{\mathcal{E}}{N\eta} + \frac{\rho}{N\eta} + \frac{1}{(N\eta)^2}\,. \tag{18.29}$$

Up to the technical factor of $N^{\gamma/4}$ the right hand side coincides with the right hand side of the cubic equation defining $\mathcal{E}$. On the other hand, the right hand side of the cubic equation (18.9) for the quantity Θ from Theorem 18.2 is of the same form as the left hand side of (18.29). Therefore, we infer

$$|\Theta^3 + \pi_2\Theta^2 + \pi_1\Theta| \prec \frac{\mathcal{E}}{N\eta} + \frac{\rho}{N\eta} + \frac{1}{(N\eta)^2}\,. \tag{18.30}$$

We will argue that on appropriately chosen domains out of the three summands in the cubic expression in Θ always one is the biggest by far. Therefore, the error function $\mathcal{E}$, defined by (18.26), is essentially the best bound on Θ that one may hope to deduce from (18.30). Indeed, since Θ is by definition an average of $\mathbf{g} - \mathbf{m}$, we expect $\Theta \prec \mathcal{E}$.

We will now prove (18.22). To this end we gradually improve the bound on Θ. The sequence of deterministic bounds on this quantity is defined as

$$\Phi_0 := 1\,, \qquad \Phi_{k+1} := \max\{ N^{-\varepsilon}\Phi_k\,,\, N^{9\varepsilon}\mathcal{E}\}\,, \tag{18.31}$$

where $\varepsilon \in (0, \widetilde{\varepsilon})$ is a fixed positive constant. From here on until the end of this section the threshold function N_0 from the definition of the stochastic domination (cf. Definition 15.5) as well as the definition of 'a.w.o.p.' (cf. Definition 15.7) may depend on ε in addition to p, P, L, μ and γ. At the end of the proof we will remove this dependence. The following lemma is essential for doing one step in the upcoming iteration.

LEMMA 18.5 (Improving bound through cubic). *Suppose that for all $z \in \tau_0 + [-\delta_*, \delta_*] + \mathrm{i}[N^{\gamma-1}, \delta_*]$ and some $k \in \mathbb{N}$ the quantity Θ fulfils*

$$\bigl|\Theta(z)^3 + \pi_2(z)\,\Theta(z)^2 + \pi_1(z)\,\Theta(z)\bigr| \prec \frac{\rho(z) + \Phi_k(\omega,\eta)}{N\eta} + \frac{1}{(N\eta)^2}\,. \tag{18.32}$$

Then $\Theta(z) \prec \Phi_{k+1}(\omega,\eta)$.

We will postpone the proof of this lemma until the end of this section. First we show how to use this result in the proof of the main theorem. Fix an integer $k \geq 0$ and assume that $\Theta + |\langle \mathbf{g} - \mathbf{m}\rangle| \prec \Phi_k$ is already proven. For $k = 0$ this follows from the rough bound on Λ in Lemma 18.4, $\Lambda \prec 1 = \Phi_0$. Then we see from (18.28) that

$$\|\mathbf{d}\|_\infty + \Lambda_{\mathrm{o}} \prec \sqrt{\frac{\rho + \Phi_k}{N\eta}} + \frac{1}{N\eta}\,. \tag{18.33}$$

The right hand side is a deterministic bound on the off-diagonal error Λ_{o}. We apply the fluctuation averaging, Theorem 17.5, to the right hand side of the cubic equation (18.9). In this way we see that the hypothesis (18.32) of Lemma 18.5 is satisfied. Therefore, the bound on Θ is improved to

$$\Theta(z) \prec \Phi_{k+1}(\omega,\eta)\,. \tag{18.34}$$

In order to improve the bound on $|\langle \mathbf{g} - \mathbf{m}\rangle|$ as well, we use the bound (18.8b) from Theorem 18.2 for averages of $\mathbf{g} - \mathbf{m}$ against bounded vectors. Since by Lemma 18.4 the deviation function Λ is bounded by a small constant, the indicator function in (18.8b) is a.w.o.p. non-zero. Choosing $\mathbf{w} = (1, \dots, 1)$, we find that

$$\bigl|\langle \mathbf{g} - \mathbf{m}\rangle\bigr| \lesssim \Theta + \|\mathbf{d}\|_\infty^2 + |\langle \widetilde{\mathbf{w}}, \mathbf{d}\rangle|\,, \qquad \text{a.w.o.p.}\,, \tag{18.35}$$

where $\widetilde{\mathbf{w}} = \mathbf{T}\mathbf{w}$ is a bounded, $\|\widetilde{\mathbf{w}}\|_\infty \lesssim 1$, deterministic vector. Together with the bound (18.33) we apply the fluctuation averaging again,

$$\bigl|\langle \mathbf{g} - \mathbf{m}\rangle\bigr| \prec \Phi_{k+1} + \frac{\rho + \Phi_k}{N\eta} + \frac{1}{(N\eta)^2} \lesssim N^{-\varepsilon}\Phi_k + \Phi_{k+1} \lesssim \Phi_{k+1}\,. \tag{18.36}$$

This concludes one step in the iteration, i.e., we have shown $\Theta + |\langle \mathbf{g} - \mathbf{m} \rangle| \prec \Phi_{k+1}$.

We repeat this step finitely many times and each time improve Φ_k by a factor of $N^{-\varepsilon}$ until it reaches its target value $N^{9\varepsilon}\mathcal{E}$ and is not improved anymore. At that stage we have

$$\Theta + |\langle \mathbf{g} - \mathbf{m} \rangle| \prec_\varepsilon N^{9\varepsilon}\mathcal{E} .$$

The subindex ε indicates that the threshold N_0 from the stochastic domination may depend on ε. But since ε was arbitrary, we infer $\Theta + |\langle \mathbf{g} - \mathbf{m} \rangle| \prec \mathcal{E}$, where now and until the start of the proof of Lemma 18.5 below the stochastic domination is ε-independent. By (18.28) we conclude

$$\|\mathbf{d}\|_\infty + \Lambda_{\mathrm{o}} \prec \sqrt{\frac{\rho}{N\eta}} + \frac{1}{N\eta} + \mathcal{E} . \tag{18.37}$$

For the bound on the diagonal contribution, Λ_{d}, we use (18.8a) to get

$$\Lambda_{\mathrm{d}} \lesssim \Theta + \|\mathbf{d}\|_\infty \prec \sqrt{\frac{\rho}{N\eta}} + \frac{1}{N\eta} + \mathcal{E} .$$

Finally, with the help of (18.8b), (18.37) and the fluctuation averaging, we prove the bound on averages of $\mathbf{g} - \mathbf{m}$ against any bounded, $\|\mathbf{w}\|_\infty \leq 1$, deterministic vector,

$$|\langle \mathbf{w}, \mathbf{g} - \mathbf{m} \rangle| \prec \frac{\rho}{N\eta} + \frac{1}{(N\eta)^2} + \Theta \prec \frac{\rho}{N\eta} + \frac{1}{(N\eta)^2} + \mathcal{E} .$$

This finishes the proof of Theorem 15.6 apart from the proof of Lemma 18.5 which we will tackle now.

$\square$

Proof of Lemma 18.5. The spectral parameter $z = \tau_0 + \theta\,\omega + \mathrm{i}\eta$ lies inside the δ_*-neighbourhood of τ_0. We fix $\omega \in [-\delta_*, \delta_*]$ and show that the claim holds for any choice of $\eta \in [N^{\gamma-1}, \delta_*]$ We split the interval of possible values of η into two or three regimes, depending on the case we are treating.

- **Edge:** If $\tau_0 \in \partial \operatorname{supp} \rho$ is an edge of a gap of size $\Delta := \Delta_0(\tau_0)$, then we define

$$\mathbb{D}_1(\omega) := \left\{ \eta \in [N^{\gamma-1}, \delta_*] : \frac{(|\omega| + \eta)^{1/2}}{(|\omega| + \eta + \Delta)^{1/6}} \geq N^{-5\varepsilon}\Phi_k(\omega, \eta) \right\},$$
$$\mathbb{D}_2(\omega) := \left\{ \eta \in [N^{\gamma-1}, \delta_*] : N^{5\varepsilon}\frac{(|\omega| + \eta)^{1/2}}{(|\omega| + \eta + \Delta)^{1/6}} \leq \Phi_k(\omega, \eta) \leq N^{2\varepsilon}(|\omega| + \eta + \Delta)^{1/3} \right\},$$
$$\mathbb{D}_3(\omega) := \left\{ \eta \in [N^{\gamma-1}, \delta_*] : (|\omega| + \eta + \Delta)^{1/3} \leq N^{-2\varepsilon}\Phi_k(\omega, \eta) \right\}.$$

  If any of the two regimes $\mathbb{D}_l(\omega)$ with $l = 2, 3$ consists of a single point only, then we set $\mathbb{D}_l(\omega) := \emptyset$.

- **Internal minimum:** If $\tau_0 \in \mathbb{M} \setminus \partial \operatorname{supp} \rho$, then we set $\mathbb{D}_2(\omega) := \emptyset$ and define

$$\mathbb{D}_1(\omega) := \left\{ \eta \in [N^{\gamma-1}, \delta_*] : \rho(\tau_0) + (|\omega| + \eta)^{1/3} \geq N^{-2\varepsilon}\Phi_k(\omega, \eta) \right\},$$
$$\mathbb{D}_3(\omega) := \left\{ \eta \in [N^{\gamma-1}, \delta_*] : \rho(\tau_0) + (|\omega| + \eta)^{1/3} \leq N^{-2\varepsilon}\Phi_k(\omega, \eta) \right\}.$$

  If $\mathbb{D}_3(\omega)$ consists of a single point only, then we set $\mathbb{D}_3(\omega) := \emptyset$.

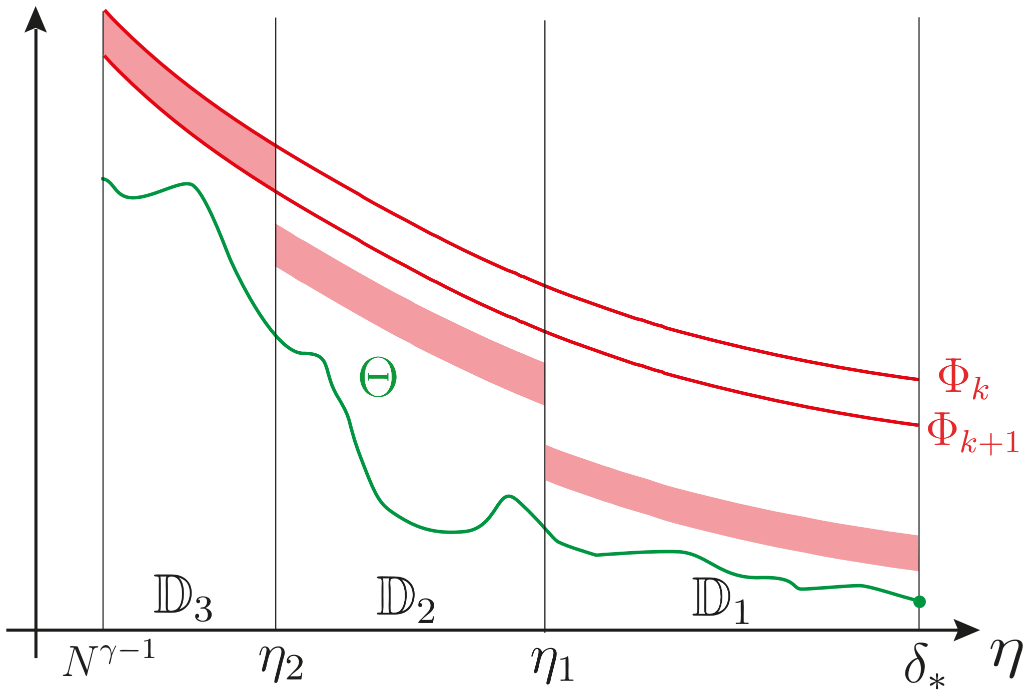

Figure 18.1: The continuous function Θ cannot cross the forbidden areas.

In the cubic equation (18.26), used to define the error function $\mathcal{E}$, the coefficients $\widetilde{\pi}_1$ and $\widetilde{\pi}_2$ on the left hand side are monotonously increasing functions of η. The linear and the constant coefficient of $\mathcal{E}$ on the right hand side are monotonously decreasing in η. Thus, $\mathcal{E}$ itself is a monotonously decreasing function of η. From this fact and the definition of the regimes $\mathbb{D}_1$, $\mathbb{D}_2$ and $\mathbb{D}_3$ we see that $\mathbb{D}_1 = [\eta_1, \delta_*]$, $\mathbb{D}_2 = [\eta_2, \eta_1]$ and $\mathbb{D}_3 = [N^{\gamma-1}, \eta_2]$ for some $\eta_1, \eta_2 \in [N^{\gamma-1}, \delta_*]$. Here, we interpret $\mathbb{D}_2 = \emptyset$ if $\eta_1 \le \eta_2$ and $\mathbb{D}_3 = \emptyset$ if $\eta_2 \le N^{\gamma-1}$.

Now we define a z-dependent indicator function

$$\chi(\omega,\eta) \;:=\; \begin{cases} \mathbb{1}\big(N^{-7\varepsilon}\Phi_k(\omega,\eta) \le \Theta(\tau_0+\theta\omega+\mathrm{i}\eta) \le N^{-6\varepsilon}\Phi_k(\omega,\eta)\big) & \text{if } \eta\in\mathbb{D}_1(\omega) \\ \mathbb{1}\big(N^{-4\varepsilon}\Phi_k(\omega,\eta) \le \Theta(\tau_0+\theta\omega+\mathrm{i}\eta) \le N^{-3\varepsilon}\Phi_k(\omega,\eta)\big) & \text{if } \eta\in\mathbb{D}_2(\omega) \\ \mathbb{1}\big(N^{-\varepsilon}\Phi_k(\omega,\eta) \le \Theta(\tau_0+\theta\omega+\mathrm{i}\eta) \le \Phi_k(\omega,\eta)\big) & \text{if } \eta\in\mathbb{D}_3(\omega) \end{cases}. \tag{18.38}$$

This function fixes the values of Θ to a small interval just below the deterministic control parameter Φ_k. We will prove that Θ cannot take these values, i.e. $\chi = 0$ a.w.o.p.. Figure 18.1 illustrates this argument. Compared to Figure 6.1 in [26] we see that instead of two there are now three domains, $\mathbb{D}_1(\omega)$, $\mathbb{D}_2(\omega)$ and $\mathbb{D}_3(\omega)$, to be distinguished. The reason for this extra complication is that (18.9) is cubic in Θ, compared to the quadratic equation for $[v]$ that appeared in the proof of Lemma 6.2 in [26]. To see that $\chi = 0$, first note that the choice of the domains, $\mathbb{D}_l$, ensures that there is always one summand on the left hand side of the cubic equation (18.9) for Θ which dominates the two others by a factor N^{ε}, whenever χ does not vanish. In fact, by construction we have:

Claim: The random functions Θ and χ satisfy a.w.o.p.

$$\big(\Theta(z)^3 + \widetilde{\pi}_2(\omega,\eta)\,\Theta(z)^2 + \widetilde{\pi}_1(\omega,\eta)\,\Theta(z)\big)\chi(\omega,\eta) \;\lesssim\; \big|\Theta(z)^3 + \pi_2(z)\,\Theta(z)^2 + \pi_1(z)\,\Theta(z)\big|. \tag{18.39}$$

We will verify this fact at the end of the proof of this lemma. Now we will simply use it. By the definition of the indicator function χ we have $\Theta\chi \ge N^{-7\varepsilon}\Phi_k$ Using the assumption (18.32) of the lemma and (18.39) we conclude that

$$\big(\mathcal{R}^3+\widetilde{\pi}_2\mathcal{R}^2+\widetilde{\pi}_1\mathcal{R}\big)\chi \,\le\, N^{8\varepsilon}\frac{\mathcal{R}}{N\eta}+\frac{\widetilde{\rho}}{N\eta}+\frac{1}{(N\eta)^2}\,, \qquad \text{a.w.o.p.}\,, \qquad \mathcal{R}:=N^{-8\varepsilon}\Phi_k\,.$$

Here we gave up one factor of N^ε to get an inequality instead of the stochastic domination. By the choice of ε we have $8\varepsilon\le\gamma/2$. Thus, we see from the cubic equation, (18.26), defining $\mathcal{E}$, and from its monotonicity property that a.w.o.p. $N^{-8\varepsilon}\Phi_k\chi\le\mathcal{E}$. But by the definition of Φ_k in (18.31) we know that $\Phi_k>N^{8\varepsilon}\mathcal{E}$. Together these two inequalities imply

$$\chi(\omega,\eta)\,=\,0\,, \qquad \eta\in[N^{\gamma-1},\delta_*]\,, \qquad \text{a.w.o.p.}\,. \tag{18.40}$$

Now we successively, for $l=1,2,3$, apply Lemma C.1 on the connected domains $\tau_0+\theta\omega+\mathrm{i}\mathbb{D}_l(\omega)$ with the choices $\varphi:=\Theta$ and

$$\Phi(\tau_0+\theta\omega+\mathrm{i}\eta)\,:=\,\begin{cases}N^{-6\varepsilon}\Phi_k(\omega,\eta) & \text{if } \ l=1\,,\\ N^{-3\varepsilon}\Phi_k(\omega,\eta) & \text{if } \ l=2\,,\\ \Phi_k(\omega,\eta) & \text{if } \ l=3\,,\end{cases} \qquad z_0\,:=\,\begin{cases}\tau_0+\theta\omega+\mathrm{i}\delta_* & \text{if } \ l=1\,,\\ \tau_0+\theta\omega+\mathrm{i}\eta_1 & \text{if } \ l=2\,,\\ \tau_0+\theta\omega+\mathrm{i}\eta_2 & \text{if } \ l=3\,,\end{cases}$$

where as explained after the definition of $\mathbb{D}_1$, $\mathbb{D}_2$ and $\mathbb{D}_3$ above we have $\mathbb{D}_1=[\eta_1,\delta_*]$, $\mathbb{D}_2=[\eta_2,\eta_1]$ and $\mathbb{D}_3=[N^{\gamma-1},\eta_2]$. The condition (C.1) of the lemma is satisfied because of the definition of Θ in (18.7), the Hölder-continuity of the solution of the QVE, the weak Lipschitz-continuity of $\mathbf{g}$ with Lipschitz-constant N^2 and the Hölder-continuity of $\mathbf{s}$ from (18.11a). The gap condition, (C.2), holds because of (18.40) and the definition of χ and Φ for an appropriate choice of the exponent D_3.

The condition, $\varphi(z_0)\le\Phi(z_0)$ a.w.o.p., necessary for the application of Lemma C.1 on the first domain, $\tau_0+\theta\omega+\mathrm{i}\mathbb{D}_1(\omega)$, is obtained form Proposition 17.1. With Lemma C.1 we propagate the bound to all $z\in\tau_0+\theta\omega+\mathrm{i}\mathbb{D}_1(\omega)$. Now we apply Lemma C.1 on the second domain $\tau_0+\theta\omega+\mathrm{i}\mathbb{D}_2(\omega)$, provided $\mathbb{D}_2(\omega)$ is not empty. The bound (C.3) for the new $z_0=\tau_0+\theta\omega+\mathrm{i}\eta_1$ is obtained from the previous step. Finally, we apply Lemma C.1 to $\tau_0+\theta\omega+\mathrm{i}\mathbb{D}_3(\omega)$, in case it is not empty, with the new choice $z_0=\tau_0+\theta\omega+\mathrm{i}\eta_2$. Altogether, we applied the lemma at most three times. Through this procedure we prove that a.w.o.p. $\Theta(z)\le\Phi(z)$ for all $z\in\tau_0+\theta\omega+\mathrm{i}[N^{\gamma-1},\delta_*]$. On the third domain, $\tau_0+\theta\omega+\mathrm{i}\mathbb{D}_3(\omega)$, we use that a.w.o.p. $\chi=0$ (cf. (18.40)) and thus a.w.o.p. $\Theta(z)\le N^{-\varepsilon}\Phi_k$. Altogether we showed that in the δ_*-neighbourhood of τ_0,

$$\text{a.w.o.p.} \qquad \Theta(z)\,\le\,N^{-\varepsilon}\Phi_k\,\le\,\Phi_{k+1}\,.$$

This finishes the proof of Lemma 18.5 up to verifying the claim (18.39).

Proof of the claim: For the proof of this inequality one verifies cases by cases that on $\mathbb{D}_1$ the term $\widetilde{\pi}_1\Theta\sim|\pi_1|\Theta$ is bigger than the two other terms, $\widetilde{\pi}_2\Theta^2$ and Θ^3 by a factor of N^ε. If $\mathbb{D}_3$ is not empty then the term Θ^3 is the biggest in that regime. If $\mathbb{D}_2$ is not empty, then $|\pi_2|\sim\widetilde{\pi}_2$ and $\widetilde{\pi}_2\Theta^2$ is the biggest term by a factor of N^ε.

As an example we demonstrate these relations in a few cases:

- **Well inside a gap:** If $\tau_0\in\partial\operatorname{supp}\rho$ and $\omega\in[c_*\Delta,\Delta/2]$ then $\mathbb{D}_2(\omega)=\emptyset$. We now check that on $\mathbb{D}_1(\omega)$ the linear term in Θ is the biggest while on $\mathbb{D}_3(\omega)$ the cubic term dominates. First, let $\eta\in\mathbb{D}_1(\omega)$. Then the following chain of inequalities hold,

$$\widetilde{\pi}_1\Theta\,\sim\,|\pi_1|\Theta\,\sim\,(\Delta+\eta)^{2/3}\Theta\,\gtrsim\,N^{-5\varepsilon}(\Delta+\eta)^{1/3}\,\Phi_k\Theta\,\sim\,N^{-5\varepsilon}\widetilde{\pi}_2\,\Phi_k\Theta\,\gtrsim\,N^{-10\varepsilon}\,\Phi_k^2\Theta\,.$$

  Here, we used (18.25), (18.12b), the definition of $\mathbb{D}_1(\omega)$ and (18.23c) in the form $\widetilde{\pi}_2\sim(\Delta+\eta)^{1/3}$. By definition of χ and since $\widetilde{\pi}_k\gtrsim|\pi_k|$ for $k=1,2$ we also get

$$N^{-5\varepsilon}\widetilde{\pi}_2\,\Phi_k\Theta\chi\,\ge\,N^{\varepsilon}\widetilde{\pi}_2\Theta^2\chi\,\gtrsim\,N^{\varepsilon}|\pi_2|\Theta^2\chi\,, \qquad N^{-10\varepsilon}\,\Phi_k^2\Theta\chi\,\ge\,N^{2\varepsilon}\Theta^3\chi\,.$$

We conclude that on $\mathbb{D}_1(\omega)$ the linear term in Θ dominates the others,

$$\widetilde{\pi}_1\Theta\chi \gtrsim N^{\varepsilon}(\Theta^3+\widetilde{\pi}_2\Theta^2)\chi\,.$$

Suppose now that $\eta \in \mathbb{D}_3(\omega)$. In this case, using the choice of the indicator function χ,

$$\Theta^3\chi \geq N^{-\varepsilon}\Phi_k\Theta^2\chi \geq N^{-2\varepsilon}\Phi_k^2\Theta\chi\,.$$

By definition of $\mathbb{D}_3(\omega)$ and (18.23c) we find that

$$N^{-\varepsilon}\Phi_k\Theta^2 \gtrsim N^{\varepsilon}(\Delta+\eta)^{1/3}\Theta^2 \sim N^{\varepsilon}\widetilde{\pi}_2\Theta^2\,, \qquad N^{-2\varepsilon}\Phi_k^2\Theta \gtrsim N^{2\varepsilon}(\Delta+\eta)^{2/3}\Theta \sim N^{2\varepsilon}\widetilde{\pi}_1\Theta\,.$$

Altogether we find that the cubic term dominates the two others,

$$\Theta^3\chi \gtrsim N^{\varepsilon}(\widetilde{\pi}_2\Theta^2+\widetilde{\pi}_1\Theta)\chi\,.$$

- **Inside a gap close to an edge on $\mathbb{D}_2$:** If $\tau_0 \in \partial\operatorname{supp}\rho$, $\omega \in [0, c_*\Delta]$ and $\eta \in \mathbb{D}_2(\omega)$, then we will show the quadratic term in Θ dominates the two other terms. We have

  $$|\pi_2|\Theta^2 \sim \widetilde{\pi}_2\Theta^2 \sim (\Delta+\eta)^{1/3}\Theta^2 \gtrsim N^{-2\varepsilon}\Phi_k\Theta^2,$$

  where in the inequality we used the definition of $\mathbb{D}_2(\omega)$. The choice of χ guarantees that $\Phi_k\chi \geq N^{3\varepsilon}\Theta\chi$. Thus, the quadratic term is larger than the cubic term by a factor of N^{ε}. On the other hand

  $$(\Delta+\eta)^{1/3}\Theta^2\chi \gtrsim N^{-4\varepsilon}(\Delta+\eta)^{1/3}\Phi_k\Theta \gtrsim N^{\varepsilon}(\omega+\eta)^{1/2}(\Delta+\eta)^{1/6}\Theta \sim N^{\varepsilon}\widetilde{\pi}_1\Theta \sim N^{\varepsilon}|\pi_1|\Theta\,.$$

  Here, in the first inequality we used the indicator function χ and in the second inequality the definition of $\mathbb{D}_2(\omega)$. Altogether, we arrive at

  $$\widetilde{\pi}_2\Theta^2\chi \gtrsim N^{\varepsilon}(\Theta^3+\widetilde{\pi}_1\Theta)\chi\,.$$

- **Internal minimum on $\mathbb{D}_1$:** If $\tau_0 \in \mathbb{M}\setminus\partial\operatorname{supp}\rho$ and $\eta \in \mathbb{D}_1(\omega)$, then the linear term is the biggest,

  $$|\pi_1|\Theta \sim \widetilde{\pi}_1\Theta \sim \big(\rho(\tau_0)^2+(|\omega|+\eta)^{2/3}\big)\Theta \gtrsim N^{-2\varepsilon}\big(\rho(\tau_0)+(|\omega|+\eta)^{1/3}\big)\Phi_k\Theta\,.$$

  Here, we used (18.25) and the definitions of $\widetilde{\pi}_1$ and $\mathbb{D}_1(\omega)$, respectively. Since $\Phi_k\chi \geq N^{6\varepsilon}\Theta\chi$ and by the definition of $\widetilde{\pi}_2$ this shows that the linear term is larger than the quadratic term by a factor of $N^{4\varepsilon}$. In order to compare the linear with the cubic term we estimate further. By definition of $\mathbb{D}_1(\omega)$,

  $$N^{-2\varepsilon}\big(\rho(\tau_0)+(|\omega|+\eta)^{1/3}\big)\Phi_k\Theta \geq N^{-4\varepsilon}\Phi_k^2\Theta\,.$$

  Again we use the lower bound on $\Phi_k\chi$ and get

  $$N^{-4\varepsilon}\Phi_k^2\Theta\chi \geq N^{8\varepsilon}\Theta^3\chi\,.$$

  Thus we showed that on the domain $\mathbb{D}_1(\omega)$

  $$\widetilde{\pi}_1\Theta\chi \gtrsim N^{\varepsilon}(\Theta^3+\widetilde{\pi}_2\Theta^2)\chi\,.$$

The other cases are proven similarly. This completes the proof of (18.39). $\square$

19 Rigidity and delocalisation of eigenvectors

19.1 Proof of Corollary 15.8

Here we explain how the local law, Theorem 15.6, is used to estimate the difference between the cumulative density of states and the eigenvalue distribution function of the random matrix $\mathbf{H}$. The following auxiliary result shows that the difference between two probability measures can be estimated in terms of the difference of their respective Stieltjes transforms. For completeness the proof is given in the appendix. It uses a Cauchy integral formula that was also applied in the construction of the Helffer-Sjöstrand functional calculus (cf. [21]) and it appeared in different variants in [37], [28] and [36].

LEMMA 19.1 (Bounding measures by Stieltjes transforms). *There is a universal constant $C > 0$, such that for any two probability measures, ν_1 and ν_2, on the real line and any three numbers $\eta_1, \eta_2, \varepsilon \in (0, 1]$ with $\varepsilon \geq \max\{\eta_1, \eta_2\}$, the difference between the two measures evaluated on the interval $[\tau_1, \tau_2] \subseteq \mathbb{R}$ with $\tau_1 < \tau_2$ satisfies*

$$|\nu_1([\tau_1, \tau_2]) - \nu_2([\tau_1, \tau_2])| \leq C \big(\nu_1([\tau_1 - \eta_1, \tau_1] \cup [\tau_2, \tau_2 + \eta_2]) + J_1 + J_2 + J_3\big) . \tag{19.1}$$

Here, the three contributions to the error, J_1, J_2 and J_3, are defined as

$$\begin{aligned}
J_1 &:= \int_{\tau_1-\eta_1}^{\tau_1} \mathrm{d}\omega \left(\operatorname{Im} m_{\nu_1}(\omega + \mathrm{i}\eta_1) + |m_{\nu_1-\nu_2}(\omega + \mathrm{i}\eta_1)| + \frac{1}{\eta_1} \int_{\eta_1}^{2\varepsilon} \mathrm{d}\eta |m_{\nu_1-\nu_2}(\omega + \mathrm{i}\eta)| \right) , \\
J_2 &:= \int_{\tau_2}^{\tau_2+\eta_2} \mathrm{d}\omega \left(\operatorname{Im} m_{\nu_1}(\omega + \mathrm{i}\eta_2) + |m_{\nu_1-\nu_2}(\omega + \mathrm{i}\eta_2)| + \frac{1}{\eta_2} \int_{\eta_2}^{2\varepsilon} \mathrm{d}\eta |m_{\nu_1-\nu_2}(\omega + \mathrm{i}\eta)| \right) , \\
J_3 &:= \frac{1}{\varepsilon} \int_{\tau_1-\eta_1}^{\tau_2+\eta_2} \mathrm{d}\omega \int_{\varepsilon}^{2\varepsilon} \mathrm{d}\eta |m_{\nu_1-\nu_2}(\omega + \mathrm{i}\eta)| ,
\end{aligned} \tag{19.2}$$

where m_ν denotes the Stieltjes transform of ν for any signed measure ν.

We will now apply this lemma to prove Corollary 15.8 with the choices of the measures

$$\nu_1(\mathrm{d}\omega) := \rho(\omega)\mathrm{d}\omega , \qquad \nu_2(\mathrm{d}\omega) := \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}(\mathrm{d}\omega) . \tag{19.3}$$

As a first step we show that a.w.o.p. there are no eigenvalues with an absolute value larger or equal than 10, i.e.,

$$\#\{i : |\lambda_i| \geq 10\} = 0 \qquad \text{a.w.o.p.} . \tag{19.4}$$

We focus on the eigenvalues $\lambda_i \geq 10$. The ones with $\lambda_i \leq -10$ are treated in the same way. We will show first that there are no eigenvalues in a small interval around τ with $\tau \geq 10$. In fact, we prove that for $\gamma \in (0, 1/3)$,

$$\#\{i : \tau \leq \lambda_i \leq \tau + N^{-1}\} \prec N^{-\gamma}. \tag{19.5}$$

For this we apply Lemma 19.1 with the same choices of the measures ν_1 and ν_2 as in (19.3) and with

$$\eta_1 := \eta_2 := \varepsilon := N^{\gamma-1} , \qquad \tau_1 := \tau , \qquad \tau_2 := \tau + N^{-1}. \tag{19.6}$$

Theorem 15.6 and the Lipschitz-continuity of $\langle \mathbf{g} \rangle$ with Lipschitz-constant bounded by N^2, as well as the uniform 1/3-Hölder-continuity of $\langle \mathbf{m} \rangle$ imply that

$$\sup \big|\langle \mathbf{g}(\omega + \mathrm{i}\eta)\rangle - \langle \mathbf{m}(\omega + \mathrm{i}\eta)\rangle\big| \prec \frac{1}{N} + N^{-2\gamma}, \tag{19.7}$$

where the supremum is taken over $\omega \in [\tau - N^{\gamma-1}, \tau + 2N^{\gamma-1}]$ and $\eta \in [N^{\gamma-1}, 2N^{\gamma-1}]$. Plugging this bound into the definition of J_1, J_2 and J_3 from (19.2) and using (19.1) and the fact that $\rho = 0$ in this regime shows the validity of (19.5).

We conclude that a.w.o.p. there are no eigenvalues in an interval of length N^{-1} to the right of τ. By using a union bound this implies that

$$\#\{i : 10 \le \lambda_i \le N\} \;=\; 0 \qquad \text{a.w.o.p.}\,.$$

The eigenvalues larger than N are treated by the following simple argument,

$$\max_{i=1,\dots,N} \lambda_i^2 \;\le\; \sum_{i=1}^N \lambda_i^2 \;=\; \sum_{i,j=1}^N |h_{ij}|^2 \;\prec\; N\,.$$

Thus (19.4) holds true.

Now we apply Lemma 19.1 to prove (15.23). In case $|\tau| \ge 10$ the bound (15.23) follows because a.w.o.p. there are no eigenvalues of $\mathbf{H}$ with absolute value larger or equal than 10. Thus, we fix $\tau \in (-10, 10)$ and make the choices

$$\eta_1 \;:=\; \eta_2 \;:=\; N^{\gamma-1}\,, \qquad \tau_1 \;:=\; -10\,, \qquad \tau_2 \;:=\; \tau\,, \qquad \varepsilon \;:=\; 1\,. \tag{19.8}$$

Again we use (15.17) from Theorem 15.6, the Lipschitz-continuity of $\langle \mathbf{g} \rangle$ and the Hölder-continuity of $\langle \mathbf{m} \rangle$ to see that uniformly for all $\eta \ge N^{\gamma-1}$,

$$\sup_{\omega \in [0,\eta_1]} \big| \langle \mathbf{g}(\tau_1 - \omega + \mathrm{i}\eta) \rangle - \langle \mathbf{m}(\tau_1 - \omega + \mathrm{i}\eta) \rangle \big| \prec \frac{1}{N} + \frac{1}{(N\eta)^2}\,.$$

Here we evaluated $\Delta(\tau_1) = 1$ and thus $\kappa \lesssim \eta + (N\eta)^{-1}$. With J_1 defined as in (19.2) we infer $J_1 \prec N^{-1}$. Theorem 15.6 also implies the bound

$$\sup_{\omega \in [-20,20]} \sup_{\eta \in [1,2]} \big| \langle \mathbf{g}(\omega + \mathrm{i}\eta) \rangle - \langle \mathbf{m}(\omega + \mathrm{i}\eta) \rangle \big| \prec \frac{1}{N}\,,$$

since in this regime $\kappa \lesssim 1$, thus showing that $J_3 \prec N^{-1}$. We are left with estimating the three terms constituting J_2. The first and second of these terms are estimated trivially by using the boundedness of their integrands. Therefore, we conclude that

$$\left| \int_{-10}^{\tau} \rho(\omega)\mathrm{d}\omega - \frac{\#\{i : -10 \le \lambda_i \le \tau\}}{N} \right| \prec N^{\gamma-1} + R(\tau)\,, \tag{19.9}$$

where the error term, R, is defined as

$$R(\tau) \;:=\; N^{1-\gamma} \int_0^{N^{\gamma-1}} \mathrm{d}\omega \int_{N^{\gamma-1}}^{2} \mathrm{d}\eta \, \min\biggl\{ \frac{1}{N\eta(\Delta(\tau+\omega)^{1/3} + \rho(\tau+\omega+\mathrm{i}\eta))}\,, \frac{1}{(N\eta)^{1/2}} \biggr\}. \tag{19.10}$$

This expression is derived by using the bound (15.19) on κ for the integrand of the third contribution to J_2.

To estimate R further we distinguish three cases, depending on whether τ is away from $\mathbb{M}$, close to an edge or close to a local minimum in the interior of $\operatorname{supp}\rho$. In each of these cases we prove

$$R(\tau) \;\prec\; \min\biggl\{ \frac{1}{N(\Delta(\tau)^{1/3} + \rho(\tau))}\,, \frac{1}{N^{4/5}} \biggr\}. \tag{19.11}$$

Away from $\mathbb{M}$: In case $\operatorname{dist}(\tau, \mathbb{M}) \geq \delta_*$, with δ_* the size of the neighbourhood around the local minima from Theorem 18.1, we have $\Delta^{1/3} + \rho \sim 1$ and thus the η-integral in (19.10) yields a factor comparable to $N^{-1} \log N$. Thus, $R(\tau) \prec N^{-1}$.

Close to an edge: Let $\operatorname{dist}(\tau, \{\alpha_k, \beta_k\}) \leq \delta_*$. Then from the size of ρ at an internal edge, at the extreme edges and inside the gap (cf. (18.5b), (18.5d) and (18.5c) from Theorem 18.1) we see that

$$\Delta(\tau+\omega)^{1/3} + \rho(\tau+\omega+\mathrm{i}\eta) \sim \big(\Delta(\tau) + \operatorname{dist}(\tau, \{\alpha_k, \beta_k\}) + \eta\big)^{1/3}.$$

for any $\omega \in [0, N^{\gamma-1}]$ and $\eta \in [N^{\gamma-1}, 2]$. With this the size of R is given by

$$R(\tau) \sim \int_{N^{\gamma-1}}^{2} \mathrm{d}\eta \min\bigg\{ \frac{1}{N\eta(\Delta(\tau) + \operatorname{dist}(\tau, \{\alpha_k, \beta_k\}) + \eta)^{1/3}}, \frac{1}{(N\eta)^{1/2}} \bigg\}.$$

Integrating over η yields that

$$R(\tau) \lesssim \min\bigg\{ \frac{\log N}{N(\Delta(\tau) + \operatorname{dist}(\tau, \{\alpha_k, \beta_k\}))^{1/3}}, \frac{1}{N^{4/5}} \bigg\}.$$

Now (19.11) follows by using the size of ρ from Theorem 18.1 again.

Close to an internal local minimum: Suppose $|\tau - \tau_0| \leq \delta_*$ for some $\tau_0 \in \mathbb{M} \setminus \partial \operatorname{supp} \rho$. Then using the size of ρ from (18.5e) of Theorem 18.1 we see that

$$R(\tau) \sim \int_{N^{\gamma-1}}^{2} \mathrm{d}\eta \min\bigg\{ \frac{1}{N\eta(\rho(\tau_0) + |\tau-\tau_0|^{1/3} + \eta^{1/3})}, \frac{1}{(N\eta)^{1/2}} \bigg\}.$$

The bound (19.11) follows by performing the integration over η.

This finishes the proof of (19.11). We insert this bound into (19.9) and use that γ was arbitrary. Thus, we find

$$\bigg| \int_{-10}^{\tau} \rho(\omega)\, \mathrm{d}\omega - \frac{\#\{\, i : -10 \leq \lambda_i \leq \tau\}}{N} \bigg| \prec \min\bigg\{ \frac{1}{N(\Delta(\tau)^{1/3} + \rho(\tau))}, \frac{1}{N^{4/5}} \bigg\}.$$

This finishes the proof of (15.23) since there are no eigenvalues below -10.

Now we prove (15.24). Let $\tau \in \mathbb{R} \setminus \operatorname{supp} \rho$. Suppose that for some $k = 1, \ldots, K$ we have $|\tau - \beta_k| = \operatorname{dist}(\tau, \partial \operatorname{supp} \rho)$. The case when τ is closer to the set $\{\alpha_k\}$ than to $\{\beta_k\}$ is treated similarly. Suppose further that

$$\tau \geq \alpha_k + \delta_k,$$

where δ_k are defined as in (15.25) and $\delta_0 = N^{\gamma-2/3}$. Note that there is nothing to show if $k > 1$ and the size of the gap, $\alpha_k - \beta_{k-1}$, is smaller than $2\delta_k$, i.e., if such a τ does not exist. In particular, we have $\alpha_k - \beta_{k-1} = \Delta(\tau) \gtrsim N^{-1/2}$. We will show that a.w.o.p. there are no eigenvalues in an interval of length $N^{-2/3}$ to the right of τ, i.e.

$$\#\big\{\, i : \tau \leq \lambda_i \leq \tau + N^{-2/3}\big\} = 0 \qquad \text{a.w.o.p.}. \tag{19.12}$$

We apply Lemma 19.1 with the same choices of the measures ν_1 and ν_2 as in (19.3). Additionally, we set

$$\eta_1 := \eta_2 := \varepsilon := N^{-2/3}, \qquad \tau_1 := \tau, \qquad \tau_2 := \tau + N^{-2/3}. \tag{19.13}$$

We use the local law, Theorem 15.6, to estimate the differences between the Stieltjes transforms of the two measures for the integrands in the definition of the three error terms, J_1, J_2 and J_3 from (19.2). By the definition of δ_k the condition (15.20) is satisfied inside the integrals and we use the improved bound, (15.21), on κ. Indeed, we find

$$\sup \big|\langle \mathbf{g}(\omega + \mathrm{i}\eta)\rangle - \langle \mathbf{m}(\omega + \mathrm{i}\eta)\rangle\big| \prec \frac{1}{N\delta_k \Delta(\tau)^{1/3}} + \frac{1}{N^{2/3}\,\delta_k^{1/2}\Delta(\tau)^{1/6}},$$

where the supremum is taken over $\omega \in [\tau - N^{-2/3}, \tau + 2N^{-2/3}]$ and $\eta \in [N^{-2/3}, 2N^{-2/3}]$. With this, the definition of δ_k and the size of ρ from (18.5c) and (18.5d) we infer

$$J_1 + J_2 + J_3 \prec N^{-1-\gamma/2}.$$

From this (19.12) follows. The claim, (15.24), is now a consequence of a simple union bound taken over the events in (19.12) with different choices of τ. This finishes the proof of Corollary 15.8. $\square$

19.2 Proof of Corollary 15.9

Here we show how we get the rigidity, Corollary 15.9, from Corollary 15.8. Fix a $\tau \in [\alpha_1, \beta_K]$. We define the fluctuation to the left, δ_-, and to the right, δ_+, of the eigenvalue $\lambda_{i(\tau)}$ as

$$\delta_+(\tau) := \inf\bigg\{ \delta \geq 0 : 2 + \Big|\#\{i : \lambda_i \leq \tau + \delta\} - N\int_{-\infty}^{\tau+\delta} \rho(\omega)\mathrm{d}\omega\Big| \leq N\int_{\tau}^{\tau+\delta} \rho(\omega)\mathrm{d}\omega \bigg\} \tag{19.14a}$$

$$\delta_-(\tau) := \inf\bigg\{ \delta \geq 0 : 1 + \Big|\#\{i : \lambda_i \leq \tau - \delta\} - N\int_{-\infty}^{\tau-\delta} \rho(\omega)\mathrm{d}\omega\Big| \leq N\int_{\tau-\delta}^{\tau} \rho(\omega)\mathrm{d}\omega \bigg\}. \tag{19.14b}$$

We show now that with this definition,

$$\lambda_{i(\tau)} \in [\tau - \delta_-(\tau), \tau + \delta_+(\tau)]\,. \tag{19.15}$$

We start with the upper bound on $\lambda_{i(\tau)}$. By definition of $i(\tau)$ we find the inequality

$$\#\{i : \lambda_i \leq \lambda_{i(\tau)}\} = i(\tau) \leq 1 + N\int_{-\infty}^{\tau} \rho(\omega)\mathrm{d}\omega = 1 + N\int_{-\infty}^{\tau+\delta_+} \rho(\omega)\mathrm{d}\omega - N\int_{\tau}^{\tau+\delta_+} \rho(\omega)\mathrm{d}\omega\,.$$

The definition of $\delta_+ = \delta_+(\tau)$ implies that

$$\#\{i : \lambda_i \leq \lambda_{i(\tau)}\} < \#\{i : \lambda_i \leq \tau + \delta_+\}\,.$$

By monotonicity of the cumulative eigenvalue distribution, we conclude that $\lambda_{i(\tau)} \leq \tau + \delta_+$. Thus, the upper bound is proven.

Now we show the lower bound. We start similarly,

$$\#\{i : \lambda_i \leq \lambda_{i(\tau)}\} = i(\tau) \geq N\int_{-\infty}^{\tau} \rho(\omega)\mathrm{d}\omega = N\int_{-\infty}^{\tau-\delta_-} \rho(\omega)\mathrm{d}\omega + N\int_{\tau-\delta_-}^{\tau} \rho(\omega)\mathrm{d}\omega\,.$$

By definition of δ_- we get

$$\#\{i : \lambda_i \leq \lambda_{i(\tau)}\} \geq 1 + \liminf_{\varepsilon\downarrow 0} \#\{i : \lambda_i \leq \tau - \delta_- - \varepsilon\}\,.$$

Here the lim inf is necessary, since the cumulative eigenvalue distribution is not continuous from the left. We conclude that $\lambda_{i(\tau)} \geq \tau - \delta_- - \varepsilon$ for all $\varepsilon > 0$ and therefore the lower bound is proven.

Now we start with the proof of (15.29). For this we show that for any τ that is well inside the support of the density of states, i.e., that satisfies (15.27), we have

$$\delta_-(\tau) + \delta_+(\tau) \prec \delta \,, \qquad \delta := \min\biggl\{ \frac{1}{\rho(\tau)(\Delta(\tau)^{1/3} + \rho(\tau))N} \,, \frac{1}{N^{3/5}} \biggr\} . \tag{19.16}$$

If τ is in the bulk, i.e., $\mathrm{dist}(\tau, \mathbb{M}) \geq \delta_*$, then $\delta \sim N^{-1}$ and thus (19.16) follows from (15.23). We distinguish the two remaining cases, namely whether τ is close to an edge or to a local minimum inside the interior of $\operatorname{supp} \rho$.

Close to an edge: Suppose that $\tau \in [\beta_k - \delta_*, \beta_k - \varepsilon_k]$. The case when τ is closer to $\{\alpha_k\}$ than to $\{\beta_k\}$ is treated similarly. By the definition of ε_k in (15.28) and by the size of ρ from (18.5d) and (18.5b) in Theorem 18.1 we see that $\varepsilon_k \gtrsim N^\gamma \delta$. Using Corollary 15.8 we find for any $\varepsilon \in (0, \gamma/2)$ that

$$\biggl|\#\{i : \lambda_i \leq \tau + N^\varepsilon \delta\} - N \int_{-\infty}^{\tau + N^\varepsilon \delta} \rho(\omega) \mathrm{d}\omega \biggr| \prec \min\bigl\{ (\Delta(\tau) + \beta_k - \tau)^{-1/3}, N^{1/5} \bigr\} .$$

On the other hand

$$N \int_{\tau}^{\tau + N^\varepsilon \delta} \rho(\omega) \mathrm{d}\omega \sim \frac{N^{1+\varepsilon} \delta\, (\beta_k - \tau)^{1/2}}{(\Delta(\tau) + \beta_k - \tau)^{1/6}} \gtrsim N^\varepsilon \min\bigl\{ (\Delta(\tau) + \beta_k - \tau)^{-1/3}, N^{1/5} \bigr\} .$$

Here we used the size of ρ from Theorem 18.1, the definition of δ and $\beta_k - \tau \geq \varepsilon_k$. Since ε was arbitrary we conclude that $\delta_+(\tau) \prec \delta$. The bound, $\delta_-(\tau) \prec \delta$, is shown in the same way.

Close to internal local minima: Suppose $|\tau - \tau_0| \leq \delta_*$ for some $\tau_0 \in \mathbb{M} \setminus \partial \operatorname{supp} \rho$. Then by (18.5e) with $\Delta(\tau_0) = 0$ and the definition of δ in (19.16) we have

$$\delta \sim \min\Bigl\{ \frac{1}{(\rho(\tau_0)^3 + |\tau - \tau_0|)^{2/3} N}, \frac{1}{N^{3/5}} \Bigr\} .$$

We apply (15.23) from Corollary 15.8 and, using (18.5e) again, we get

$$\biggl|\#\{i : \lambda_i \leq \tau + N^\varepsilon \delta\} - N \int_{-\infty}^{\tau + N^\varepsilon \delta} \rho(\omega) \mathrm{d}\omega \biggr| \prec \min\bigl\{ (\rho(\tau_0)^3 + |\tau + N^\varepsilon \delta - \tau_0|)^{-1/3}, N^{1/5} \bigr\}. \tag{19.17}$$

On the other hand we find

$$N \int_{\tau}^{\tau + N^\varepsilon \delta} \rho(\omega) \mathrm{d}\omega \sim N^{1+\varepsilon} \delta\, (\rho(\tau_0)^3 + |\tau - \tau_0| + N^\varepsilon \delta)^{1/3}. \tag{19.18}$$

We will now verify that for large enough N,

$$N^{\varepsilon/2} \min\bigl\{ (\rho(\tau_0)^3 + |\tau + N^\varepsilon \delta - \tau_0|)^{-1/3}, N^{1/5} \bigr\} \lesssim N^{1+\varepsilon} \delta\, (\rho(\tau_0)^3 + |\tau - \tau_0| + N^\varepsilon \delta)^{1/3}. \tag{19.19}$$

We distinguish three cases. First let us consider the regime where $\rho(\tau_0)^3 + |\tau - \tau_0| \leq N^{-3/5}$. Then we have $\delta = N^{-3/5}$ and

$$N^{1+\varepsilon} \delta\, (\rho(\tau_0)^3 + |\tau - \tau_0| + N^\varepsilon \delta)^{1/3} \sim N^{4\varepsilon/3} N^{1/5} .$$

Now we treat the situation where, $N^{-3/5} < \rho(\tau_0)^3 + |\tau - \tau_0| \leq N^{3\varepsilon/2 - 3/5}$. In this case

$$N^{1+\varepsilon} \delta\, (\rho(\tau_0)^3 + |\tau - \tau_0| + N^\varepsilon \delta)^{1/3} \gtrsim \frac{N^\varepsilon}{(\rho(\tau_0)^3 + |\tau - \tau_0|)^{1/3}} \geq N^{\varepsilon/2} N^{1/5} .$$

Finally, we consider $\rho(\tau_0)^3 + |\tau - \tau_0| > N^{3\varepsilon/2-3/5}$. Then for large enough N we find on the one hand

$$\min\left\{\rho(\tau_0)^3 + |\tau + N^\varepsilon\delta - \tau_0|)^{-1/3},\, N^{1/5}\right\} \sim \frac{1}{(\rho(\tau_0)^3 + |\tau - \tau_0|)^{1/3}}\,,$$

and on the other hand

$$N^{1+\varepsilon}\delta\,(\rho(\tau_0)^3 + |\tau - \tau_0| + N^\varepsilon\delta)^{1/3} \gtrsim \frac{N^\varepsilon}{(\rho(\tau_0)^3 + |\tau - \tau_0|)^{1/3}}\,.$$

Thus, (19.19) holds true and since ε was arbitrary, we infer from (19.17) and (19.18) that $\delta_+(\tau) \prec \delta$. Along the same lines we prove $\delta_-(\tau) \prec \delta$. Thus (19.16) and with it (15.29) are proven.

The statement about the fluctuation of the eigenvalues at the leftmost edge, (15.30) follows directly from (15.29) and (15.24) in Corollary 15.8. Indeed, for $\tau \in [\alpha_1, \alpha_1 + \varepsilon_0)$ we have $\lambda_{i(\tau)} \le \lambda_{i(\alpha_1+\varepsilon_0)}$ and from (15.29) with $\Delta(\tau) = 1$, as well as $\rho(\alpha_1 + \varepsilon_0) \sim \varepsilon_0^{1/2}$, and from the definition of ε_0 we see that

$$\lambda_{i(\alpha_1+\varepsilon_0)} \le \alpha_1 + \varepsilon_0 + N^{\gamma-2/3} \le \tau + 2N^{\gamma-2/3} \qquad \text{a.w.o.p.}\,.$$

On the other hand, (15.24) shows that a.w.o.p. $\lambda_{i(\tau)} \ge \alpha_1 - N^{\gamma-2/3}$. Since γ was arbitrary, (15.30) follows. The rigidity at the rightmost edge, (15.31), is proven along the same lines.

The claim, (15.32), about the remaining eigenvalues follows from a similar argument. For $\tau \in (\beta_k - \varepsilon_k, \alpha_{k+1} + \varepsilon_k)$, as a consequence of (15.24), we have

$$\lambda_{i(\tau)} \in [\lambda_{i(\beta_k-\varepsilon_k)}, \beta_k + \delta_k] \cup [\alpha_{k+1} - \delta_k, \lambda_{i(\alpha_{k+1}+\varepsilon_k)}] \qquad \text{a.w.o.p.}\,.$$

From (15.29) and the definition of ε_k we infer $\lambda_{i(\beta_k-\varepsilon_k)} \ge \beta_k - 2\varepsilon_k$ a.w.o.p., as well as $\lambda_{i(\alpha_{k+1}+\varepsilon_k)} \le \alpha_{k+1} + 2\varepsilon_k$ a.w.o.p., which finishes the proof of (15.32). □

19.3 Proof of Corollary 15.11

The delocalisation of eigenvectors is a simple consequence of the local law, Theorem 15.6. The following argument is taken from [26]. We use the resolvent identity,

$$G_{ii}(z) = \sum_{j=1}^{N} \frac{|u_j(i)|^2}{\lambda_j - z}\,,$$

where $\mathbf{u}_j = (u_j(1), \dots, u_j(N))$ is the eigenvector corresponding to the eigenvalue λ_j. We evaluate this at $z := \lambda_j + \mathrm{i}N^{\gamma-1}$ with $\gamma > 0$ as in the statement of Theorem 15.6. The local law implies the boundedness of $G_{ii}(z)$ which then implies

$$1 \gtrsim |G_{ii}(z)| \ge \frac{|u_j(i)|^2}{|\lambda_j - z|} = N^{1-\gamma}\,|u_j(i)|^2,$$

Where we kept only a single summand from the resolvent identity. Since $\gamma > 0$ was arbitrary we conclude

$$|u_j(i)| \prec N^{-1/2}.$$

□

20 Isotropic law and universality

20.1 Proof of Theorem 15.12

Given the entrywise local law, Theorem 15.6, the proof of the isotropic law follows exactly as Section 7 in [12], where the same argument was presented for generalised Wigner matrices (this argument itself mimicked the detailed proof of the isotropic law for sample covariance matrices in Section 5 of [12]). The only difference is that in our case $G_{ii}(z)$ is close to $m_i(z)$, the solution to the QVE, which now genuinely depends on i, while in [12] we had $G_{ii} \approx m_{sc}$ for every i, where m_{sc} is the solution to (1.3). However, the diagonal resolvent elements played no essential role in [12]. We now explain the small modifications.

Recall from Section 5.2 of [12] that by polarisation it is sufficient to prove (15.33) for ℓ^2-normalised vectors $\mathbf{w} = \overline{\mathbf{v}}$. We can then write

$$\sum_{i,j=1}^{N} \overline{v_i}\, G_{ij} v_j - \sum_{i=1}^{N} m_i |v_i|^2 \;=\; \sum_i (G_{ii} - m_i)|v_i|^2 + \mathcal{Z}, \qquad \mathcal{Z} := \sum_{i \neq j}^{N} \overline{v_i}\, G_{ij} v_j\,.$$

The first term containing the diagonal elements G_{ii} is clearly bounded by the right hand side of (15.33) by Theorem 15.6. This is the first instant where the nontrivial i-dependence of m_i is used.

The main technical part of the proof in [12] is then to control $\mathcal{Z}$, the contribution of the off diagonal terms. We can follow this proof in our case to the latter; the nontrivial i-dependence of m_i requires a slight modification only at one point. To see this, we recall the main structure of the proof. For any even p, the moment

$$\mathbb{E}|\mathcal{Z}|^p = \mathbb{E} \sum_{b_{11} \neq b_{12}} \cdots \sum_{b_{p1} \neq b_{p2}} \Big(\prod_{k=1}^{p/2} \bar{v}_{b_{k1}} G_{b_{k1} b_{k2}} v_{b_{k2}} \Big) \Big(\prod_{k=p/2+1}^{p} \bar{v}_{b_{k1}} G^*_{b_{k1} b_{k2}} v_{b_{k2}} \Big) \tag{20.1}$$

is computed. Using the resolvent identity (16.9) (and a similar one for the reciprocals of the diagonal elements) we successively expand the resolvents until each of them appears in a *maximally expanded* form, i.e. until no further b_{k1} or b_{k2} can be added as an upper index to them by these resolvent identities (see Definition 5.4 of [12]). Next we use (16.3) to each maximally expanded off-diagonal resolvent entry and (17.16) to the reciprocals of the diagonal entries. In this way, only resolvent entries of those minors appear that do not contain any b_{kj} index as rows or columns; in other words, the v-indices and the $G^{(\#)}$-indices are decoupled; only explicit h-terms represent the connections between them. We can now take partial expectation for the rows and columns of these h-terms. In this way we guarantee that v-indices are paired, i.e. the $2p$-fold summation in (20.1) effectively becomes a p-fold summation. This renders the uncontrolled ℓ^1-norm of $\mathbf{v}$ to an ℓ^2-norm which is one by normalisation. Along this procedure we need to replace reciprocals of diagonal resolvent entries $1/G_{ii}$ (that arise from (16.9)) by their deterministic approximation $1/m_i$, see the analogous formulas (5.41)–(5.42) in [12]. Here we use the self-consistent equation (15.3) to conclude from (17.16) that

$$\frac{1}{G_{ii}} = \frac{1}{m_i} + h_{ii} + s_{ii} m_i - \sum_{a,b}^{(i)} \Big(h_{ia} G^{(i)}_{ab} h_{bi} - s_{ia} m_a \delta_{ab} \Big). \tag{20.2}$$

Taking the inverse of this formula and expanding around the leading term $1/m_i \sim 1$, we get a geometric series expansion for G_{ii}. The terms $h_{ii} \prec N^{-1/2}$ and $s_{ii} m_i \lesssim N^{-1}$ are negligible. The term in the square bracket is small by the large deviation estimates (16.7a)–(16.7c) and by the fact that the local law Theorem 15.6 applied to the minor $H^{(i)}$ yields that $G^{(i)}_{aa}$ is close to m_a^i, the solution to the self-consistent equation for the minor. Finally, this latter is close to m_a by the stability of (15.3), since for any fixed i, the $(N-1)$-vectors $(m_a^i)_{a\neq i}$ and $(m_a)_{a\neq i}$ satisfy almost the same self consistent equation, with an additive perturbation of order $1/N$ between them. The proof in [12] did not use the specific form of the subtracted term $s_{ia}m_a\delta_{ab}$ in (20.2), just the fact that the subtraction made (16.7c) applicable for the double summation in (20.2). After this slight modification, the rest of the proof in [12] goes through without any further changes. □

20.2 Proof of Theorem 15.14

For the proof of Theorem 15.14 we follow the method developed in [33, 36, 26]. Theorem 2.1 from [34] was designed for proving universality for a random matrix with a small independent Gaussian component and densities of state that may differ from Wigner's semicircle law. The main theorem in [34] asserts that if local laws hold in a sufficiently strong sense then bulk universality holds locally for matrices with a small Gaussian component. We remark that a similar approach was independently developed in [47] that can also be easily used to conclude bulk universality from Theorem 15.6, but here we follow [34]. In Section 2.5 of [34] a recipe was given how to use this theorem to establish universality for a quite general class of random matrix models even without the Gaussian component, as long as uniform local laws on the optimal scale are known and the matrix satisfies a condition as in Definition 15.13 that allows for an application of the moment matching (Lemma 6.5 in [36]) and the Green's function comparison theorem (Theorem 2.3 in [36]). Following this recipe it remains to show that the local law, Theorem 15.6, holds for all matrices with the same variance matrix as

$$\mathbf{H}_t \;=\; \mathrm{e}^{-t/2}\mathbf{H}_0 + (1-\mathrm{e}^{-t})^{1/2}\mathbf{U},$$

for all $t\in[0,T]$ and T is a small negative power of N, i.e., $T=N^{-\varepsilon}$. Here, $\mathbf{U}$ is the standard Gaussian ensemble (GUE or GOE), $\mathbf{H}_0$ has independent entries and is independent of $\mathbf{U}$ and $\mathbf{H}_T$ has variance matrix $\mathbf{S}$. The bounded moment condition (D) is automatically satisfied for $\mathbf{H}_t$ by the construction of $\mathbf{H}_0$.

The variance matrix of $\mathbf{H}_t$ is

$$\mathbf{S}_t \;=\; \mathrm{e}^{-t}\mathbf{S}_0 + (1-\mathrm{e}^{-t})\mathbf{S}_{\mathrm{G}}\,,$$

where $\mathbf{S}_{\mathrm{G}}$ is the variance matrix of the standard Gaussian ensemble and $\mathbf{S}_0$ is given by

$$\mathbf{S}_0 \;=\; \mathrm{e}^{T}\mathbf{S} - (\mathrm{e}^{T}-1)\mathbf{S}_{\mathrm{G}}\,.$$

Since condition (A) is simply a normalisation, it suffices to verify (B) and (C) from (15.6) and (15.5) for the variance matrices $\mathbf{S}_t$. The uniform primitivity, assumption (B) is satisfied because $\mathbf{S}$ is q-full (cf. Definition 15.13) and $T\ll 1$. In particular, the matrices $\mathbf{S}_t$ are uniformly $q/2$-full. This also implies the boundedness of the corresponding solutions, $\mathbf{m}^{(t)}$, to the QVE in a neighbourhood of the origin, $z=0$, by Remark 9.2 in Part II. The validity of (C) away from $z=0$ follows from Remark 6.11 in Part II and from the L^2-bound,

$$\frac{1}{N}\sum_{i=1}^{N}\big|m_i^{(t)}(z)\big|^2 \;\le\; \frac{3}{|z|}\,,\qquad z\in\mathbb{H}\,,\; t\in[0,T]\,.$$

This bound holds for the solution of the QVE that corresponds to very general variance matrices and is part of the statement of Theorem 6.1 in Part II. The 3 in this bound replaces the original 2 in order to compensate for the potentially violated normalisation of the variance matrices. This shows that Theorem 15.6 can be applied to $\mathbf{H}_t$ and thus universality is proven. □

21 Gaussian translation invariant model

In this section we will prove that Theorem 15.16 applies for dependent Gaussian random matrices introduced in Subsection 15.2.

Let $\widetilde{A}$ denote the integral operator acting on functions $h : [0,1] \to \mathbb{C}$ via

$$\widetilde{A}h(\phi) := \int_0^1 \widetilde{a}(\phi,\theta)h(\theta)\,\mathrm{d}\theta\,, \tag{21.1}$$

where the kernel $\widetilde{a} : [0,1]^2 \to \mathbb{C}$ is defined as

$$\widetilde{a}(\phi,\theta) := \sum_{k,l=0}^{N-1} a_{kl}\, e_k(\phi)\, e_{-l}(\theta)\,, \qquad \phi,\theta \in [0,1]\,, \tag{21.2}$$

and $e_k : \mathbb{R} \to \mathbb{C}$ is the exponential function $e_k(\phi) := \mathrm{e}^{\mathrm{i}2\pi k\phi}$. Here, we identified $\mathbb{T}$ with the set of integers $\{0,1,2,\dots,N-1\}$. We remark that $\widetilde{a}(\phi,\theta) \geq 0$ for all $\phi,\theta \in [0,1]$. This follows from (15.38b) and the Bochner inequality (cf. (21.11) below). Note that $\widehat{a}_{\phi\theta} = N^{-1}\widetilde{a}(\phi,\theta)$ if $\phi,\theta \in \mathbb{S}$ and we consider $\mathbb{S}$ as canonically embedded in $[0,1]$. For $k \in \mathbb{N}$ let $\widetilde{a}^k(\phi,\theta)$ denote the kernel of $\widetilde{A}^k$.

The following proposition is a generalisation of Theorem 15.16 to polynomially decaying correlations.

Proposition 21.1 (Generalisation to good correlation matrices). *Let* $\mathbf{H}$ *and* $\mathbf{A}$ *be related by* (15.38b). *Assume that* $\mathbf{A}$ *is* **good** *in the sense that*

(i) *There is* $\alpha > 0$ *and* $L \in \mathbb{N}$ *such that* $(\widetilde{a}^L)(\phi,\theta) \geq \alpha$ *for every* $\phi,\theta \in [0,1]$*;*

(ii) *There exists an integer* $\beta \geq 1$ *such that*

$$\sum_{x,y} (1+|x|+|y|)^{\beta}|a_{xy}| \;\leq\; 1\,. \tag{21.3}$$

(iii) *The integral operator* $\widetilde{A}$ *is block fully indecomposable: There exist two constants* $\varphi > 0$, $K \in \mathbb{N}$, *a fully indecomposable matrix* $\mathbf{Z} = (Z_{ij})_{i,j=1}^K$, *with* $Z_{ij} \in \{0,1\}$, *and a measurable partition* $\mathcal{D} := \{D_j\}_{j=1}^K$ *of* $[0,1]$, *such that for every* $1 \leq i,j \leq K$ *the following holds:*

$$|D_j| \;=\; \frac{1}{K}\,, \qquad \text{and} \qquad \widetilde{a}(\phi,\theta) \;\geq\; \varphi Z_{ij}\,, \quad \text{whenever} \quad (\phi,\theta) \in D_i \times D_j\,. \tag{21.4}$$

Then the conclusions of Theorem 15.16 hold except that the exponential decay of off-diagonal resolvent elements (15.48) *is replaced by the weaker decay:*

$$|q_x(z)| \;\lesssim\; (1+|x|)^{-\beta}\,, \qquad x \in \mathbb{T}\,. \tag{21.5}$$

If $\mathbf{A}$ satisfies (15.39) and (15.40) then the comparison relations (cf. Convention 15.15) hold w.r.t. the model parameters ν,ξ. On the other hand, if $\mathbf{A}$ is good then L,α,β,K,φ are considered as model parameters.

The proof of Theorem 15.16 splits into three relatively independent parts. In the first part we show how to make $\mathbf{H}$ into an almost Wigner type matrix by changing basis. In the second part we describe how the proofs for Wigner type matrices are modified in order to accommodate some extra dependence of the transformed matrix. In the third part we show that the assumptions on the correlation matrix $\mathbf{A}$ imply that the QVE (15.42) has a bounded and sufficiently regular solution $\mathbf{m}$ using the general theory developed in Part II of this work. Finally, in Subsection 21.4 we combine these steps with the main results of this paper to prove Theorem 15.16.

21.1 Mapping H into Wigner type matrix by change of basis

The (discrete) Fourier transform of a matrix $\mathbf{T} = (t_{ij})_{i,j\in\mathbb{T}}$ is another matrix $\widehat{\mathbf{T}} = (\widehat{t}_{\phi\theta})_{\phi,\theta\in\mathbb{S}}$ defined by

$$\widehat{t}_{\phi\theta} := \frac{1}{N}\sum_{x,y\in\mathbb{T}} \mathrm{e}^{\mathrm{i}2\pi(\phi x-\theta y)} t_{xy}\,. \tag{21.6}$$

Since the mapping $\mathbf{T}\mapsto\widehat{\mathbf{T}}$ corresponds to the conjugation by the unitary matrix $\mathbf{F} = (f_{xy})_{x,y\in\mathbb{T}}$, with elements

$$f_{xy} := \frac{1}{N^{1/2}}\mathrm{e}^{\mathrm{i}2\pi xy}\,, \tag{21.7}$$

the matrices $\mathbf{T}$ and $\widehat{\mathbf{T}} = \mathbf{F}\mathbf{T}\mathbf{F}^*$ have the same spectrum,

$$\mathrm{Spec}(\mathbf{T}) = \mathrm{Spec}(\widehat{\mathbf{T}})\,. \tag{21.8}$$

The next result shows that the discrete Fourier transform maps Gaussian translation invariant random matrices into Wigner type random matrices with an extra dependence.

Definition 21.2 (4-fold correlated ensemble). *A random matrix* $\mathbf{H}$ *indexed by a torus is* **4-fold correlated** *if* h_{ij} *and* h_{kl} *are independent unless*

$$(k,l) \in \big\{(i,j),(j,i),(-i,-j),(-j,-i)\big\}\,. \tag{21.9}$$

Lemma 21.3 (Fourier transform). *Let* $\mathbf{H}$ *be a (not necessarily Gaussian) random matrix satisfying* (15.38)*. Then the elements of its Fourier transform* $\widehat{\mathbf{H}}$ *satisfy*

$$\mathbb{E}\,\widehat{h}_{ij} = 0 \tag{21.10a}$$

$$\mathbb{E}\,\widehat{h}_{ij}\overline{\widehat{h}_{kl}} = \widehat{a}_{ij}\,\delta_{ik}\,\delta_{jl} + \widehat{b}_{ij}\,\delta_{i,-l}\,\delta_{j,-k}\,, \tag{21.10b}$$

for every $i,j,k,l\in\mathbb{S}$*. If additionally* $\mathbf{H}$ *is Gaussian, then* $\widehat{\mathbf{H}}$ *is* 4*-fold correlated.*

We remark that if $\mathbf{A}$ is good, then (21.3) implies $\widehat{a}_{ij}\le N^{-1}$.

Proof. The proof of (21.10) is a straightforward computation. We omit further details. From (21.10b) we see that covariances between $\mathrm{Re}\,h_{ij}$, $\mathrm{Im}\,h_{ij}$ and $\mathrm{Re}\,h_{kl}$, $\mathrm{Im}\,h_{kl}$ can be non-zero if and only if (21.9) holds. Since $\widehat{\mathbf{H}}$ is Gaussian this implies the statement about the independence. □

The following result shows a practical way to construct real symmetric random matrices with translation invariant correlation structure. A similar, but slightly more complicated convolution representation exists for complex Hermitian random matrices.

Lemma 21.4. *Suppose a real symmetric matrix* $\mathbf{A}$ *satisfies the following Bochner type condition*

$$\sum_{i,j,k,l\in\mathbb{T}} \overline{q_{ij}}\,a_{i-k,j-l}\,q_{kl} \ge 0\,, \tag{21.11}$$

for any matrix $\mathbf{Q} = (q_{ij})_{i,j\in\mathbb{T}}$*, and define a symmetric filter matrix* $\mathbf{R} = (r_{ij})_{i,j\in\mathbb{T}}$*, by*

$$r_{xy} := \frac{1}{N^{3/2}}\sum_{\phi,\theta\in\mathbb{S}} \mathrm{e}^{-\mathrm{i}2\pi(x\phi-y\theta)}\sqrt{\widehat{a}_{\phi\theta}}\,. \tag{21.12}$$

If $\mathbf{W}$ *is GOE random matrix, then the random matrix* $\mathbf{H}$ *with elements,*

$$h_{ij} := \sum_{k,l\in\mathbb{T}} r_{i-k,j-l}\,w_{kl}\,, \tag{21.13}$$

has the correlation structure (15.38) *with* $\mathbf{B} = \mathbf{A}$.

21.2 Local law for 4-fold correlation

In this subsection we sketch how to prove a local law for the elements of the resolvent

$$\widehat{\mathbf{G}}(z) := (\widehat{\mathbf{H}} - z)^{-1}, \tag{21.14}$$

much the same way as we did for the Wigner type matrices. Indeed, the analysis is the same as before, but due to the extra correlation between (ϕ, θ) and $(-\phi, -\theta)$ we have to remove both the rows and columns corresponding to indices ϕ and $-\phi$ from $\widehat{\mathbf{H}}$ in order to make it independent of a given row ϕ. We state a local law for a general self-adjoint random matrix with independent entries apart from a possible correlation of the entries with indices (i, j) and $(-i, -j)$.

Theorem 21.5 (Local law for 4-fold correlation). *Let $\mathbf{H} = \mathbf{H}^* = (h_{ij})_{i,j\in\mathbb{T}}$ be a self-adjoint 4-fold correlated random matrix with centred entries. Let $\mathbf{S}$ be its variance matrix, defined as in* (15.2)*. Suppose $\mathbf{H}$ satisfies assumptions (A)-(D) and*

$$\mathbb{E}\, h_{ij} h_{-j,-i} = 0, \qquad i \neq j. \tag{21.15}$$

Then the conclusions of Theorem 15.6 hold for the resolvent elements of $\mathbf{H}$.

Proof. We sketch the proof of Theorem 21.5 by following the proof of the local law, Theorem 15.6, for random matrices without the 4-fold correlation and pointing out the necessary modifications. For simplicity we restrict to bounded values of z, i.e., we assume $|z| \leq 10$, say. The independence of the entries of $\mathbf{H}$ was used in the proofs of Lemma 16.1 and of the fluctuation averaging, Theorem 17.5. We will now show, how these two results are established for the 4-fold correlated case.

Adaptation of Lemma 16.1: The diagonal resolvent elements $G_{ii} = g_i$ still satisfy the perturbed QVE (16.1). This equation is derived using the Schur formula (17.16) with the perturbation

$$d_k = \sum_{i,j}^{(k)} h_{ki} G_{ij}^{(k)} h_{jk} - \sum_i s_{ki} G_{ii} - h_{kk}. \tag{21.16}$$

We will derive a different representation for $\mathbf{d}$, from which we will establish the bound

$$|d_k|\, \mathbb{1}(\Lambda \leq \lambda_*) \prec \Lambda_{\mathrm{o}} + \frac{1}{\sqrt{N}}. \tag{21.17}$$

This is in analogy to (16.19) in the proof of Lemma 16.1. To apply large deviation estimates in expressions such as (21.16) we expand the resolvent elements $G_{ij}^{(k)}$ such that their dependence on both h_{ki} and h_{jk} in the sum becomes explicit.

If $k = -k$ then we can proceed as in the standard case. In the case $k \neq -k$ we split the sum into parts according to whether the summed over indices i, j coincide with $-k, k$ or not:

$$\begin{aligned} \sum_{i,j}^{(k)} h_{ki} G_{ij}^{(k)} h_{jk} &= \sum_{i,j}^{[k]} h_{ki} G_{ij}^{(k)} h_{jk} \\ &\quad + h_{k,-k} \Bigg(\sum_j^{(k)} G_{-k,j}^{(k)} h_{jk} \Bigg) + \Bigg(\sum_i^{(k)} h_{ki} G_{i,-k}^{(k)} \Bigg) h_{-k,k} - h_{k,-k} G_{-k,-k}^{(k)} h_{-k,k}. \end{aligned} \tag{21.18}$$

Here, the upper index $[k]$ on the sum indicates that it runs over all indices except k and $-k$. Then we use the resolvent identity (16.9) for removing the $-k$ index from $G^{(k)}_{ij}$ and find

$$\sum_{i,j}^{[k]} h_{ki} G^{(k)}_{ij} h_{jk} \,=\, \sum_{i,j}^{[k]} h_{ki} G^{[k]}_{ij} h_{jk} + \frac{1}{G^{(k)}_{-k,-k}} \Bigg(\sum_{i}^{[k]} h_{ki} G^{(k)}_{i,-k} \Bigg) \Bigg(\sum_{j}^{[k]} G^{(k)}_{-k,j} h_{jk} \Bigg) . \tag{21.19}$$

Now we apply the general resolvent identities

$$\sum_{i}^{(k)} h_{ki} G^{(k)}_{ij} \,=\, -\frac{G_{kj}}{G_{jj}} \qquad \text{and} \qquad \sum_{j}^{(k)} G^{(k)}_{ij} h_{jk} \,=\, -\frac{G_{ik}}{G_{ii}} \,, \tag{21.20}$$

to all the terms in the parenthesis of (21.18) and (21.19). With the notation $\mathbf{G}^{[k]} := \mathbf{G}^{(-k,k)}$ we arrive at the result

$$\begin{aligned} \sum_{i,j}^{(k)} h_{ki} G^{(k)}_{ij} h_{jk} \,=\, & \sum_{i,j}^{[k]} h_{ki} G^{[k]}_{ij} h_{jk} + \frac{1}{G^{(k)}_{-k,-k}} \bigg(\frac{G_{k,-k}}{G_{kk}} + h_{k,-k} G^{(k)}_{-k,-k} \bigg) \bigg(\frac{G_{-k,k}}{G_{kk}} + h_{-k,k} G^{(k)}_{-k,-k} \bigg) \\ & - h_{k,-k} \frac{G_{-k,k}}{G_{kk}} - h_{-k,k} \frac{G_{k,-k}}{G_{kk}} - h_{k,-k} G^{(k)}_{-k,-k} h_{-k,k} \,. \end{aligned}$$

Therefore, we have shown that the diagonal resolvent elements satisfy the perturbed QVE (16.1) where the error vector $\mathbf{d}$ is given by

$$\begin{aligned} d_k \,=\, & \sum_{i \neq j}^{[k]} h_{ki} G^{[k]}_{ij} h_{jk} + \sum_{i}^{[k]} (|h_{ki}|^2 - s_{ki}) G^{[k]}_{ii} + \sum_{i}^{[k]} s_{ki} (G^{[k]}_{ii} - G_{ii}) - h_{kk} \\ & - s_{kk} G_{kk} - s_{k,-k} G_{-k,-k} + \frac{1}{G^{(k)}_{-k,-k}} \bigg(\frac{G_{k,-k}}{G_{kk}} + h_{k,-k} G^{(k)}_{-k,-k} \bigg) \bigg(\frac{G_{-k,k}}{G_{kk}} + h_{-k,k} G^{(k)}_{-k,-k} \bigg) \\ & - h_{k,-k} \frac{G_{-k,k}}{G_{kk}} - h_{-k,k} \frac{G_{k,-k}}{G_{kk}} - h_{k,-k} G^{(k)}_{-k,-k} h_{-k,k} \,. \end{aligned} \tag{21.21}$$

Then we follow the strategy of the proof of Lemma 16.1. We estimate the first two terms by using large deviation bounds. The other summands are bounded directly in terms of Λ_{o} and inverse powers of N using the still valid bound (16.10) for removing upper indices. In this way we arrive at (21.17).

To estimate the generic off-diagonal elements, G_{kl} with $k \neq l$ and $k \neq -l$, we modify the proof of Lemma 16.1 in a similar fashion. Starting from the identity (16.3) we remove the indices $-k$ and $-l$ from the resolvent elements $G^{(kl)}_{ij}$ there. For that we use the resolvent identity (16.9). Afterwards we use (21.20) in the same way as we did for d_k, apply the large deviation estimate and arrive at

$$|G_{kl}|\, \mathbb{1}(\Lambda \leq \lambda_*) \,\prec\, \sqrt{\frac{\operatorname{Im}\langle \mathbf{g} \rangle}{N\eta}} + \frac{1}{\sqrt{N}} \,.$$

Finally, we treat the special off-diagonal elements, $G_{k,-k}$, separately. The case $k = -k$ is already dealt with in the diagonal case so we assume here $k \neq -k$. From (16.3) we get

$$G_{k,-k} \,=\, G_{kk} G^{(k)}_{-k,-k} \sum_{i,j}^{[k]} h_{ki} G^{[k]}_{ij} h_{j,-k} - G_{kk} h_{k,-k} \,. \tag{21.22}$$

The entries of $\mathbf{H}$ appearing in the sum are independent provided $j \neq -i$. Thus we split the double sum into two parts,

$$\sum_{i,j}^{[k]} h_{ki} G_{ij}^{[k]} h_{j,-k} = \sum_{i \neq -j}^{[k]} h_{ki} G_{ij}^{[k]} h_{j,-k} + \sum_{i}^{[k]} h_{ki} h_{-i,-k} \, G_{i,-i}^{[k]} . \tag{21.23}$$

Here the first term on the right hand side can be bounded using the large deviation estimate (16.7b) as in the proof of Lemma 16.1. For the last term on the other hand we use (21.15). In particular, the family of random variables $(h_{ki} h_{-i,-k})_i$ are centred and independent. Thus, we can apply another large deviation estimate and find

$$\left| \sum_{i}^{[k]} h_{ki} \overline{h}_{-k,-i} \, G_{i,-i}^{[k]} \right| \prec \frac{1}{\sqrt{N}} \left(\frac{1}{N} \sum_{i}^{[k]} \left| G_{i,-i}^{[k]} \right|^2 \right)^{1/2} .$$

Altogether we see that also the resolvent elements $G_{k,-k}$ are small,

$$|G_{k,-k}| \, \mathbb{1}(\Lambda \leq \lambda_*) \prec \sqrt{\frac{\operatorname{Im}\langle \mathbf{g} \rangle}{N\eta}} + \frac{1}{\sqrt{N}} .$$

Therefore, the analog of Lemma 16.1 is proven.

Adaptation of Theorem 17.5: The fluctuation averaging is proven again following closely the arguments in the proof of Theorem 4.7 in [26]. One simply changes the equivalence relation given within the proof in the following way: For a given $\mathbf{k} = (k_1, \ldots, k_p) \in \mathbb{T}^p$ and $r, s \in \{1, \ldots, p\}$ we define $r \sim s$ if $k_r = k_s$ or $k_r = -k_s$. This means that for each 'lone index' k one removes the index $-k$ in addition to k from the other resolvent elements within the same monomial. For a more detailed description of the necessary modifications see the proof of Theorem 4.6 in [3].

Finally we explain the proof of the isotropic law under the 4-fold correlation structure.

Adaptation of Theorem 15.12: As we explained in the proof of Theorem 15.12, the main step is to estimate a high moment of $\mathcal{Z}$ via a resolvent expansion that decouples the set of indices b_{kj} of the vector $\mathbf{v}$ from the indices of resolvents of maximally expanded minors. In this way, by taking partial expectation of the expanded rows and columns, we collapse the $2p$-fold summation in (20.1) to a p-fold summation. The same procedure works under the 4-fold correlation structure with two minor modifications. First, we change the definition of being maximally expanded; we require that whenever we build a minor $H^{(k)}$ by removing the k-th row/column, we also remove its companion index $-k$. This will guarantee the independence of the expanded rows and columns from the remaining resolvents of minors. Second, when we perform the partial expectations, more pairing may occur, since an expanded matrix element h_{ab} may be paired not only with h_{ba} but also with $h_{-a,-b}$ and $h_{-b,-a}$. This results in not more than three times as many terms as before, but the $2p$-fold summation is still reduced to a p-fold summation. To illustrate this difference, as a toy second moment calculation (cf. (5.1) of [12]) without the 4-fold correlation we have

$$\mathbb{E} \sum_{a \neq b} \sum_{c \neq d} \sum_{xy} \left(\bar{v}_a h_{ax} G_{xy}^{(abcd)} h_{yb} v_b \right) \left(\bar{v}_c h_{cu} G_{uv}^{(abcd)*} h_{vd} v_d \right)$$

$$= \sum_{a \neq b} |v_a|^2 |v_b|^2 \mathbb{E} \frac{1}{N^2} \sum_{xy} |G_{xy}^{(abcd)}|^2 \lesssim \|\mathbf{v}\|_2^4,$$

where only the $a = d$, $b = c$ index-pairing is possible. With the 4-fold correlation a may be paired with $-d$ and b with $-c$ as well. This gives additional terms of the form

$$\sum_{a \neq b} \big(|v_a|^2 + |v_a||v_{-a}|\big)\big(|v_b|^2 + |v_b||v_{-b}|\big)$$

but with a simple Schwarz inequality they all can be estimated by powers of ℓ^2-norms. With these two modifications, the original proof of the isotropic law goes through. $\square$

21.3 Properties of QVE

In this subsection we show that a good correlation matrix $\mathbf{A}$ implies that the corresponding QVE (15.42) has a well behaving uniformly bounded solution everywhere. Then we prove that an exponentially decaying and non-resonant correlation matrix is good. Finally, we show that the quantity $q_{x-y}(z)$ describing the limit of the off-diagonal resolvent elements G_{x-y} (cf. (15.46)) has the right decay properties in $|x - y|$.

Recall the definition (21.1) of the integral operator $\widetilde{A}$. We consider the continuous version,

$$-\frac{1}{\widetilde{m}(z)} = z + \widetilde{A}\widetilde{m}(z)\,, \tag{21.24}$$

of the discrete QVE (15.42).

In the following we will use several results from Part II of this work on the properties of the general QVE defined on a probability space $(\mathfrak{X}, \pi)$ with an operator S in two different setups. When we discuss the discrete QVE (15.42) the setup is

$$\mathfrak{X} := \mathbb{S}\,, \qquad \pi := \frac{1}{N}\sum_{\phi \in \mathbb{S}} \delta_\phi \qquad \text{and} \qquad S := N\widehat{\mathbf{A}} \;\text{ i.e., } S_{\phi\theta} := N\,\widehat{a}_{\phi\theta}\,. \tag{21.25a}$$

For the continuous QVE (21.24) the setup is

$$\mathfrak{X} := [0,1]\,, \qquad \pi(\mathrm{d}\phi) := \mathrm{d}\phi \qquad \text{and} \qquad S := \widetilde{A} \;\text{ i.e., } S_{\phi\theta} := \widetilde{a}(\phi,\theta)\,. \tag{21.25b}$$

In the sequel L^p-norms and the scalar products are understood in the appropriate probability space $(\mathfrak{X}, \pi)$.

Lemma 21.6 (Good correlation matrix implies bounded solution). *If $\mathbf{A}$ is good then the continuous QVE* (21.24) *has a unique and uniformly bounded solution $\widetilde{m}(z) : [0,1] \to \overline{\mathbb{H}}$ that satisfies*

$$\sup_{\phi \in [0,1]} \Big(|\widetilde{m}(z;\phi)| + |\partial_\phi \widetilde{m}(z;\phi)|\Big) \lesssim 1\,, \qquad \forall\, z \in \mathbb{H}\,. \tag{21.26}$$

Furthermore, the discrete QVE (15.42) *also has a unique bounded solution $\mathbf{m}$ that is close to $\widetilde{m}$ in the sense that*

$$\sup_{\phi \in \mathbb{S}} |m_\phi(z) - \widetilde{m}(z;\phi)| \lesssim \min\left\{\frac{1}{\sqrt{N}}, \frac{1}{\mathrm{dist}(z,\{\sigma_*, -\sigma_*\})^{1/2} N}\right\}, \qquad \forall\, z \in \mathbb{H}\,. \tag{21.27}$$

Proof. We will first consider the continuous QVE. A straightforward computation using the definition (21.2) yields

$$|\widetilde{a}(\phi,\theta)| + |\partial_\phi \widetilde{a}(\phi,\theta)| \le 2\pi \sum_{x,y=0}^{N-1} (1 + |x|)\,|a_{xy}| \lesssim 1\,. \tag{21.28}$$

For the norm of an operator from $\mathrm{L}^2[0,1]$ to $\mathrm{L}^\infty[0,1]$ we find

$$\|\widetilde{A}\|^2_{\mathrm{L}^2\to\mathrm{L}^\infty} \,=\, \operatorname*{ess\,sup}_{\phi} \int_0^1 |\widetilde{a}(\phi,\theta)|^2 \mathrm{d}\theta \,\lesssim\, 1\,. \tag{21.29}$$

Next we show that $m(z)$ is uniformly bounded for $z \neq 0$. Indeed, using (21.28) we get

$$\|\widetilde{a}(\phi_1,\,\bullet\,) - \widetilde{a}(\phi_2,\,\bullet\,)\|_2 \,\le\, C_2|\phi_1-\phi_2|\,, \qquad \forall\,\phi_1,\phi_2 \in [0,1]\,.$$

From this it follows that

$$\lim_{\varepsilon\to 0}\inf_{\phi_1\in[0,1]} \int_0^1 \frac{\mathrm{d}\phi_2}{(\varepsilon + \|\,\widetilde{a}(\phi_1,\,\bullet\,) - \widetilde{a}(\phi_2,\,\bullet\,)\|_2)^2} \,=\, \infty\,.$$

This bound together with (21.29) makes it possible to apply (i) of Theorem 4.1. The theorem shows that $\|m(z)\|_\infty \le C(\delta)$ for any $|z| \ge \delta$ with $C(\delta)$ depending on $\delta > 0$

Now we apply the results of Part II using the setup (21.25b). The property (iii) of $\mathbf{A}$ in the definition of goodness (cf. Proposition 21.1) is equivalent to property **B1.** in Part II. Hence by (ii) of Theorem 9.1 from Part II of this work $\widetilde{m}(z)$ is uniformly bounded in some c_2-sized neighbourhood of $z=0$. Combining this with the uniform bound away from $z=0$ we get the uniform bound for $|\widetilde{m}(z;\phi)|$. In order to estimate also the derivative $\partial_\phi\widetilde{m}(z;\phi)$ we differentiate the continuous QVE (21.24) and get

$$\partial\widetilde{m}(z;\phi) = \widetilde{m}(z;\phi)^2 \int_0^1 \mathrm{d}\theta\, \widetilde{m}(z;\theta)\, \partial_\phi \widetilde{a}(\theta,\phi)\,. \tag{21.30}$$

Using (21.28) and the uniform boundedness of $\widetilde{m}$ we finish the proof of (21.26).

In order to prove the boundedness of the solution $\mathbf{m}$ as well we consider (15.42) as a perturbation of (21.24). Given $\mathbf{m}$ we first embed $\mathbb{S}$ into $[0,1]$ canonically, and define piecewise constant functions

$$\begin{aligned} g(z;\phi) &:= m_{N^{-1}\lfloor N\phi\rfloor}(z) \\ t(\phi,\theta) &:= N\,\widehat{a}_{N^{-1}\lfloor N\phi\rfloor,\, N^{-1}\lfloor N\theta\rfloor}\,, \end{aligned} \tag{21.31}$$

for every $\phi,\theta \in [0,1]$. Notice that $t(\phi,\theta) = \widetilde{a}(\phi,\theta)$, when $\phi,\theta \in \mathbb{S}$. Together with (21.28) this implies

$$|\,t(\phi,\theta) - \widetilde{a}(\phi,\theta)| \,\lesssim\, N^{-1}\,, \qquad \phi,\theta \in [0,1]\,. \tag{21.32}$$

In terms of these quantities (15.42) can be written as

$$-\frac{1}{g} \,=\, z + Tg\,, \tag{21.33}$$

where T is the integral operator with kernel $t(\phi,\theta)$. We clearly have

$$-\frac{1}{g} \,=\, z + \widetilde{A}g + d\,, \qquad \text{where} \quad d := (T-\widetilde{A})g\,. \tag{21.34}$$

The function d is a small perturbation of the continuous QVE. To see this we use (21.32) to get

$$\|d\|_\infty \le \|T-\widetilde{A}\|_{\mathrm{L}^2\to\mathrm{L}^\infty}\|g\|_2 \,\lesssim\, N^{-1}\|g\|_2\,. \tag{21.35}$$

Clearly, $\|T\|_{\mathrm{L}^2\to\mathrm{L}^\infty} \sim 1$ as well. Hence, we know from the general theory (cf. the bound (6.9) of Theorem 6.1 in Part II) that $\|g(z)\|_2 \le 2/|z|$. Using (21.32) we see that for sufficiently large

N the operator T is also block fully indecomposable with the same matrix $\mathbf{Z}$ and the partition $\mathcal{D}$ as $\widetilde{A}$. Thus we get $\|g(z)\|_2 \lesssim 1$ for all z by (ii) of Theorem 9.1 in Part II. Combining this with (21.35) yields

$$\|d(z)\|_\infty \lesssim N^{-1} . \tag{21.36}$$

We show that this implies that also the solutions g and $\widetilde{m}$ are close in the sense of (21.27). For this purpose we use Theorem 6.10 from Part II. By the rough stability statement of that theorem we find

$$\|g(z) - \widetilde{m}(z)\|_\infty \mathbb{1}\big(\|g(z) - \widetilde{m}(z)\|_\infty \leq \lambda_*\big) \lesssim N^{-1}, \qquad \operatorname{dist}(z, \{\sigma_*, -\sigma_*\}) \geq c_0 , \tag{21.37}$$

where c_0 is an arbitrarily small constant. This means that we get stability as long as we stay away from the edges σ_* and $-\sigma_*$ of the support of the density of states. The necessary initial bound inside the indicator function is satisfied for large enough values of $|z|$, since

$$\|g(z)\|_\infty + \|\widetilde{m}(z)\|_\infty \lesssim |z|^{-1} , \qquad |z| \geq C_1 .$$

Here C_1 is a sufficiently large constant. This bound follows from the Stieltjes transform representation of both the solution of the discrete and the continuous QVE (cf. Theorem 6.1 of Part II). We use continuity of g and $\widetilde{m}$ in z and (21.37) to propagate the initial bound from the regime of large values of $|z|$ to all $z \in \mathbb{H}$ with $\operatorname{dist}(z, \{\sigma_*, -\sigma_*\}) \geq c_0$. In particular, (21.37) remains true even without the indicator function, i.e.,

$$\|g(z) - \widetilde{m}(z)\|_\infty \lesssim N^{-1}, \qquad \operatorname{dist}(z, \{\sigma_*, -\sigma_*\}) \geq c_0 . \tag{21.38}$$

It remains to show (21.27) close to the edges. We restrict to the case $|z - \sigma_*| \leq c_0$, close to the right edge. The left edge is treated in the same way. For the following analysis we use Theorem 18.2 in the continuous setup (cf. Proposition 13.1 in Part II). The theorem yields

$$\|g - \widetilde{m}\|_\infty \mathbb{1}\big(\|g - \widetilde{m}\|_\infty \leq \lambda_*\big) \lesssim \Theta + N^{-1} , \tag{21.39}$$

where the quantity $\Theta = \Theta(z) \geq 0$ satisfies

$$\big|\Theta^3 + \pi_2 \Theta^2 + \pi_1 \Theta\big| \, \mathbb{1}\big(\|g - \widetilde{m}\|_\infty \leq \lambda_*\big) \lesssim N^{-1}. \tag{21.40}$$

By (18.10) the coefficients of the cubic equation satisfy

$$|\pi_1| \sim \frac{\eta}{\rho} + \rho , \qquad \text{and} \qquad |\pi_2| \sim 1 ,$$

because $\sigma \sim 1$ if c_0 is sufficiently small. From (18.5d) in Theorem 18.1 we see how ρ behaves in the neighbourhood of the right edge. Thus, we infer that for small enough λ_* and $z = \sigma_* + \omega + \mathrm{i}\eta$,

$$\Theta \, \mathbb{1}\big(\|g - \widetilde{m}\|_\infty \leq \lambda_*\big) \lesssim |\pi_1| + N^{-1/2} \sim (|\omega| + \eta)^{1/2} + N^{-1/2} ,$$

Plugging this back into (21.39) yields

$$\|g - \widetilde{m}\|_\infty \mathbb{1}\big(\|g - \widetilde{m}\|_\infty \leq \lambda_*\big) \lesssim (|\omega| + \eta)^{1/2} + N^{-1/2}.$$

By choosing the size of the neighbourhood, c_0, around the edge small enough we ensure that this bound implies a gap in the possible values of the continuous function $z \mapsto \|g(z) - \widetilde{m}(z)\|_\infty$. Since on the boundary, $|z - \sigma_*| = c_0$, the initial bound, $\|g - \widetilde{m}\|_\infty \leq \lambda_*$, holds by (21.38), it propagates to all z with $|z-\sigma_*| \le c_0$. Thus (21.39) and (21.40) remain true without the indicator functions.

It still remains to bound Θ in (21.39). Since $|\pi_2| \sim 1$, we may absorb the cubic term in Θ in (21.40). We find that Θ satisfies

$$|\Theta^2 + \varpi_1 \Theta| \lesssim N^{-1}\,, \qquad |\varpi_1| \sim (|\omega|+\eta)^{1/2}\,., \tag{21.41}$$

where $\varpi := \pi_1/(\pi_2+\Theta)$. Now we distinguish two cases. First let $(|\omega|+\eta)^{1/2} \le C_2 N^{-1/2}$ for some sufficiently large constant C_2. Then (21.41) implies

$$\Theta \lesssim \frac{1}{\sqrt{N}} \lesssim \min\biggl\{ \frac{1}{\sqrt{N}}\,, \frac{1}{(|\omega|+\eta)^{1/2} N} \biggr\}.$$

If on the other hand $(|\omega|+\eta)^{1/2} \ge C_2 N^{-1/2}$, then (21.41) implies

$$\Theta\, \mathbb{1}(\Theta \le N^{-1/2}) \lesssim \frac{1}{(|\omega|+\eta)^{1/2} N}\,. \tag{21.42}$$

By choosing C_2 large enough this becomes

$$\Theta\, \mathbb{1}(\Theta \le N^{-1/2}) \le \frac{1}{2\sqrt{N}}\,.$$

In particular, there is a gap in the possible values of Θ which allows us to propagate the bound $\Theta \le N^{-1/2}$ from the boundary of the neighborhood of the edge to all z with $|z-\sigma_*| \le c_0$. Thus, (21.42) holds without the indicator function and we see that

$$\Theta \lesssim \min\biggl\{ \frac{1}{\sqrt{N}}\,, \frac{1}{(|\omega|+\eta)^{1/2} N} \biggr\}.$$

Using this bound in (21.39) without the indicator function proves the bound (21.27) at the right edge. □

Lemma 21.7 (Non-resonant and exponentially decaying correlation matrix is good). *If* $\mathbf{A}$ *satisfies* (15.39) *and* (15.40) *then* $\mathbf{A}$ *is good. In particular, the constants* $L, \alpha, \beta, K, \varphi$ *depend only on the model parameters* ξ, ν.

The proof of this lemma relies on the following technical result that is proven in the appendix. Let us denote the complex strip of width $\nu > 0$ by

$$\mathbb{R}_\nu = \mathbb{R} + \mathrm{i}(-\nu, +\nu)\,. \tag{21.43}$$

Lemma 21.8 (Jensen-Poisson bound). *Suppose* $f : \mathbb{R}_\nu \to \mathbb{C}$ *is an analytic function satisfying*

$$\sup_{\zeta \in \mathbb{R}_\nu} |f(\zeta)| \le C_1 \qquad \text{and} \qquad \int_0^1 |f(\phi)|\,\mathrm{d}\phi \ge 1\,. \tag{21.44}$$

Then for every $\varepsilon > 0$ *there exists* $\delta > 0$ *depending only on* ε, ν, C_1 *such that*

$$\bigl|\bigl\{\phi \in [0,1] : |f(\phi)| \ge \delta\bigr\}\bigr| \ge 1-\varepsilon\,. \tag{21.45}$$

Proof of Lemma 21.7. The non-resonance condition (15.40) guarantees that the $\mathrm{L}^1[0,1]$-norms of the row functions $\theta \mapsto \widetilde{a}(\phi,\theta)$ are uniformly bounded from below. Indeed, since $\widetilde{a}(\phi,\theta) \ge 0$, we have

$$\|\widetilde{a}(\phi, \bullet)\|_1 = \int_0^1 \widetilde{a}(\phi,\theta)\,\mathrm{d}\theta = \sum_j \mathrm{e}^{\mathrm{i}2\pi j\phi} a_{j0} \ge \xi\,. \tag{21.46}$$

The conditions (i) and (iii) of the good correlation matrix for $\mathbf{A}$ rely on the following property of $\widetilde{a}$: For any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\big|\big\{\theta \in [0,1] : \widetilde{a}(\phi,\theta) \geq \delta\big\}\big| \,\geq\, 1-\varepsilon\,, \qquad \forall\, \phi \in [0,1]\,. \tag{21.47}$$

This property of $\widetilde{a}$ follows by applying Lemma 21.8 to $f(\zeta) = \widetilde{a}(\phi,\zeta)$ with ϕ fixed.

The condition (i) holds with $L = 2$. Indeed, choosing $\varepsilon = 1/3$, we see that

$$\widetilde{a}^2(\phi,\theta) \,\geq\, \int_{D(\phi,\theta)} \widetilde{a}(\phi,\varphi)\widetilde{a}(\varphi,\theta)\mathrm{d}\varphi \,\geq\, \frac{\delta^2}{3}\,,$$

since the measure of the connecting set

$$D(\phi,\theta) \,:=\, \big\{\varphi : \widetilde{a}(\phi,\varphi) > \delta\big\} \cap \big\{\varphi : \widetilde{a}(\varphi,\theta) > \delta\big\}\,,$$

is at least $1 - 2\,\varepsilon = 1/3$ by (21.47) and the union bound.

Next we show that $\widetilde{A}$ is a block fully indecomposable operator. To this end we pick $\delta > 0$ and $\varepsilon > 0$ such that (21.47) holds. From (21.28) we see that

$$|\widetilde{a}(\phi_1,\theta_1) - \widetilde{a}(\phi_2,\theta_2)| \,\lesssim\, |\phi_1 - \phi_2| + |\theta_1 - \theta_2|\,, \tag{21.48}$$

for every $\phi_1, \phi_2, \theta_1, \theta_2 \in [0,1]$. Let $K \in \mathbb{N}$ be so large that

$$\big|\widetilde{a}(\phi_1,\theta_1) - \widetilde{a}(\phi_2,\theta_2)\big| \,\leq\, \frac{\delta}{2}\,, \quad \text{provided} \quad |\phi_1 - \phi_2| + |\theta_1 - \theta_2| \leq \frac{1}{K}\,.$$

Let us define a partition $\mathcal{D} = \{D_k\}_{k=1}^K$ of $[0,1]$ and a matrix $\mathbf{Z} = (z_{ij})_{i,j=1}^K$, by

$$D_k := \left[\frac{k-1}{K}, \frac{k}{K}\right) \qquad \text{and} \qquad z_{ij} := \mathbb{1}\bigg\{\max_{(\phi,\theta)\in D_i\times D_j} \widetilde{a}(\phi,\theta) \geq \delta\bigg\}\,. \tag{21.49}$$

Clearly, we have

$$\widetilde{a}(\phi,\theta) \,\geq\, \frac{\delta}{2}\, z_{ij}\,, \qquad (\phi,\theta) \in D_i \times D_j\,. \tag{21.50}$$

We will now show that $\mathbf{Z}$ is FID by proving that if there are two sets I and J such that $z_{ij} = 0$, for all $i \in I$ and $j \in J$, then

$$|I| + |J| \,\leq\, K - 1\,. \tag{21.51}$$

Denoting $D_I := \cup_{i\in I} D_i$, we have $\widetilde{a}(\phi,\theta) \leq \delta$ for $(\phi,\theta) \in D_I \times D_J$. Thus (21.47) implies

$$\frac{|I|}{K} = |D_I| \leq \varepsilon\,, \qquad \text{and} \qquad \frac{|J|}{K} = |D_I| \leq \varepsilon\,. \tag{21.52}$$

Choosing $\varepsilon \leq 1/3$ we see that $|I| + |J| \leq (2/3)K$, and (21.51) follows. Since $\mathbf{Z}$ is a FID, matrix we see that $\widetilde{A}$ is block fully indecomposable. $\square$

Lemma 21.9 (Decay of correlations). *If a non-resonant correlation matrix* $\mathbf{A}$ *is exponentially decaying, then* $q_x(z)$*, defined in* (15.47)*, also decays exponentially in* $|x|$ *such that* (15.48) *holds. If* $\mathbf{A}$ *is good, then* (21.5) *holds instead.*

PROOF. Recall that $\widetilde{m}(z)$ is the bounded solution of the continuous QVE (21.24). We will first prove the appropriate decay properties for the quantity

$$\widetilde{q}_x(z) := \langle e_x, \widetilde{m}(z) \rangle \,, \qquad x \in \mathbb{Z} \,, \tag{21.53}$$

instead of for $q_x(z)$.

Let us first assume that $\mathbf{A}$ is non-resonant and exponentially decaying. From (15.39) it follows that $\widetilde{a}$ from (21.2) has an analytic extension to the complex strip $\mathbb{R}_\nu$ (cf. (21.43)) where $\nu > 0$ is the exponent from (15.39). We will now show that $\widetilde{q}_x(z)$ decays exponentially in this case. To see this we consider the function $\Gamma(z) : \mathbb{R}_\nu \to \mathbb{C}$, defined by

$$\Gamma(z;\zeta) := -\left(z + \int_0^1 \widetilde{a}(\zeta,\phi) m(z;\phi) \,\mathrm{d}\phi \right)^{-1} . \tag{21.54}$$

In particular, it follows that $\widetilde{m}(z;\phi) = \Gamma(z;\phi)$ for every $\phi \in [0,1]$. Because $\widetilde{a}$ is uniformly continuous and the expression inside the parenthesis on the right hand side of (21.54) is bounded away from zero by a constant of size $(\sup_z \|m(z)\|_\infty)^{-1}$ there exists a constant $\nu' < \nu$ such that $|\Gamma(z;\zeta)| \le C_0$ for $\zeta \in \mathbb{R}_{\nu'}$. Since $\widetilde{a} : \mathbb{R}^2_{\nu'} \to \mathbb{C}$ is analytic also $\Gamma(z) : \mathbb{R}_{\nu'} \to \mathbb{C}$ is analytic. For any $x \in \mathbb{Z}$ we thus get by a contour deformation

$$\begin{aligned} \mathrm{e}^{2\pi\nu' x} \widetilde{q}_x(z) \;=\; \mathrm{e}^{2\pi\nu' x} \langle e_x, \widetilde{m}(z) \rangle \;&=\; \int_0^1 \mathrm{e}^{-\mathrm{i}2\pi x (\phi + \mathrm{i}\nu')} \,\Gamma(z;\phi) \,\mathrm{d}\phi \\ &=\; \int_0^1 \mathrm{e}^{-\mathrm{i}2\nu x \phi} \,\Gamma(z; \phi - \mathrm{i}\nu') \,\mathrm{d}\phi \,, \end{aligned} \tag{21.55}$$

where the integrals over the vertical line segments joining ± 1 and $\pm 1 - \mathrm{i}\nu'$ cancel each other due to periodicity of the integrand in the horizontal direction. Since $x \in \mathbb{Z}$ was arbitrary taking absolute values of (21.55) yields the exponential decay,

$$|\widetilde{q}_x(z)| \;\le\; \left(\sup_{\zeta \in \mathbb{R}_{\nu'}} |\Gamma(z;\zeta)| \right) \mathrm{e}^{-2\pi\nu'|x|} \;\le\; C_0 \,\mathrm{e}^{-2\pi\nu'|x|} \,, \qquad x \in \mathbb{T} \,. \tag{21.56}$$

Next we prove that if $\mathbf{A}$ is good, then also $\widetilde{q}_x(z)$ decays like $|x|^{-\beta}$ for large $|x|$. To this end let ∂ denote the derivative w.r.t. the variable in $[0,1]$. Using $e_x(\phi) = \mathrm{e}^{\mathrm{i}2\pi x \phi}$ we get

$$|x|^k |\widetilde{q}_x(z)| \;=\; (2\pi)^{-k} |\langle \partial^k e_x, \widetilde{m}(z) \rangle| \;=\; (2\pi)^{-k} |\langle e_x, \partial^k \widetilde{m}(z) \rangle| \le \|\partial^k \widetilde{m}(z)\|_\infty \,, \quad \forall x \in \mathbb{Z} \,. \tag{21.57}$$

Thus we need to show that $\|\partial^\beta \widetilde{m}(z)\|_\infty \lesssim 1$ uniformly in z. The proof is by induction on the number of derivatives of $\widetilde{m}$. It is based on

$$\partial^k \widetilde{m}(z;\phi) \;=\; \partial_\phi^{k-1} \left(\widetilde{m}(z;\phi)^2 \int \mathrm{d}\theta \,\widetilde{m}(z;\theta) \,\partial_\phi \widetilde{a}(\theta,\phi) \right),$$

which follows from (21.30), and the following consequence of (21.3):

$$\max_{j=0}^{\beta} \sup_{\phi,\theta \in [0,1]} |\partial_\phi^j \widetilde{a}(\phi,\theta)| \lesssim 1 \,. \tag{21.58}$$

Now we will show that $q_x(z)$ inherits its decay properties from $\widetilde{q}_x(z)$ so that

$$|q_x(z)| \;\le\; |\widetilde{q}_x(z)| + \frac{C}{\sqrt{\mathrm{dist}(z, \{\sigma_*, -\sigma_*\}) \,N}} \,, \qquad x \in \mathbb{T} \,. \tag{21.59}$$

This implies the estimates (15.48) and (21.5) under the corresponding assumptions on $\mathbf{A}$, respectively. Indeed, for any $x \in \{0,1,2,\dots,N-1\}$ we have

$$\begin{aligned}\big|\widetilde{q}_x(z)-q_x(z)\big| &\le \left|\int_0^1 \mathrm{e}^{-\mathrm{i}2\pi x\phi}\,\widetilde{m}(z;\phi)\mathrm{d}\phi - \frac{1}{N}\sum_{\theta\in\mathbb{S}}\mathrm{e}^{-\mathrm{i}2\pi x\theta}m_\theta(z)\right| \\ &\le \sum_{j=0}^{N-1}\int_{j/N}^{(j+1)/N}\Big|\mathrm{e}^{-\mathrm{i}2\pi x\phi}\widetilde{m}(z;\phi)-\mathrm{e}^{-\mathrm{i}2\pi x\frac{j}{N}}m_{j/N}(z)\Big| \\ &\lesssim \frac{1+x}{\operatorname{dist}(z,\{\sigma_*,-\sigma_*\})^{1/2}N}\,,\end{aligned} \tag{21.60}$$

where we have used (21.27) and (21.26). This proves (21.59) for $x \le N^{1/2}$.

For large $x \ge N^{1/2}$ we bound the size of $q_x = q_x(z)$ using the summation of parts

$$q_x \,=\, \frac{1}{N}\sum_{j=0}^{N-1}\mathrm{e}^{-\mathrm{i}2\pi x\frac{j}{N}}m_{j/N} \,=\, -\frac{1}{N}\sum_{j=1}^{N-2}\big(m_{(j+1)/N}-m_{j/N}\big)\sum_{k=0}^{j}\mathrm{e}^{-\mathrm{i}2\pi x\frac{j}{N}} \,+\, \mathcal{O}\Big(\frac{1}{N}\Big)\,,$$

where we have dropped two boundary terms of size $\mathcal{O}(N^{-1})$. Here $|m_{(j+1)/N}-m_{j/N}| \le C/N$, while the geometric sum is $\mathcal{O}(N/x)=\mathcal{O}(N^{1/2})$. Thus estimating each term in the sum over j separately shows that $|q_x(z)| \lesssim N^{-1/2}$. □

21.4 Proof of Theorem 15.16

Proof of Theorem 15.16 and Proposition 21.1. Let $\mathbf{H}$ be a Gaussian random matrix with translationally invariant correlation satisfying (15.38) for a good correlation matrix $\mathbf{A}$. In particular, if $\mathbf{A}$ is non-resonant and exponentially decaying with parameters ν and ξ then $\mathbf{A}$ is also good with the parameters L,α,β,K,φ depending only on ν and ξ (cf. Lemma 21.7). From now on we consider the case where $\mathbf{A}$ is good.

Since $\mathbf{A}$ is good Lemma 21.6 implies that the corresponding QVE (15.42) has a bounded solution $\mathbf{m}=(m_\phi)_{\phi\in\mathbb{S}}$. Moreover, since $\widetilde{a}=\widetilde{a}^{(N)} : [0,1]^2 \to [0,\infty)$ is a continuous function (uniformly in N) and $|N\widehat{a}_{\phi\theta}-\widetilde{a}(\phi,\theta)| \lesssim N^{-1}$ for $\phi,\theta\in\mathbb{S}$ (cf. (21.32)) we see that

$$\sup_{D\subset\mathbb{S}}\inf_{\substack{\phi_1\in D\\ \phi_2\notin D}}\sum_{\theta\in\mathbb{S}}\big|\widehat{a}_{\phi_1,\theta}-\widehat{a}_{\phi_2,\theta}\big| \,\le\, \frac{C}{N}+\sup_{D\subset[0,1]}\inf_{\substack{\phi_1\in D\\ \phi_2\notin D}}\int_0^1\big|\widetilde{a}(\phi_1,\theta)-\widetilde{a}(\phi_2,\theta)\big|\mathrm{d}\theta \,\lesssim\, \frac{1}{N}\,,$$

Now we apply the general theory from Part II of this work in the discrete setup (21.25a). Theorem 6.9 of Part II implies that ρ has the properties (i-iii) and (v) stated in Theorem 15.16. The property (iv) follows from (ii) of Theorem 6.2 of Part II.

By Lemma 21.3 the Fourier transform $\widehat{\mathbf{H}}$ of $\mathbf{H}$ has the correlation structure (21.10). In particular, $\widehat{\mathbf{H}}$ has the 4-fold symmetry. Moreover, from (21.10b) we read off that

$$\mathbb{E}\,\widehat{h}_{\phi\theta}\widehat{h}_{-\theta,\phi}=0\,,\qquad \forall\,\phi,\theta\in\mathbb{S}\,,\quad \phi\neq\theta\,.$$

Hence, Theorem 21.5 with $\widehat{\mathbf{H}}$ playing the role of $\mathbf{H}$ and $\mathbf{S}:=\widehat{\mathbf{A}}$ yields the averaged local law (15.41). Note that we have $\kappa(z)\le 1$ in (15.17) since $\operatorname{supp}\rho$ is a single interval and hence $\Delta(z)=1$ whenever $\tau=\operatorname{Re} z$ is close to the edges $\pm\sigma_*$.

In order to get (15.46) we use the isotropic local law (Theorem 15.12). Indeed, fix two arbitrary elements x and y of $\mathbb{T}$ and define two unit vectors $\mathbf{v}$ and $\mathbf{w}$ of $\mathbb{C}^{\mathbb{T}}$ by setting

$$v_\phi := N^{-1/2}\mathrm{e}^{\mathrm{i}2\pi x\phi} \qquad\text{and}\qquad w_\theta := N^{-1/2}\mathrm{e}^{\mathrm{i}2\pi y\theta}\,,$$

for every $\phi, \theta \in \mathbb{S}$. From (21.6) and (15.47) we see that

$$G_{xy}(z) = \langle \mathbf{v}, \widehat{\mathbf{G}}(z)\mathbf{w} \rangle \qquad \text{and} \qquad q_x(z) = \langle \mathbf{v}, \mathrm{diag}(\mathbf{m}(z))\mathbf{w} \rangle \,.$$

Thus the isotropic local law (15.33) implies (15.46). Furthermore, if $\mathbf{A}$ is non-resonant and exponentially decaying then $q_x(z)$ decays exponentially by Lemma 21.9. This proves (15.48). On the other hand, if $\mathbf{A}$ is good then Lemma 21.9 says that $q_x(z)$ satisfies (21.5). This completes the proof. □

A Appendix Part I

A.1 Non-centred diagonal entries

In this section we will discuss a generalisation to the model introduced in Section 2. This generalisation concerns assumption 2., namely that the entries of the random matrix, $\mathbf{H}$, are centred. In general, relaxing this assumption leads to a matrix equation for all N^2 resolvent entries instead of the vector equation (4.9) that only involves the diagonal entries of $\mathbf{G}$. But if the assumption $\mathbb{E}\, h_{ij} = 0$ is dropped only on the diagonal, then our theory can be applied essentially unchanged.

THEOREM A.1 (Non-centred diagonal). *Let $\mathbf{H}$ be a real symmetric random matrix that satisfies assumptions 1., 3. and 4. and suppose that the off-diagonal entries are centred,*

$$\mathbb{E}\, h_{ij} \;=\; g(i/N)\, \delta_{ij} \,, \qquad i,j = 1, \dots, N \,,$$

where $g : [0,1] \to \mathbb{R}$ is a Hölder-continuous function with Hölder-exponent $1/2$. Then the conclusions of Theorems 15.6 and 15.14 hold, except that m solves the equation

$$-\frac{1}{m(x;z)} \;=\; z - g(x) + \int_0^1 s(x,y) m(y;z) \,, \qquad x \in [0,1] \,,\; z \in \mathbb{H} \,, \tag{A.1}$$

instead of (6.5).

The reason for the extra term, $g(x)$, in the QVE (A.1) as compared to (6.5) can be seen from the resolvent expansion carried out in Section 4.1. Indeed, the term $-h_{kk}$ in the Schur's complement formula (4.4), was considered part of the error d_k because it satisfied $|h_{kk}| \prec N^{-1/2}$. However, now h_{kk} is not centred anymore and thus satisfies

$$|h_{kk} - g(k/N)| \;\prec\; \frac{1}{\sqrt{N}} \,.$$

Thus the diagonal resolvent entries now almost fulfil the modified QVE (A.1).

Other than the derivation of the QVE the proof of Theorem A.1 requires a change only in the argument from Section 4.3 that leads to the $\mathrm{L}^2[0,1]$-bound on m. The modification of this argument we present here is simpler than the original argument. In particular, it does not require a separate treatment of the neighbourhood of $z = 0$. The drawback of this approach is that it cannot be applied in the general setting that is considered in Part II of this work where the kernel $s(x,y)$ is block fully indecomposable.

The original argument fails because in(4.15) the z on the left hand side is replaced by $z - g(x)$ which cannot be inverted for $z \neq 0$. The new argument that we use to prove Theorem A.1 has three steps. First we prove an $\mathrm{L}^1[0,1]$-bound for m. Then this bound is used to show that m is uniformly bounded away from zero. Finally, we prove the $\mathrm{L}^2[0,1]$-bound by making use of the lower bound on $|m|$.

The operator F, defined in (4.14), still satisfies the bound $\sup \operatorname{Spec}(F) \leq 1$. The argument leading to this conclusion only makes use of the imaginary part of the QVE, which remains unchanged if z is replaced by $z-g(x)$. The spectral bound on F is used in the following simple argument, where we evaluate the quadratic form corresponding to F on the constant function

$$1 \geq \int_0^1 (F(z)1)(x) = \int_0^1 \int_0^1 s(x,y)|m(x;z)||m(y;z)|\,\mathrm{d}x\mathrm{d}y \geq \inf_{u,v} s(u,v)\Big(\int_0^1 |m(x;z)|\mathrm{d}x\Big)^2 .$$

Thus, $m(\,\cdot\,;z)$ is uniformly bound in $\mathrm{L}^1[0,1]$ uniformly in z on a compact set of the upper half plane. From the QVE (6.5) and the boundedness of the kernel s, we see that this $\mathrm{L}^1[0,1]$-bound implies a uniform upper bound on the inverse of m. Therefore, m is bounded away from zero.

Finally we evaluate $F(z)$ again on the constant function and use that the spectral bound implies a bound on $F(z)$ as an operator on $\mathrm{L}^2[0,1]$ as well,

$$1 \geq \|F(z)1\|_2^2 = \int_0^1\int_0^1\int_0^1 |m(x;z)||m(y;z)||m(u;z)|^2 s(x,u)s(u,y)\,\mathrm{d}u\,\mathrm{d}x\,\mathrm{d}y .$$

On the right hand side we make use of the uniform lower bounds on $|m|$ and s and find

$$\int_0^1 |m(u;z)|^2\mathrm{d}u \leq \frac{1}{\inf_{u,v} s(u,v)^2 \inf_x |m(x;z)|^2} .$$

B Appendix Part II

B.1 Behaviour of $\operatorname{Im}\langle m\rangle$

Corollary B.1 (Scaling relations). *Suppose the assumptions of Theorem 6.4 are satisfied. There exists a positive threshold $\varepsilon \sim 1$ such that for $\mathbb{M}$, defined as in the statement of that theorem, and any $\eta \in (0,\varepsilon]$, the average generating density has the following growth behaviour close to the points in $\mathbb{M}$:*

(a) **Support around an edge:** *At the edges α_i, β_{i-1} with $i=2,\dots,K'$,*

$$\big\langle \operatorname{Im} m(\alpha_i+\omega+\mathrm{i}\eta)\big\rangle \sim \big\langle \operatorname{Im} m(\beta_{i-1}-\omega+\mathrm{i}\eta)\big\rangle \sim \frac{(\omega+\eta)^{1/2}}{(\alpha_i-\beta_{i-1}+\omega+\eta)^{1/6}}, \qquad \omega\in[0,\varepsilon] .$$

(b) **Inside a gap:** *Between two neighbouring edges β_{i-1} and α_i with $i=2,\dots,K'$,*

$$\big\langle \operatorname{Im} m(\tau+\mathrm{i}\eta)\big\rangle \sim \frac{\eta}{(\alpha_i-\beta_{i-1}+\eta)^{1/6}}\bigg(\frac{1}{(\tau-\beta_{i-1}+\eta)^{1/2}}+\frac{1}{(\alpha_i-\tau+\eta)^{1/2}}\bigg), \quad \tau\in[\beta_{i-1},\alpha_i] .$$

(c) **Support around an extreme edge:** *Around the extreme points α_1 and $\beta_{K'}$ of $\operatorname{supp} v$:*

$$\big\langle \operatorname{Im} m(\alpha_1+\omega+\mathrm{i}\eta)\big\rangle \sim \big\langle \operatorname{Im} m(\beta_{K'}-\omega+\mathrm{i}\eta)\big\rangle \sim \begin{cases} (\omega+\eta)^{1/2}, & \omega\in[0,\varepsilon], \\ \dfrac{\eta}{(|\omega|+\eta)^{1/2}}, & \omega\in[-\varepsilon,0] . \end{cases}$$

(d) **Close to a local minimum:** *In a neighbourhood of the local minima $\{\gamma_k\}$ in the interior of the support of the generating density,*

$$\big\langle \operatorname{Im} m(\gamma_k+\omega+\mathrm{i}\eta)\big\rangle \sim \big\langle v(\gamma_k)\big\rangle + (|\omega|+\eta)^{1/3}, \qquad \omega\in[-\varepsilon,\varepsilon] .$$

All constants hidden behind the comparison relations depend on the parameters $(\rho, L, \|S\|_{\mathrm{L}^2 \to \mathscr{B}}, \Phi)$.

PROOF. To prove the lemma, we will use Theorem 6.4 and the Stieltjes transform representation of the solution of QVE. We start with the claim about the growth behaviour around the points $\{\gamma_k\}$. By the description of the shape of the generating density in Theorem 6.4 and because of $\Psi_{\min}(\lambda) \sim \min\{\lambda^2, |\lambda|^{1/3}\}$ (cf. (6.13b)), we have for small enough $\varepsilon \sim 1$:

$$\langle v(\gamma_k + \omega)\rangle \sim \rho_k + \min\{\omega^2/\rho_k^5, |\tau|^{1/3}\} \sim \rho_k + |\omega|^{1/3}, \qquad \omega \in [-2\varepsilon, 2\varepsilon].$$

The constant ρ_k is comparable to $\langle v(\gamma_k)\rangle$ by (6.19c). Thus, we find

$$\big\langle \mathrm{Im}\, m(\gamma_k + \omega + \mathrm{i}\eta)\big\rangle = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\eta\,\langle v(\tau)\rangle\, \mathrm{d}\tau}{\eta^2 + (\gamma_k + \omega - \tau)^2} \sim \langle v(\gamma_k)\rangle + \int_{-2\varepsilon}^{2\varepsilon} \frac{\eta\,|\tau|^{1/3}\,\mathrm{d}\tau}{\eta^2 + (\omega - \tau)^2}, \qquad \omega \in [-\varepsilon, \varepsilon].$$

The claim follows because the last integral is comparable to $(|\omega| + \eta)^{1/3}$ for any $\varepsilon \sim 1$.

Let us now consider the case, in which an edge is close by. We treat only the case of a right edge, i.e., the vicinity of β_i for $i = 1, \dots, K'$. For the left edge the argument is the same. Here, Theorem 6.4 and $\Psi_{\mathrm{edge}}(\lambda) \sim \min\{\lambda^{1/2}, \lambda^{1/3}\}$ (cf. (6.13a)) imply for small enough $\varepsilon \sim 1$:

$$\langle v(\beta_i - \omega)\rangle \sim \min\{\Delta^{-1/6}\omega^{1/2}, \omega^{1/3}\}, \qquad \omega \in [0, 2\varepsilon].$$

The positive constant Δ is comparable to the gap size, $\Delta \sim \alpha_{i+1} - \beta_i$, if β_i is not the rightmost edge, i.e., $i \neq K'$. In case $i = K'$, we have $\Delta \sim 1$. Let us set $\widetilde{\varepsilon} := \varepsilon$ in case $i = K'$, and $\widetilde{\varepsilon} := \min\{\varepsilon, (\alpha_{i+1} - \beta_i)/2\}$ otherwise. Then we find

$$\big\langle \mathrm{Im}\, m(\beta_i + \omega + \mathrm{i}\eta)\big\rangle = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\eta\,\langle v(\tau)\rangle\, \mathrm{d}\tau}{\eta^2 + (\beta_i + \omega - \tau)^2} \sim \eta\int_0^{2\varepsilon} \frac{\min\{\Delta^{-1/6}\tau^{1/2}, \tau^{1/3}\}}{\eta^2 + (\omega + \tau)^2}\,\mathrm{d}\tau, \quad \omega \in [-\varepsilon, \widetilde{\varepsilon}].$$

The contribution to the integral in the middle, coming from the other side α_{i+1} of the gap (β_i, α_{i+1}), is not larger than the last expression, because the growth of the average generating density is the same on both sides of the gap. For the last integral we find

$$\eta\int_0^{2\varepsilon} \frac{\min\{\Delta^{-1/6}\tau^{1/2}, \tau^{1/3}\}}{\eta^2 + (\omega + \tau)^2}\,\mathrm{d}\tau \sim \begin{cases} \dfrac{\eta}{(\Delta + \eta)^{1/6}(\omega + \eta)^{1/2}}, & \omega \in [0, \widetilde{\varepsilon}], \\[2ex] \dfrac{(|\omega| + \eta)^{1/2}}{(\Delta + |\omega| + \eta)^{1/6}}, & \omega \in [-\varepsilon, 0]. \end{cases}$$

This holds for any $\varepsilon \sim 1$ and thus the claim of the lemma follows. □

B.2 Proofs of auxiliary results in Section 7

PROOF OF PROPERTY 3 OF LEMMA 7.1. Let $(\zeta_1, \zeta_2) \in K$ be a location of the maximum of D restricted to K. If (ζ_1, ζ_2) is an extreme point, the claim is shown. Suppose therefore that (ζ_1, ζ_2) is not an extreme point of K. It is easy to see that in this case there exists $(\omega_1, \omega_2) \in \mathbb{C}^2$ such that $(\zeta_1 + \tau\omega_1, \zeta_2 + \tau\omega_2) \in K$ for values of τ in a closed interval $[\alpha, \beta]$, with 0 in its interior, such that $(\zeta_1 + \alpha\omega_1, \zeta_2 + \alpha\omega_2)$ is an extreme point of K.

Since (ζ_1, ζ_2) is a local maximum of D, the function

$$\varphi(\tau) := D(\zeta_1 + \tau\omega_1, \zeta_2 + \tau\omega_2), \tag{B.2}$$

has a local maximum at $\tau = 0$, and thus fulfils

$$\dot{\varphi}(0) = 0 \qquad \text{and} \qquad \ddot{\varphi}(0) \leq 0. \tag{B.3}$$

Here the dot denotes the derivative with respect to the argument τ. We will show that this already implies that φ is constant on $[\alpha, \beta]$. The maximum is therefore also attained at $(\zeta_1 + \alpha\,\omega_1, \zeta_2 + \alpha\,\omega_2)$, which was assumed to be an extreme point of K. This proves Property 3 of the lemma.

We shorten the notation by writing

$$\varrho_j \;:=\; \frac{\operatorname{Im}\omega_j}{\operatorname{Im}\zeta_j}\,, \tag{B.4}$$

for $j = 1, 2$. First let us assume $\omega_1 = \omega_2$. Using the definition of D from (7.1) we get

$$\varphi(\tau) \;=\; \frac{1}{(1+\varrho_1\tau)(1+\varrho_2\tau)}\,\frac{|\zeta_1 - \zeta_2|}{(\operatorname{Im}\zeta_1)(\operatorname{Im}\zeta_2)}\,. \tag{B.5}$$

Now we compute the first and second derivative,

$$\begin{aligned} \dot\varphi(\tau) \;&=\; -\varphi(\tau)\left(\frac{\varrho_1}{1+\varrho_1\tau} + \frac{\varrho_2}{1+\varrho_2\tau}\right)\,, \\ \ddot\varphi(\tau) \;&=\; 2\,\varphi(\tau)\left(\frac{\varrho_1^2}{(1+\varrho_1\tau)^2} + \frac{\varrho_2^2}{(1+\varrho_2\tau)^2} + \frac{\varrho_1\,\varrho_2}{(1+\varrho_1\tau)(1+\varrho_2\tau)}\right)\,. \end{aligned}$$

From the evaluation of these functions at $\tau = 0$ and (B.3) we see that $\varphi(0) = 0$ or $\varrho_1 = \varrho_2 = 0$. In both cases we can easily conclude that $\varphi(\tau)$ is constant.

Therefore, we may now assume $\omega_1 \neq \omega_2$ and write φ in the form

$$\varphi(\tau) \;=\; \frac{|\omega_1 - \omega_2|^2}{(\operatorname{Im}\zeta_1)(\operatorname{Im}\zeta_2)}\,\frac{|\xi|^2 + 2\tau\operatorname{Re}\xi + \tau^2}{(1+\varrho_1\tau)(1+\varrho_2\tau)}\,, \qquad \xi \;:=\; \frac{\zeta_1 - \zeta_2}{\omega_1 - \omega_2}\,. \tag{B.6}$$

We take the first derivative

$$\dot\varphi(\tau) \;=\; \left(\frac{2\operatorname{Re}\xi + 2\tau}{|\xi|^2 + 2\tau\operatorname{Re}\xi + \tau^2} - \frac{\varrho_1}{1+\varrho_1\tau} - \frac{\varrho_2}{1+\varrho_2\tau}\right)\varphi(\tau)\,. \tag{B.7}$$

First we treat the case $\xi = 0$, i.e. $\zeta_1 = \zeta_2$. In this case the second derivative, evaluated at $\tau = 0$, is

$$\ddot\varphi(0) \;=\; 2\,\frac{|\omega_1 - \omega_2|^2}{(\operatorname{Im}\zeta_1)(\operatorname{Im}\zeta_2)} \;>\; 0\,. \tag{B.8}$$

But this contradicts (B.3). Therefore, we have $\xi \neq 0$. In particular, $\varphi(0) > 0$. We compute the second derivative, evaluate it at $\tau = 0$ and use $\dot\varphi(0) = 0$. This way we find

$$\ddot\varphi(0) \;=\; \frac{\varphi(0)}{|\xi|^4}\Big(2|\xi|^2 - (2\operatorname{Re}\xi)^2 + (\varrho_1 + \varrho_2)^2|\xi|^4 - 2\,\varrho_1\,\varrho_2|\xi|^4\Big)\,. \tag{B.9}$$

Now we use the identity,

$$2\operatorname{Re}\xi - (\varrho_1 + \varrho_2)|\xi|^2 \;=\; 0\,, \tag{B.10}$$

coming from $\dot\varphi(0) = 0$ and (B.7). We plug this into (B.9) and get

$$\ddot\varphi(0) \;=\; 2\,\frac{\varphi(0)}{|\xi|^2}\,(1 - \varrho_1\,\varrho_2|\xi|^2)\,. \tag{B.11}$$

Since $\varphi(0) > 0$, we have, in addition to (B.10), the condition

$$1 - \varrho_1\,\varrho_2|\xi|^2 \;\leq\; 0\,. \tag{B.12}$$

We solve (B.10) for ϱ_2 and plug the result into (B.12) to get

$$0 \,\geq\, 1 - \varrho_1 \left(\frac{2\mathrm{Re}\,\xi}{|\xi|^2} - \varrho_1 \right) |\xi|^2 \,=\, |\xi|^2 \left(\varrho_1 - \frac{\mathrm{Re}\,\xi}{|\xi|^2} \right)^2 + \frac{(\mathrm{Im}\,\xi)^2}{|\xi|^2} \,.$$

Thus $\mathrm{Im}\,\xi = 0$, and

$$\varrho_1 \,=\, \varrho_2 \,=\, \frac{\mathrm{Re}\,\xi}{|\xi|^2} \,=\, \frac{1}{\xi} \,.$$

Plugging this back into (B.6) yields that φ is constant. □

Proof of Lemma 7.6. Recall that T is a generic bounded symmetric operator on $\mathrm{L}^2 = \mathrm{L}^2(\mathfrak{X};\mathbb{C})$ that preserves non-negative functions. Moreover, the following is assumed:

$$\exists\, h \in \mathrm{L}^2 \quad \text{s.t.} \quad \|h\|_2 = 1\,, \quad Th \leq h\,, \quad \text{and} \quad \varepsilon := \inf_{x \in \mathfrak{X}} h_x > 0\,. \tag{B.13}$$

We show that $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2} \leq 1$. Let us derive a contradiction by assuming $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2} > 1$. We have

$$T^n h \leq h\,, \qquad \forall\, n \in \mathbb{N}\,. \tag{B.14}$$

Indeed, $Th \leq h$ is true by definition, and if $T^n h \leq h$ for some $n \geq 2$, then also

$$T^{n+1} h \,=\, Th - T(h - T^n h) \,\leq\, h\,,$$

since $T(h - T^n h) \geq 0$ because of $h - T^n h \geq 0$.

Now, the assumption $\|T\|_{\mathrm{L}^2\to\mathrm{L}^2} > 1$ implies:

$$\exists\, u \in \mathscr{B} \quad \text{s.t.} \quad \|u\|_2 = 1\,, \quad u \geq 0\,, \quad \text{and} \quad \langle u, Tu \rangle > 1$$

Since T is positive $\langle u, Tu \rangle \leq \langle |u|, T|u| \rangle$ so we may assume $u \geq 0$. Moreover, by standard density arguments we may assume $\|u\|_{\mathscr{B}} < \infty$ as well.

Since $\langle u, Tu \rangle > 1$, we obtain by inserting u-projections between the T's:

$$\langle u, T^n u \rangle \,\geq\, \langle u, Tu \rangle \langle u, T^{n-1} u \rangle \,\geq\, \cdots \,\geq\, \langle u, Tu \rangle^n \to \infty \qquad \text{as } n \to \infty \tag{B.15}$$

The contradiction follows now by combining (B.14) and (B.15):

$$\langle h, u \rangle \,\geq\, \langle T^n h, u \rangle \,=\, \langle h, T^n u \rangle \,\geq\, \langle h, u \rangle \langle u, T^n u \rangle\,. \tag{B.16}$$

The left hand side is less than $\|h\|_2 \|u\|_2 = 1$. On the other hand, since $h \geq \varepsilon$, $u \geq 0$ and $\|u\|_2 = 1$ we have $\langle h, u \rangle > 0$. Thus (B.15) implies that the right side of (B.16) approaches infinity as n grows. □

B.3 Proofs of auxiliary results in Section 8

Proof of Lemma 8.6. First we note that h is bounded away from zero by

$$h \,=\, Th \,\geq\, \varepsilon \int_{\mathfrak{X}} \pi(\mathrm{d}x)\, h_x\,. \tag{B.17}$$

Let u be orthogonal to h in L^2. Then we compute

$$\begin{aligned}\langle u, (1 \pm T)u \rangle \;&=\; \frac{1}{2}\int \pi(\mathrm{d}x) \int \pi(\mathrm{d}y)\, T_{xy} \left(u_x \sqrt{\frac{h_y}{h_x}} \,\pm\, u_y \sqrt{\frac{h_x}{h_y}}\, \right)^2 \\ &\geq\; \frac{\varepsilon}{2\Phi^2}\int \pi(\mathrm{d}x) \int \pi(\mathrm{d}y)\, h_x\, h_y \left(u_x^2\, \frac{h_y}{h_x} \,+\, u_y^2\, \frac{h_x}{h_y} \,\pm\, 2\, u_x\, u_y \right) \\ &=\; \frac{\varepsilon}{\Phi^2}\int \pi(\mathrm{d}x)\, u_x^2\,,\end{aligned}$$

where in the inequality we used $T_{xy} \geq \varepsilon \geq \varepsilon\, h_x h_y/\Phi^2$ for almost all $x, y \in \mathfrak{X}$. Now we read off that

$$\int_{\mathfrak{X}} \pi(\mathrm{d}x)\, u_x\, (Tu)_x \;\leq\; \Big(1 - \frac{\varepsilon}{\Phi^2}\Big)\, \|u\|_2^2 \qquad \text{and} \qquad \int_{\mathfrak{X}} \pi(\mathrm{d}x)\, u_x\, (Tu)_x \;\geq\; -\Big(1 - \frac{\varepsilon}{\Phi^2}\Big)\, \|u\|_2^2\,. \tag{B.18}$$

This shows the gap in the spectrum of the operator T. □

Proof of Lemma 8.9. Proving the claim (8.43) is equivalent to proving that

$$\|(U-T)w\|_2 \;\geq\; \frac{\|T\|_{\mathrm{L}^2 \to \mathrm{L}^2}}{50}\, \mathrm{Gap}(T)\, |1 - \|T\|_{\mathrm{L}^2\to\mathrm{L}^2} \langle h, Uh \rangle|\, \|w\|_2\,, \qquad \forall\, w \in \mathrm{L}^2\,. \tag{B.19}$$

To this end, fix $w \in \mathrm{L}^2$ with $\|w\|_2 = 1$. We decompose w according to the spectrum of T,

$$w \;=\; \alpha\, h + \beta\, u\,.$$

Here u and h are both normalised and orthogonal to each other. The coefficients, α and β, can be assumed to be non-negative and satisfy $\alpha^2 + \beta^2 = 1$. During this proof we will omit the specification of norms, since every calculation is in L^2. We will show the claim in three separate regimes:

(i) $4\beta \,\geq\, |1 - \|T\|\, \langle h, Uh\rangle|^{1/2}$

(ii) $4\beta \,<\, |1 - \|T\|\, \langle h, Uh\rangle|^{1/2}$ and $|1 - \|T\|\, \langle h, Uh\rangle| \,\geq\, 1 - |\langle h, Uh\rangle|^2$

(iii) $4\beta \,<\, |1 - \|T\|\, \langle h, Uh\rangle|^{1/2}$ and $|1 - \|T\|\, \langle h, Uh\rangle| \,<\, 1 - |\langle h, Uh\rangle|^2$

In the first regime, (i) the triangle inequality yields

$$\|(U-T)w\| \;\geq\; \|w\| - \|Tw\| \;=\; 1 - \big(\alpha^2\, \|T\|^2 + \beta^2 \|Tu\|^2\big)^{1/2}\,.$$

We use the simple inequality, $1 - \sqrt{1-\tau} \geq \tau/2$, valid for every $\tau \in [0,1]$, and find

$$\begin{aligned} 2\,\|(U-T)w\| \;&\geq\; 1 - \alpha^2 \|T\|^2 - \beta^2 \|Tu\|^2 \;=\; 1 - \|T\|^2 + \beta^2 (\|T\|^2 - \|Tu\|^2) \\ &\geq\; 1 - \|T\| + \beta^2 \|T\|\, \mathrm{Gap}(T)\,. \end{aligned}$$

The definition of the first regime implies the desired bound.

In the second regime, (ii), we project the left hand side of (B.19) onto the h-direction,

$$\|(U-T)w\| \;=\; \|(1-U^*T)w\| \;\geq\; |\langle h, (1-U^*T)w\rangle| \;=\; |\alpha\, \langle h, (1-U^*T)h\rangle + \beta\, \langle h, (1-U^*T)u\rangle|\,.$$

Using the orthogonality of h and u, we find

$$\|(U-T)w\| \;\geq\; \alpha\, |1 - \|T\|\, \langle h, Uh\rangle| - \beta\, |\langle Uh, Tu\rangle|\,.$$

We will now use that since $\beta \le 1/2$, the coefficient $\alpha \ge \sqrt{3}/2 \ge 1/2$ is bounded away from zero in this regime. Furthermore, we estimate

$$|\langle Uh, Tu\rangle|^2 \;\le\; \|Tu\|^2 \sup_v |\langle Uh, v\rangle|^2 \;\le\; \|Uh\|^2 - |\langle Uh, h\rangle|^2 \;=\; 1 - |\langle h, Uh\rangle|^2\,,$$

where the supremum is taken over normalised vectors v with $\langle h, v\rangle = 0$. By the definition of the second regime we see that

$$4\beta\,|\langle Uh, Tu\rangle| \;\le\; |1 - \|T\|\,\langle h, Uh\rangle|\,.$$

We conclude that

$$\|(U-T)w\| \;\ge\; \frac{1}{4}\bigl|1 - \|T\|\,\langle h, Uh\rangle\bigr|\,.$$

Finally, we treat the last regime, (iii). Here we project the left hand side of (B.19) onto the orthogonal complement of h and get

$$\|(U-T)w\| \;\ge\; \sup_v |\langle v, (U-T)w\rangle| \;=\; \sup_v \bigl|\alpha\langle v, (U-T)h\rangle + \beta\langle v, (U-T)u\rangle\bigr|\,,$$

where again the v's are normalised and orthogonal to h. We conclude that

$$\|(U-T)w\| \;\ge\; \sup_v \bigl(\alpha\,|\langle v, Uh\rangle| - \beta\,|\langle v, (U-T)u\rangle|\bigr) \;\ge\; \alpha \sup_v |\langle v, Uh\rangle| - \beta \sup_v |\langle v, (U-T)u\rangle|\,.$$

We use $\alpha \ge \sqrt{3}/2$ and $\sup_v |\langle v, Uh\rangle|^2 = 1 - |\langle h, Uh\rangle|^2$, as well as the definition of the third regime to see that

$$\alpha\,\sup_v|\langle v, Uh\rangle| \;\ge \frac{\sqrt{3}}{2}\,\bigl|1 - \|T\|\,\langle h, Uh\rangle\bigr|^{1/2}\,.$$

On the other hand

$$\beta\,\sup_v |\langle v, (U-T)u\rangle| \;\le\; 2\beta \;\le\; \frac{1}{2}\bigl|1 - \|T\|\,\langle h, Uh\rangle\bigr|^{1/2}\,.$$

Therefore, we arrive at

$$\|(U-T)w\| \;\ge\; \frac{\sqrt{3}-1}{2}\bigl|1 - \|T\|\,\langle h, Uh\rangle\bigr|^{1/2} \;\ge\; \frac{1}{8}\bigl|1 - \|T\|\,\langle h, Uh\rangle\bigr|\,.$$

In the last inequality we used that $|1 - \|T\|\,\langle h, Uh\rangle| \;\le\; 2$.

□

B.4 Variational bounds when $\mathrm{Re}\, z = 0$

Proof of Lemma 9.7. Applying Jensen's inequality on the definition (9.16) of J_η yields,

$$J_\eta(w) \;\ge\; \langle w, Sw\rangle \;-\; 2\log\langle w\rangle + 2\eta\,\langle w\rangle\,. \tag{B.20}$$

The lower bound shows that the functional J_η is indeed well defined and takes values in $(-\infty, +\infty]$. Evaluating J_η on a constant function shows that it is not identically $+\infty$.

Next we show that J_η has a unique minimiser on the space L^1_+ (cf. definition (9.15)) of positive integrable functions. As the first step, we show that we can restrict our attention to functions, which satisfy the upper bound $w \le 1/\eta$. To this end, pick $w \in \mathrm{L}^1_+$, such that the set $\{x : w_x \ge \eta^{-1}\}$ has positive π-measure, and define the one parameter family of L^1_+-functions

$$w(\tau) \;:=\; w \;-\; \tau\,(w - \eta^{-1})_+\,, \qquad 0 \le \tau \le 1\,, \tag{B.21}$$

where $\phi_+ := \max\{0, \phi\}$, $\phi \in \mathbb{R}$. It follows that $w(\tau) \le w(0) = w$ and $J_\eta(w(\tau)) < \infty$ for every $\tau \in [0, 1]$. We will show that

$$J_\eta\big(\min(w, \eta^{-1})\big) \;=\; J_\eta(w(1)) \;<\; J_\eta(w)\,. \tag{B.22}$$

For this we compute

$$\frac{\mathrm{d}}{\mathrm{d}\tau} J_\eta(w(\tau)) \;=\; -2\,\Big\langle \Big(Sw(\tau) + \eta - \frac{1}{w(\tau)}\Big)\big(w - \eta^{-1}\big)_+ \Big\rangle\,. \tag{B.23}$$

Since $w \ge 0$ and therefore $Sw \ge 0$, the integrand is positive on the set of x where $w_x > 1/\eta$. Thus, the derivative (B.23) is strictly positive for $\tau \in [0, 1)$. We conclude that the minimiser must be bounded from above by η^{-1}.

Now we use a similar argument to see that we may further restrict the search of the minimiser to functions which satisfy also the lower bound $w \ge \eta/(1+\eta^2)$. To this end, fix $w \in \mathrm{L}^1_+$ satisfying $J_\eta(w) < \infty$ and $\|w\|_\infty \le \eta^{-1}$. Suppose $w < \eta/(1 + \eta^2)$, on some set of positive π-measure, and set

$$w(\tau) \;:=\; w \,+\, \Big(\frac{\eta}{1 + \eta^2} - w\Big)_+ \tau\,, \tag{B.24}$$

so that $w = w(0) \le w(\tau)$, and $J_\eta(w(\tau)) < \infty$, for every $\tau \in [0, 1]$. Differentiation yields,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\, J_\eta(w(\tau)) \;\le\; 2\Big\langle \Big(\frac{1}{\eta} \,+\, \eta \,-\, \frac{1}{w(\tau)}\Big)\Big(\frac{\eta}{1 + \eta^2} - w\Big)_+ \Big\rangle\,, \tag{B.25}$$

where the term η^{-1} originates from $\|Sw(\tau)\|_{\mathscr{B}} \le \|S\|_{\mathscr{B}\to\mathscr{B}}\|w(\tau)\|_{\mathscr{B}} \le \eta^{-1}$. Since $\eta^{-1} + \eta = (\eta/(1 + \eta^2))^{-1}$, and $w < \eta/(1 + \eta^2)$ on a positive set of positive measure, we again conclude that $J_\eta(w(1)) < J_\eta(w)$.

Consider now a sequence $(w^{(n)})_{n\in\mathbb{N}}$ in L^1_+ that satisfies

$$\lim_{n\to\infty} J_\eta(w^{(n)}) \;=\; \inf_w J_\eta(w) \qquad \text{and} \qquad \frac{\eta}{1 + \eta^2} \;\le\; w^{(n)} \;\le\; \frac{1}{\eta}\,. \tag{B.26}$$

Obviously, $w^{(n)}$ also constitutes a bounded sequence of L^2_+. Consequently, there is a subsequence, denoted again by $(w^{(n)})_{n\in\mathbb{N}}$, that converges weakly to an element $w^\star$ of L^2_+. This weak limit also satisfies

$$\frac{\eta}{1 + \eta^2} \;\le\; w^\star_x \;\le\; \frac{1}{\eta}\,, \qquad \forall\, x \in \mathfrak{X}\,. \tag{B.27}$$

In order to conclude that $w^\star$ is indeed a minimiser of J_η we will show that J_η is weakly continuous in L^2_+ at all points $w^\star$ satisfying the bounds (B.27). To this end, we consider the three term constituting J_η separately. Evidently the averaging $u \mapsto \langle u\rangle$ is weakly continuous. For the quadratic form we first compute for any sequence $w^{(n)}$ converging weakly to $w^\star$:

$$\big|\langle w^{(n)}, Sw^{(n)}\rangle - \langle w^\star\, Sw^\star\rangle\big| \;\le\; \big(\|w^{(n)}\|_2 + \|w^\star\|_2\big)\,\|S(w^{(n)} - w^\star)\|_2\,. \tag{B.28}$$

Since the L^2-norm is lower-semicontinuous and $\|w^\star\|_2 \le \|w^\star\|_{\mathscr{B}} \le \eta^{-1}$, we infer

$$\limsup_{n\to\infty} \big|\langle w^{(n)}, Sw^{(n)}\rangle - \langle w^\star\, Sw^\star\rangle\big| \;\le\; \frac{2}{\eta}\, \limsup_{n\to\infty}\, \big\|S(w^{(n)} - w^\star)\big\|_2\,. \tag{B.29}$$

Using the L^2-function, $S_x : \mathfrak{X} \to [0, \infty), y \mapsto S_{xy}$, we obtain:

$$\|S(w^{(n)} - w^\star)\|_2^2 \;=\; \int_{\mathfrak{X}} \pi(\mathrm{d}x)\, \Big|\int_{\mathfrak{X}} \pi(\mathrm{d}y)\, S_{xy}\,(w^{(n)} - w^\star)_y\Big|^2 \;=\; \int_{\mathfrak{X}} \pi(\mathrm{d}x)\, \big|\langle S_x(w^{(n)} - w^\star)\rangle\big|^2\,.$$

Here the weak convergence of $w^{(n)}$ to $w^\star$ implies $h_x^{(n)} := |\langle S_x(w^{(n)} - w^\star)\rangle|^2 \to 0$ for each x separately. The uniform bound $|h_x^{(n)}| \le \|S_x\|_2^2 \|w^{(n)} - w^\star\|_2^2 \le 2(\|w^{(n)}\|_2^2 - \|w^\star\|_2^2)\|S\|^2_{\mathrm{L}^2\to\mathscr{B}}$, and the dominated convergence then yield:

$$\int_{\mathfrak{X}} \pi(\mathrm{d}x)\,\bigl|\langle S_x(w^{(n)} - w^\star)\rangle\bigr|^2 \;=\; \int_{\mathfrak{X}} \pi(\mathrm{d}x)\, h_x^{(n)} \;\to\; 0\,, \qquad \text{as } n \to \infty\,.$$

Hence the last term of (B.28) converges to zero as n goes to infinity, and we have shown that the quadratic form is indeed weakly continuous at $w^\star$.

Finally, we show that also the logarithmic term is weakly continuous at $w^\star$. Applying Jensen's inequality yields

$$\bigl|\langle \log w^{(n)}\rangle - \langle \log w^\star\rangle\bigr| \;=\; \Bigl|\Bigl\langle \log\Bigl(\frac{w^{(n)}}{w^\star}\Bigr)\Bigr\rangle\Bigr| \;\le\; \Bigl|\log \Bigl\langle \frac{w^{(n)}}{w^\star}\Bigr\rangle\Bigr|\,,$$

where the last average converges to 1 by the assumed weak convergence of $w^{(n)}$ to $w^\star$ and since $1/w^\star \in \mathrm{L}^2$ by the lower bound in (B.27).

We have proven the existence of a positive minimiser $w^\star \in \mathrm{L}^1$ that satisfies (B.27). In order to see that $w_x^\star = v_x(\mathrm{i}\eta)$ for a.e. $x \in \mathfrak{X}$ we evaluate a derivative of $J_\eta(w^\star + \tau h)|_{\tau=0}$ for an arbitrary $h \in \mathscr{B}$. This derivative must vanish by the definition of $w^\star$, and therefore

$$(Sw^\star)_x + \eta - \frac{1}{w_x^\star} \;=\; 0\,, \qquad \text{for } \pi\text{-a.e. } x \in \mathfrak{X}\,. \tag{B.30}$$

Since Sw, with $w \in \mathrm{L}^2$, is insensitive to changing the values of w_x, for $x \in I$, whenever $I \subseteq \mathfrak{X}$ is of measure zero, we may modify $w^\star$ on the zero measure set where the equation of (B.30) is not satisfied, so that the equality holds everywhere. Since (B.30) equals QVE at $z = \mathrm{i}\eta$ Theorem 6.1 implies that (B.30) has $v(\mathrm{i}\eta)$ as the unique solution. We conclude that $w_x^\star = v_x(\mathrm{i}\eta)$ for a.e. $x \in \mathfrak{X}$. □

Proof of Lemma 9.8. Since $\mathbf{Z}$ is FID, the exists by the part (ii) of Proposition 9.3 a permutation σ, such that

$$\widetilde{\mathbf{Z}} = (\widetilde{Z}_{ij})_{i,j=1}^K\,, \qquad \widetilde{Z}_{ij} := Z_{i\sigma(j)}\,,$$

has a positive main diagonal, i.e., $\widetilde{Z}_{ii} = 1$ for every i. Let us define the convex function $\Lambda : (0,\infty) \to \mathbb{R}$, by

$$\Lambda(\tau) \;:=\; \frac{\varphi}{K}\,\tau + \log\frac{1}{\tau}\,,$$

where $\varphi > 0$ and $K \in \mathbb{N}$ are from **B2.** Clearly, $\lim_{\tau\to\pm\infty} \Lambda(\tau) = \infty$. In particular,

$$\Lambda(\tau) \;\ge\; \Lambda_-\,, \tag{B.31}$$

where $|\Lambda_-| \lesssim 1$, since φ and K are considered as model parameters,

Using $\widetilde{Z}_{ii} = 1$ and $w_i \widetilde{Z}_{ij} w_{\sigma(j)} \ge 0$ in the definition (9.23) of $\widetilde{J}(\mathbf{w})$, we obtain

$$\sum_i \Lambda(w_i w_{\sigma(i)}) \;\le\; \frac{\varphi}{K}\sum_i \Bigl(w_i \widetilde{Z}_{ii} w_{\sigma(i)} - \log\bigl[\, w_i\, w_{\sigma(i)}\bigr]\Bigr) \;+\; \frac{\varphi}{K}\sum_{i\ne j} w_i \widetilde{Z}_{ij} w_{\sigma(j)} \;=\; \widetilde{J}(\mathbf{w})\,. \tag{B.32}$$

Combining the assumption $\widetilde{J}(\mathbf{w}) \le \Psi$ with the lower bounds (B.31) of Λ yields

$$w_k w_{\sigma(k)} \;\sim\; 1\,, \qquad 1 \le k \le K\,. \tag{B.33}$$

Using (B.31) together with (B.32) and the hypothesis of the lemma, $\widetilde{J}(\mathbf{w}) \leq \Psi$, we obtain an estimate for the off-diagonal terms as well:

$$\frac{\varphi}{K}\sum_{i\neq j} w_i \widetilde{Z}_{ij} w_{\sigma(j)} \;\leq\; \widetilde{J}(\mathbf{w}) - \sum_i \Lambda(w_i w_{\sigma(i)}) \;\leq\; \Psi + K|\Lambda_-| \,. \tag{B.34}$$

Since we consider (φ, K, Ψ) as model parameters, the bounds (B.33) and (B.34) together yield

$$M_{ij} := w_i \widetilde{Z}_{ij} w_{\sigma(j)} \lesssim 1\,. \tag{B.35}$$

This would imply the claim of the lemma, $\max_i w_i \lesssim 1$, provided we would have $\widetilde{Z}_{ij} \gtrsim 1$ for all i, j. To overcome this limitation we compute the $(K-1)$-th power of the matrix $\mathbf{M}$ formed by the components (B.35). This way we get to use the FID property of $\mathbf{Z}$:

$$\begin{aligned}(\mathbf{M}^{K-1})_{ij} \;&=\; \Big(\frac{\varphi}{K}\Big)^{K-1}\sum_{i_1,\dots,i_{K-2}} w_i \widetilde{Z}_{ii_1} w_{\sigma(i_1)} w_{i_1} \widetilde{Z}_{i_1 i_2} w_{\sigma(i_2)} w_{i_2} \widetilde{Z}_{i_2 i_3} w_{\sigma(i_3)} \dots w_{i_{K-2}} \widetilde{Z}_{i_{K-2} j} w_{\sigma(j)} \\ &\geq\; \Big(\frac{\varphi}{K}\Big)^{K-1}\Big(\min_k w_k w_{\sigma(k)}\Big)^{K-2} (\widetilde{\mathbf{Z}}^{K-1})_{ij}\; w_i w_{\sigma(j)}\,.\end{aligned} \tag{B.36}$$

Since $\mathbf{Z}$ is FID also $\widetilde{\mathbf{Z}}$ is FID, and therefore $\min_{i,j}(\widetilde{\mathbf{Z}}^{K-1})_{ij} \geq 1$ (cf. the statements (i) and (iii) of Proposition 9.3). Moreover, by (B.33) we have $(\min_k w_k w_{\sigma(k)})^{K-2} \sim 1$. Thus choosing $j = \sigma^{-1}(i)$, so that $w_i w_{\sigma(j)} = w_i^2$, (B.36) yields

$$w_i^2 \lesssim (\mathbf{M}^{K-1})_{i\sigma^{-1}(i)}\,.$$

This is $\mathcal{O}(1)$ by (B.35). This completes the proof. □

B.5 Cubic roots and associated auxiliary functions

Proof of Lemma 12.7 and Lemma 12.15. Let $p_k : \mathbb{C} \to \mathbb{C}$, $k \in \mathbb{N}$, denote any branch of the inverse of $\zeta \mapsto \zeta^k$ so that $p_k(\zeta)^k = \zeta$. We remark that if p_k is the standard complex power function (cf. Definition 12.5) then the conventional notation $\zeta^{1/k}$ is used instead of $p_k(\zeta)$.

The special functions Φ and $\Phi_\pm$ appearing in Lemma 12.6 and Lemma 12.13, respectively, can be stated in terms of the single function

$$\Phi(\zeta) := p_3(\,p_2(1+\zeta^2) + \zeta\,)\,, \tag{B.37}$$

by rotating ζ and Φ and choosing the functions p_2 and p_3 appropriately. For example, if $|\mathrm{Re}\,\zeta| < 1$, i.e., $\zeta \in \widehat{\mathbb{C}}_0$ (cf. (12.101)), then $\Phi(\pm\mathrm{i}\zeta)^3 = \pm\mathrm{i}\Phi_\mp(\zeta)$, with the standard definition of the complex powers. In order to treat both the lemmas in the unified way, we hence consider the generic function (B.37) that is analytic on a simple connected open set D of $\mathbb{C}$ such that $\pm\mathrm{i} \notin D$.

Straightforward estimates show that

$$|\Phi(\zeta) - \Phi(\xi)| \leq C_1|\zeta - \xi|^{1/2} \tag{B.38}$$

and

$$|\partial_\zeta \Phi(\zeta)| \leq C_3 \begin{cases} |\zeta - \mathrm{i}|^{-1/2} + |\zeta + \mathrm{i}|^{-1/2} & \text{when } |\zeta| \leq 2 \\ |\zeta|^{-2/3} & \text{when } |\zeta| > 2 \end{cases} \tag{B.39}$$

The roots $\widehat{\Omega}_a(\zeta)$ defined in both (12.38) and (12.100) are of the form:

$$\Omega(\zeta) = \alpha_1 \Phi^{(1)}(\omega_1\zeta) + \alpha_2 \Phi^{(2)}(\omega_2\zeta)\,. \tag{B.40}$$

Here $\Phi^{(1)}$ and $\Phi^{(2)}$ satisfy (B.37) but with different choices of branches and branch cuts for the square and the cubic roots. The coefficients $\alpha_1, \alpha_2, \omega_1, \omega_2 \in \mathbb{C}$ satisfy $|\alpha_k| \le 2$ and $|\omega_k| = 1$ for $k = 1, 2$.

The perturbation results of Lemma 12.7 and Lemma 12.15 now follow from (B.39) and the mean value theorem:

$$|\Phi(\zeta+\gamma) - \Phi(\zeta)| \,\le\, |\gamma| \sup_{0\le\rho\le 1} |\partial_\zeta \Phi(\zeta+\rho\gamma)| \,. \tag{B.41}$$

Indeed, Lemma 12.7 follows directly by choosing $D = \big\{\zeta \in \mathbb{C} : \mathrm{dist}(\zeta, \mathbb{G}) \le 1/4\big\}$ with $\mathbb{G}$ defined in (12.44), and $\gamma := \xi$. Since $\zeta \in \mathbb{G} \subset D$ the condition (12.45) for $c_1 = 1/12$ guarantees that $\zeta + \xi \in D$. As $\mathrm{dist}(\pm\mathrm{i}, D) = 1/4$ the estimate (12.46) follows using (B.39) in (B.41).

In order to prove (12.110) we consider the case $\zeta = \mathrm{i}(-\theta+\lambda)$ and $\gamma = \mathrm{i}\mu'\lambda$, where $\theta = \pm 1$, $|\lambda - 2\theta| \ge 6\kappa$, and $|\mu'| \le \kappa$, for some $\kappa \in (0, 1/2)$. We need to bound the distance between the argument $\zeta + \rho\gamma$, of the derivative in (B.41) to the singular points $\pm\mathrm{i}$ from below. Assume $\theta = 1$ w.l.o.g. Then the distance of $\zeta + \rho\gamma$ from $-\mathrm{i}$ is bounded from below by

$$\big|\zeta + \rho\gamma + \mathrm{i}\big| \,\ge\, |\lambda|/2 \,,$$

since $|\rho\mu'| \le \kappa \le 1/2$. Similarly, we bound the distance of $\zeta + \rho\gamma$ from $+\mathrm{i}$ from below

$$\big|\zeta + \rho\gamma - \mathrm{i}\big| \,=\, \big|2\rho\mu' + (1+\rho\mu')(\lambda-2)\big| \,\ge\, \big|(1+\rho\mu')(\lambda-2)\big| - 2\rho|\mu'| \,\ge\, \kappa + |\lambda-2|/2 \,,$$

where for the last estimate we have used the assumption $|\lambda - 2\theta| = |\lambda - 2| \ge 6\kappa$. These bounds apply for arbitrary $0 \le \rho \le 1$. Hence they can be applied to estimate the derivative in (B.41) using (B.39). This way we get

$$\big|\Phi^{(k)}(\zeta+\gamma) - \Phi^{(k)}(\zeta)\big| \,\le\, C_4 \kappa^{-1/2} \min\big\{|\lambda|^{1/2}, |\lambda|^{1/3}\big\}\, |\mu'| \,.$$

Applying this in (B.40) yields (12.110). □

B.6 Hölder continuity of Stieltjes transform

Lemma B.2 (Stieltjes transform inherits regularity). *Let $\gamma > 0$, and assume $\nu : \mathbb{R} \to \mathbb{R}$ is a Hölder-continuous function with Hölder-exponent γ, supported on $[-2, 2]$, i.e.,*

$$|\nu(\tau_2) - \nu(\tau_1)| \,\le\, C_1 \,|\tau_2 - \tau_1|^\gamma \,, \qquad \forall\, \tau_1, \tau_2 \in \mathbb{R} \,, \tag{B.42}$$

for some positive constant C_1. Let ξ be its Stieltjes transform,

$$\xi(z) \,:=\, \int_{-2}^{2} \frac{\nu(\tau)\,\mathrm{d}\tau}{\tau - z} \,.$$

Then there is a constant C_2, depending only on γ, such that

$$|\xi(z_1) - \xi(z_2)| \,\le\, C_2 \bigg(C_1 + \sup_{\tau\in\mathbb{R}} |\nu(\tau)| \bigg) |z_1 - z_2|^\gamma, \qquad \forall\, z_1,\, z_2 \in \mathbb{H} \,,\; |z_1|, |z_2| \le 4 \,. \tag{B.43}$$

Proof of Lemma B.2. We start by writing ξ in the form

$$\xi(\omega + \mathrm{i}\eta) \,=\, \int_{\mathbb{R}} \frac{(\nu(\tau) - \nu(\omega))\,\mathrm{d}\tau}{\tau - \omega - \mathrm{i}\eta} + \mathrm{i}\pi\,\nu(\omega) \,. \tag{B.44}$$

We divide the proof into two steps.

First we assume that the two points for which we compare the values of ξ have the same imaginary part, i.e. we want to show that for every ω_1, $\omega_2 \in [-4,4]$ and $\eta > 0$ we get

$$|\xi(\omega_2 + \mathrm{i}\eta) - \xi(\omega_1 + \mathrm{i}\eta)| \,\le\, C\big(C_1 + \sup_{\tau\in[0,1]} |\nu(\tau)|\big)|E_2 - E_1|^\gamma \,, \tag{B.45}$$

where the constant C only depends on γ. With the formula (B.44) we estimate the difference between $\xi(\omega_2 + \mathrm{i}\eta)$ and $\xi(\omega_1 + \mathrm{i}\eta)$ by splitting the integral into five pieces,

$$\xi(\omega_2 + \mathrm{i}\eta) - \xi(\omega_1 + \mathrm{i}\eta) \,=\, \mathrm{i}\pi\,(\nu(\omega_2) - \nu(\omega_1)) + I_1 - I_2 + I_3 + I_4 + I_5 \,,$$

where we introduced

$$\begin{aligned}
I_1 \,&:=\, \int \mathrm{d}\tau\, \frac{\nu(\tau) - \nu(\omega_2)}{\omega - \omega_2 - \mathrm{i}\eta}\, \mathbb{1}(|\tau - \omega_1| \le |\omega_2 - \omega_1|)\,,\\
I_2 \,&:=\, \int \mathrm{d}\tau\, \frac{\nu(\tau) - \nu(\omega_1)}{\omega - \omega_1 - \mathrm{i}\eta}\, \mathbb{1}(|\tau - \omega_1| \le |\omega_2 - \omega_1|)\,,\\
I_3 \,&:=\, \int \mathrm{d}\tau\, \frac{\nu(\omega_1) - \nu(\omega_2)}{\tau - \omega_1 - \mathrm{i}\eta}\, \mathbb{1}(|\tau - \omega_1| > |\omega_2 - \omega_1|)\,,\\
I_4 \,&:=\, \int \mathrm{d}\tau\, \frac{(\nu(\tau) - \nu(\omega_2))(\omega_2 - \omega_1)}{(\tau - \omega_1 - \mathrm{i}\eta)(\tau - \omega_2 - \mathrm{i}\eta)}\, \mathbb{1}(|\omega_2 - \omega_1| < |\tau - \omega_1| \le 10)\,,\\
I_5 \,&:=\, \int \mathrm{d}\tau\, \frac{(\nu(\tau) - \nu(\omega_2))(\omega_2 - \omega_1)}{(\tau - \omega_1 - \mathrm{i}\eta)(\tau - \omega_2 - \mathrm{i}\eta)}\, \mathbb{1}(|\tau - \omega_1| > 10)\,.
\end{aligned} \tag{B.46}$$

Now we establish, one by one, the following bounds on these five integrals:

$$\begin{aligned}
|I_1| \,&\le\, \frac{4C_1}{\gamma}|\omega_2 - \omega_1|^\gamma\,, \quad |I_2| \,\le \frac{2C_1}{\gamma}|\omega_2 - \omega_1|^\gamma\,, \quad |I_3| \,\le\, \pi C_1 |\omega_2 - \omega_1|^\gamma\,,\\
|I_4| \,&\le\, \frac{2C_1}{\gamma(1-\gamma)}|\omega_2 - \omega_1|^\gamma\,, \quad |I_5| \,\le\, 8 \sup_{\tau\in[0,1]} \nu(\tau)\,|\omega_2 - \omega_1|^\gamma\,.
\end{aligned} \tag{B.47}$$

All these estimates follow simply by pulling the absolute value into the integral. For I_1, I_2 and I_4 we then used that Hölder-continuity of ν. The integral in I_3 can be performed explicitly. Finally, for I_5 we used that $\nu(\tau) = 0$ in the integrand, because ν is supported on $[-2,2]$ only. The contribution of the imaginary part, η, to the absolute value was not used in any of the estimates. Putting everything together yields the result (B.45).

Now we consider the second case when the real parts of the two point, that we evaluate ξ on, are the same. We will show that for any $|\omega| \le 4$ and η, $\lambda > 0$ we have

$$|\xi(\omega + \mathrm{i}\eta) - \xi(\omega + \mathrm{i}\lambda)| \,\le\, C' C_1 |\eta - \lambda|^\gamma \,, \tag{B.48}$$

where the constant C' only depends on γ. This time the representation (B.44) yields

$$\xi(\omega + \mathrm{i}\eta) - \xi(\omega + \mathrm{i}\lambda) \,=\, J_1 - J_2 + J_3 \,, \tag{B.49}$$

where the three integrals J_1, J_2 and J_3 are given as

$$\begin{aligned}
J_1 \,&:=\, \int \mathrm{d}\tau \frac{\nu(\tau) - \nu(\omega)}{\tau - \omega - \mathrm{i}\eta}\, \mathbb{1}(|\tau - \omega| \le |\eta - \lambda|)\,,\\
J_2 \,&:=\, \int \mathrm{d}\tau \frac{\nu(\tau) - \nu(\omega)}{\tau - \omega - \mathrm{i}\lambda}\, \mathbb{1}(|\tau - \omega| \le |\eta - \lambda|)\,,\\
J_3 \,&:=\, \int \mathrm{d}\tau \frac{\mathrm{i}(\eta - \lambda)(\nu(\tau) - \nu(\omega))}{(\tau - \omega - \mathrm{i}\eta)(\tau - \omega - \mathrm{i}\lambda)}\, \mathbb{1}(|\tau - \omega| > |\eta - \lambda|)\,.
\end{aligned} \tag{B.50}$$

Simply using the Hölder-continuity of ν after pulling the absolute value into the integral yields,

$$|J_1| \,\le\, \frac{2C_1}{\gamma}|\eta - \lambda|^\gamma\,, \qquad |J_2| \,\le\, \frac{2C_1}{\gamma}|\eta - \lambda|^\gamma\,, \qquad |J_3| \,\le\, \frac{2C_1}{1-\gamma}|\eta - \lambda|^\gamma\,. \tag{B.51}$$

Now, we combine the two cases (B.45) and (B.48) to finish the proof. □

C Appendix Part III

LEMMA C.1 (Bound propagation). *Suppose C_1, D_1, D_2, D_3 and ε_1 are positive constants, depending explicitly on p, P, L, μ, γ and possible on additional parameters in some set V. Suppose further that the threshold function N_0 from Definition 15.7 depends on the same parameters. Let $\mathbb{D}^{(N)} \subseteq \mathbb{H}$ be a sequence of connected subsets of the complex upper half plane with only polynomially growing diameter, $\sup\{|z_1 - z_2| : z_1, z_2 \in \mathbb{D}^{(N)}\} \leq N^{D_1}$. Let $\varphi = (\varphi^{(N)}(z) : z \in \mathbb{D}^{(N)})_{N \in \mathbb{N}}$ be a sequence of non-negative random functions and $\Phi^{(N)} : \mathbb{D}^{(N)} \to (N^{-D_3}, \infty)$ a sequence of deterministic functions on these sets. Suppose they satisfy the following conditions:*

- *Uniformly for all $z_1, z_2 \in \mathbb{D}^{(N)}$*

$$|\varphi^{(N)}(z_1) - \varphi^{(N)}(z_2)| + |\Phi^{(N)}(z_1) - \Phi^{(N)}(z_2)| \;\leq\; C_1 N^{D_2} |z_1 - z_2|^{\varepsilon_1}\,. \tag{C.1}$$

- *Uniformly for all $z \in \mathbb{D}^{(N)}$*

$$a.w.o.p. \qquad \mathbb{1}\big(\varphi^{(N)} \in [\Phi^{(N)}(z) - N^{-D_3}, \Phi^{(N)}(z)]\big) \;=\; 0\,. \tag{C.2}$$

- *There is a sequence $z_0^{(N)} \in \mathbb{D}^{(N)}$ such that*

$$a.w.o.p. \qquad \varphi^{(N)}(z_0^{(N)}) \;\leq\; \Phi^{(N)}(z_0^{(N)})\,. \tag{C.3}$$

Then the sequence φ satisfies the bound

$$a.w.o.p. \qquad \textit{for all}\;\; z \in \mathbb{D}^{(N)} \;:\quad \varphi^{(N)}(z) \;\leq\; \Phi^{(N)}(z)\,. \tag{C.4}$$

PROOF. We will not carry the upper index N in this proof. First we choose a grid $\mathbb{G} \subseteq \mathbb{D}$ with the following properties

- The number of points in $\mathbb{G}$ is polynomially large, i.e., $|\mathbb{G}| \leq C_2 N^{D_4}$.
- The grid is connected and sufficiently dense in $\mathbb{D}$, i.e., for any two points $z_1, z \in \mathbb{G}$ there is a path $(z_i)_{i=2}^K \subseteq \mathbb{G}$, such that $\max\{|z_K - z|, |z_{i+1} - z_i|\} \;\leq\; N^{-D_5}$ for all $i = 1, \ldots, K-1$.

Here, the positive exponent D_5 is choose sufficiently large such that $C_1 N^{D_2 - \varepsilon_1 D_5} \leq N^{-D_3}/2$. Then an upper bound on the positive constants D_4 and C_2 is determined by the choice of D_5 and the diameter of $\mathbb{D}$, i.e., by D_1.

Now let $z \in \mathbb{G}$. Then we find a path $(z_i)_{i=1}^K$ in $\mathbb{G}$ that connects z_0 with $z_{K+1} := z$ in the sense of the second property of $\mathbb{G}$. We may assume the length of the path, K, to be bounded by $|\mathbb{G}|$. Inductively we show that for all $i = 0, \ldots, K+1$

$$\text{a.w.o.p.} \qquad \varphi(z_i) \;\leq\; \Phi(z_i) - N^{-D_3}.$$

For $i = 0$ this follows from (C.3) and (C.2). For all other i it follows by induction using the continuity condition (C.1), which implies $|\varphi(z_{i+1}) - \varphi(z_i)| + |\Phi(z_{i+1}) - \Phi(z_i)| \leq N^{-D_3}/2$. This shows that if $\varphi(z_i) \leq \Phi(z_i) - N^{-D_3}$, then $\varphi(z_{i+1}) \leq \Phi(z_{i+1})$ and with (C.2) even that $\varphi(z_{i+1}) \leq \Phi(z_{i+1}) - N^{-D_3}$. In particular, $\varphi(z) \leq \Phi(z) - N^{-D_3}$ a.w.o.p..

Using a union bound we infer that

$$\text{a.w.o.p.} \qquad \text{for all } z \in \mathbb{G} \qquad \varphi(z) \;\leq\; \Phi(z) - N^{-D_3}.$$

By (C.1) and since $\mathbb{G}$ is sufficiently dense in $\mathbb{D}$ this bound extends to all $z \in \mathbb{D}$ and the lemma is proven. $\square$

Proof of Lemma 19.1. For f, χ compactly supported on $\mathbb{R}$ the Cauchy integral formula holds true,

$$f(\tau) = \frac{1}{\pi}\int_{\mathbb{R}^2} \frac{\partial_{\bar{z}}\widetilde{f}(\sigma+\mathrm{i}\eta)}{\tau-\sigma-\mathrm{i}\eta}\mathrm{d}\sigma\mathrm{d}\eta = \frac{1}{2\pi}\int_{\mathbb{R}^2}\frac{\mathrm{i}\eta f''(\sigma)\chi(\eta)+\mathrm{i}(f(\sigma)+\mathrm{i}\eta f'(\sigma))\chi'(\eta)}{\tau-\sigma-\mathrm{i}\eta}\mathrm{d}\sigma\mathrm{d}\eta\,,$$
$$\widetilde{f}(\sigma+\mathrm{i}\eta) := (f(\sigma)+\mathrm{i}\eta f'(\sigma))\chi(\eta)\,.$$

For a signed measure ν on $\mathbb{R}$ this implies the formula

$$\int_{\mathbb{R}} f(\tau)\nu(\mathrm{d}\tau) = \mathrm{Re}\int_{\mathbb{R}} f(\tau)\nu(\mathrm{d}\tau) = -\frac{1}{2\pi}\big(I_1(\nu)+I_2(\nu)+I_3(\nu)\big)\,,$$

where the three integrals I_1, I_2 and I_3 are given as

$$I_1(\nu) := \int_{\mathbb{R}^2}\eta f''(\sigma)\chi(\eta)\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta)\,\mathrm{d}\sigma\mathrm{d}\eta\,,$$
$$I_2(\nu) := \int_{\mathbb{R}^2} f(\sigma)\chi'(\eta)\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta)\,\mathrm{d}\sigma\mathrm{d}\eta\,,$$
$$I_3(\nu) := \int_{\mathbb{R}^2}\eta f'(\sigma)\chi'(\eta)\mathrm{Re}\, m_\nu(\sigma+\mathrm{i}\eta)\,\mathrm{d}\sigma\mathrm{d}\eta\,,$$

and m_ν is the Stieltjes transform of ν.

Now we choose $f \ge 0$, such that $f|[\tau_1,\tau_2] = 1$ and $f|\mathbb{R}\setminus[\tau_1-\eta_1,\tau_2+\eta_2] = 0$. Furthermore, we assume that the derivatives of f satisfy

$$\|f'|[\tau_1-\eta_1,\tau_1]\|_\infty \lesssim \eta_1^{-1}\,, \qquad \|f''|[\tau_1-\eta_1,\tau_1]\|_\infty \lesssim \eta_1^{-2}\,,$$
$$\|f'|[\tau_2,\tau_2+\eta_2]\|_\infty \lesssim \eta_2^{-1}\,, \qquad \|f''|[\tau_2,\tau_2+\eta_2]\|_\infty \lesssim \eta_2^{-2}\,.$$

The function $\chi \ge 0$ is chosen to be symmetric and such that $\chi|[-\varepsilon,\varepsilon] = 1$, $\chi|\mathbb{R}\setminus[-2\varepsilon,2\varepsilon] = 0$, as well as $\|\chi'\|_\infty \lesssim \varepsilon^{-1}$. Here the constant ε is chosen to satisfy $\varepsilon \ge \max\{\eta_1,\eta_2\}$. We now derive bounds on $I_k(\nu_1-\nu_2)$ for $k = 1,2,3$.

We split the integral, I_1, into the contributions,

$$I_1(\nu) = 2\,(I_{1,1,<}(\nu)+I_{1,1,>}(\nu)+I_{1,2,<}(\nu)+I_{1,2,>}(\nu))\,,$$
$$I_{1,1,<}(\nu) := \int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma\int_0^{\eta_1}\mathrm{d}\eta\,\eta f''(\sigma)\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta)\,,$$
$$I_{1,1,>}(\nu) := \int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma\int_{\eta_1}^{2\varepsilon}\mathrm{d}\eta\,\eta f''(\sigma)\chi(\eta)\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta)\,,$$
$$I_{1,2,<}(\nu) := \int_{\tau_2}^{\tau_2+\eta_2}\mathrm{d}\sigma\int_0^{\eta_2}\mathrm{d}\eta\,\eta f''(\sigma)\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta)\,,$$
$$I_{1,2,>}(\nu) := \int_{\tau_2}^{\tau_2+\eta_2}\mathrm{d}\sigma\int_{\eta_2}^{2\varepsilon}\mathrm{d}\eta\,\eta f''(\sigma)\chi(\eta)\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta)\,.$$

For a positive measure ν the function $\eta \mapsto \eta\,\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta)$ is monotonously increasing. Thus, we estimate

$$|I_{1,1,<}(\nu)| \le \max_{\sigma\in[0,\eta_1]}|f''(\tau_1-\sigma)|\int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma\int_0^{\eta_1}\mathrm{d}\eta\,\eta_1\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta_1)$$
$$\le \int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma\,\mathrm{Im}\, m_\nu(\sigma+\mathrm{i}\eta_1)\,, \qquad \nu \ge 0\,.$$

We conclude that

$$|I_{1,1,<}(\nu_1-\nu_2)| \;\leq\; \int_{\tau_1-\eta_1}^{\tau_1} \mathrm{d}\sigma\, \big(2\,\mathrm{Im}\, m_{\nu_1}(\sigma+\mathrm{i}\eta_1)+|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta_1)|\big)\,. \tag{C.5}$$

In the same way we find

$$|I_{1,2,<}(\nu_1-\nu_2)| \;\leq\; \int_{\tau_2}^{\tau_2+\eta_2} \mathrm{d}\sigma\, \big(2\,\mathrm{Im}\, m_{\nu_1}(\sigma+\mathrm{i}\eta_2)+|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta_2)|\big)\,. \tag{C.6}$$

For the treatment of $I_{1,1,>}$ we integrate by parts, first in σ and then in η,

$$I_{1,1,>}(\nu) \;=\; -\eta_1\int_{\tau_1-\eta_1}^{\tau_1} \mathrm{d}\sigma\, f'(\sigma)\mathrm{Re}\, m_\nu(\sigma+\mathrm{i}\eta_1) - \int_{\eta_1}^{2\varepsilon}\mathrm{d}\eta \int_{\tau_1-\eta_1}^{\tau_1} \mathrm{d}\sigma\, \partial_\eta(\eta\chi(\eta))\, f'(\sigma)\mathrm{Re}\, m_\nu(\sigma+\mathrm{i}\eta)\,.$$

We use $\max_\eta|\chi(\eta)+\eta\chi'(\eta)|\lesssim 1$ and $\max_{\sigma\in[0,\eta_1]}|f'(\tau_1-\sigma)|\lesssim \eta_1^{-1}$. In this way we estimate for $\nu=\nu_1-\nu_2$,

$$I_{1,1,>}(\nu_1-\nu_2) \;\lesssim\; \int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta_1)| + \frac{1}{\eta_1}\int_{\eta_1}^{2\varepsilon}\mathrm{d}\eta \int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)|\,. \tag{C.7}$$

Going through the same steps we also arrive at

$$I_{1,2,>}(\nu_1-\nu_2) \;\lesssim\; \int_{\tau_2}^{\tau_2+\eta_2}\mathrm{d}\sigma|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta_2)| + \frac{1}{\eta_2}\int_{\eta_2}^{2\varepsilon}\mathrm{d}\eta \int_{\tau_2}^{\tau_2+\eta_2}\mathrm{d}\sigma|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)|\,. \tag{C.8}$$

We continue by estimating I_2 from above.

$$|I_2(\nu_1-\nu_2)| \;\lesssim\; \frac{1}{\varepsilon}\int_{\tau_1-\eta_1}^{\tau_2+\eta_2}\mathrm{d}\sigma \int_{\varepsilon}^{2\varepsilon}\mathrm{d}\eta|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)|\,. \tag{C.9}$$

Finally we derive a bound for I_3. We split the integral into two components,

$$\begin{aligned}
I_3(\nu) \;&=\; 2\big(I_{3,1}(\nu)+I_{3,2}(\nu)\big)\,,\\
I_{3,1}(\nu) \;&:=\; \int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma \int_{\varepsilon}^{2\varepsilon}\mathrm{d}\eta\,\eta f'(\sigma)\chi'(\eta)\mathrm{Re}\, m_\nu(\sigma+\mathrm{i}\eta)\,,\\
I_{3,2}(\nu) \;&:=\; \int_{\tau_2}^{\tau_2+\eta_2}\mathrm{d}\sigma \int_{\varepsilon}^{2\varepsilon}\mathrm{d}\eta\,\eta f'(\sigma)\chi'(\eta)\mathrm{Re}\, m_\nu(\sigma+\mathrm{i}\eta)\,.
\end{aligned}$$

We arrive at the bound

$$I_3(\nu_1-\nu_2) \;\lesssim\; \frac{1}{\eta_1}\int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma \int_{\varepsilon}^{2\varepsilon}\mathrm{d}\eta|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)| + \frac{1}{\eta_2}\int_{\tau_2}^{\tau_2+\eta_2}\mathrm{d}\sigma \int_{\varepsilon}^{2\varepsilon}\mathrm{d}\eta|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)|\,.$$

We combine this with the estimates from (C.5), (C.6), (C.7), (C.8) and (C.9). Altogether we have

$$\Big|\int_{\mathbb{R}} f(\tau)(\nu_1(\mathrm{d}\tau)-\nu_2(\mathrm{d}\tau))\Big| \;\lesssim\; J_1+J_2+J_3\,,$$

where the three terms on the right hand side are given by

$$\begin{aligned}
J_1 \;&:=\; \int_{\tau_1-\eta_1}^{\tau_1}\mathrm{d}\sigma\,\Big(\mathrm{Im}\, m_{\nu_1}(\sigma+\mathrm{i}\eta_1)+|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta_1)|+\frac{1}{\eta_1}\int_{\eta_1}^{2\varepsilon}\mathrm{d}\eta|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)|\Big)\,,\\
J_2 \;&:=\; \int_{\tau_2}^{\tau_2+\eta_2}\mathrm{d}\sigma\,\Big(\mathrm{Im}\, m_{\nu_1}(\sigma+\mathrm{i}\eta_2)+|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta_2)|+\frac{1}{\eta_2}\int_{\eta_2}^{2\varepsilon}\mathrm{d}\eta|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)|\Big)\,,\\
J_3 \;&:=\; \frac{1}{\varepsilon}\int_{\tau_1-\eta_1}^{\tau_2+\eta_2}\mathrm{d}\sigma \int_{\varepsilon}^{2\varepsilon}\mathrm{d}\eta|m_{\nu_1-\nu_2}(\sigma+\mathrm{i}\eta)|\,.
\end{aligned}$$

Now we use this bound for the smoothed out indicator function to derive a bound on the difference of number of eigenvalues in the interval $[\tau_1, \tau_2]$ and the predicted number, given by the integral over the density of states. We use

$$\nu_2([\tau_1,\tau_2]) \le \int f(\tau)\nu_1(\mathrm{d}\tau) + \int f(\tau)(\nu_2(\tau) - \nu_1(\mathrm{d}\tau)), \tag{C.10}$$

for f defined as above. Then we get

$$\nu_2([\tau_1,\tau_2]) \le \nu_1([\tau_1,\tau_2]) + \nu_1([\tau_1-\eta_1,\tau_1]\cup[\tau_2,\tau_2+\eta_2]) + \left|\int f(\tau)(\nu_2(\tau) - \nu_1(\mathrm{d}\tau))\right|.$$

Similarly we use

$$\nu_1([\tau_1,\tau_2]) \ge \int f(\tau)\nu_2(\mathrm{d}\tau) - \nu_1([\tau_1-\eta_1,\tau_1]\cup[\tau_2,\tau_2+\eta_2]),$$

to get the bound

$$\nu_1([\tau_1,\tau_2]) \ge \nu_2([\tau_1,\tau_2]) - \left|\int f(\tau)(\nu_2(\mathrm{d}\tau) - \nu_1(\mathrm{d}\tau))\right| - \nu_1([\tau_1-\eta_1,\tau_1]\cup[\tau_2,\tau_2+\eta_2]).$$

Together, the two bounds imply

$$|\nu_1([\tau_1,\tau_2]) - \nu_2([\tau_1,\tau_2])| \lesssim \nu_1([\tau_1-\eta_1,\tau_1]\cup[\tau_2,\tau_2+\eta_2]) + J_1 + J_2 + J_3.$$

□

Proof of Lemma 21.8. Let us denote the $\nu/3$-neighbourhood of $[0,1]+\mathrm{i}\nu/3$ by

$$\mathbb{K} := \big\{\zeta \in \mathbb{C} : \mathrm{dist}(\zeta,[0,1]+\mathrm{i}(\nu/3)) < \nu/3\big\}. \tag{C.11}$$

We remark that f is analytic on $\mathbb{K}$ since $\mathbb{K} \subset \mathbb{R}_{2\nu/3}$. For the same reason f satisfies

$$|f(\xi) - f(\zeta)| \le C_2\,|\xi - \zeta|, \qquad \xi,\zeta \in \mathbb{K}. \tag{C.12}$$

Since $[0,1] \subset \partial\mathbb{K}$ we have

$$\big|\{\phi \in [0,1] : |f(\phi)| < \delta\}\big| \le \big|\{\zeta \in \partial\mathbb{K} : |f(\zeta)| < \delta\}\big|. \tag{C.13}$$

We will prove (21.45) by estimating the size of the set on the right.

Let us denote the complex unit disk by $\mathbb{D} := \{\zeta \in \mathbb{C} : |\zeta| < 1\}$, and let $\zeta_0 \in \mathbb{K}$ be arbitrary. By the Riemann mapping theorem there exists a bi-holomorphic conformal map $\Phi_{\zeta_0} : \mathbb{C} \to \mathbb{C}$ satisfying

$$\Phi_{\zeta_0}(\mathbb{D}) = \mathbb{K} \qquad \text{and} \qquad \Phi_{\zeta_0}(0) = \zeta_0. \tag{C.14}$$

Since the simple connected sets $\mathbb{D}$ and $\mathbb{K}$ have smooth boundaries the conformal map Φ_{ζ_0} extends to the boundary, such that $\Phi(\partial\mathbb{D}) = \partial\mathbb{K}$, with uniformly bounded derivatives. In particular, we have

$$\frac{1}{C_3(\zeta_0)} \le |\partial\Phi_{\zeta_0}(\zeta)| \le C_3(\zeta_0), \qquad \zeta \in \mathbb{D}, \tag{C.15}$$

with the constant $C_3(\zeta_0) < \infty$ independent of f, in fact it depends only on ζ_0 only through the distance $\mathrm{dist}(\zeta_0, \partial\mathbb{K})$. From the second estimate of (21.44) we know that there are points on $\partial\mathbb{K}$ where $|f| \geq 1$. Hence using the continuity (C.12) we may choose $\zeta_0 \in \mathbb{K}$ such that

$$|f(\zeta_0)| \geq \frac{1}{2} \qquad \text{and} \qquad \mathrm{dist}(\zeta_0, \partial\mathbb{K}) \geq \min\left\{\frac{1}{2C_2}, \frac{\nu}{3}\right\}. \tag{C.16}$$

Here $\nu/3$ is the maximal distance between a point in $\mathbb{K}$ from $\partial\mathbb{K}$.

Let $\log_\pm$ be the positive and negative parts of the logarithm, so that $\log\tau = \log_+\tau - \log_-\tau$ for $\tau > 0$. Using Chebyshev's inequality we get

$$\left|\left\{\zeta \in \partial\mathbb{K} : |f(\zeta)| < \delta\right\}\right| \leq \frac{1}{\log_-\delta}\int_{\partial\mathbb{K}} \log_-|f(\zeta)|\,|\mathrm{d}\zeta|\,.$$

By parametrising the boundary of $\mathbb{K}$ using the conformal map Φ_{ζ_0} we get

$$\left|\left\{\zeta \in \partial\mathbb{K} : |f(\zeta)| < \delta\right\}\right| \leq \frac{1}{\log_-\delta}\int_0^{2\pi} \log_-|f(\Phi_{\zeta_0}(\mathrm{e}^{\mathrm{i}\tau}))|\,|\partial\Phi_{\zeta_0}(\mathrm{e}^{\mathrm{i}\tau})|\,\mathrm{d}\tau\,.$$

Using (C.15) to bound the derivative and writing $\widetilde{f} := f \circ \Phi_{\zeta_0}$ we get

$$\left|\left\{\zeta \in \partial\mathbb{K} : |f(\zeta)| < \delta\right\}\right| \leq \frac{C_3(\zeta_0)}{\log_-\delta}\int_0^{2\pi} \log_-|\widetilde{f}(\mathrm{e}^{\mathrm{i}\tau})|\,\mathrm{d}\tau\,. \tag{C.17}$$

We will now bound the last integral using the Jensen-Poisson formula,

$$\log|\widetilde{f}(0)| = \frac{1}{2\pi}\int_0^{2\pi} \log|\widetilde{f}(\mathrm{e}^{\mathrm{i}\tau})|\,\mathrm{d}\tau - \sum_{j=1}^{n} \log\frac{1}{|\alpha_j|}\,,$$

where α_j's are the zeros of $\widetilde{f}$ in the unit disk $\mathbb{D}$. The last sum is always non-negative since $|\alpha_i| \leq 1$ and can be dropped. By splitting the integral into positive and negative parts we obtain an estimate for the integral on the right hand side of (C.17)

$$\begin{aligned}\int_0^{2\pi} \log_-|\widetilde{f}(\mathrm{e}^{\mathrm{i}\tau})|\,\mathrm{d}\tau &\leq 2\pi\log\frac{1}{|\widetilde{f}(0)|} + \int_0^{2\pi} \log_+|\widetilde{f}(\mathrm{e}^{\mathrm{i}\tau})|\,\mathrm{d}\tau \\ &\leq 2\pi\log 2 + 2\pi\,\log\sup_{\omega\in\mathbb{D}}|\widetilde{f}(\omega)| \\ &\leq 2\pi\log 2C_1\,,\end{aligned}$$

where we have used (C.16) to get the second inequality. For the last bound we have used $|\widetilde{f}(\omega)| = |f(\Phi_{\zeta_0}(\omega))| \leq C_1$. Plugging this into (C.17) and recalling (C.13) we get

$$\left|\left\{\phi \in [0,1] : |f(\phi)| < \delta\right\}\right| \leq \frac{2\pi C_3(\zeta_0)\log 2C_1}{\log(1/\delta)}\,.$$

This finishes the proof as $C_3(\zeta_0)$ and C_1 are independent of δ. □

D German translation

This section contains a translation of the following content of this thesis into German:

- Structure of this work (Struktur der Arbeit)
- Introduction (Einführung)

D.1 Struktur der Arbeit

In dieser Arbeit beweisen wir das Gesetz der lokalen Eigenwertstatistik und verifizieren die Universalitätshypothese für selbstadjungierte Zufallsmatrizen mit unabhängigen Einträgen. Die Arbeit hat drei Teile. In Teil I präsentieren wir eine pädagogische Einleitung in die Problemstellung und skizzieren die Beweisstrategie, indem wir uns auf ein vereinfachtes Modell beschränken. In den Teilen II und III stellen wir die wissenschaftlichen Neuheiten dieser Arbeit vor und geben die vollständigen Beweise an. Von leichten Modifikationen abgesehen stimmen Teil II und III in Inhalt und Formulierung jeweils mit [1] und [2] überein. Einige Paragraphen aus der Einführung in Teil I können ebenfalls [1] und [2] entnommen werden. Die Hauptresultate aus Teil I, Theoreme 2.1, 2.2 und 2.4, sind Vereinfachungen der Theoreme 6.2 und 6.4 in Teil II, sowie der Theoreme 15.6 und 15.14 in Teil III, und somit auch der entsprechenden Resultate aus [1] und [2]. Teil I beinhaltet Beweisskizzen dieser vereinfachten Theoreme, welche der Beweisstrategie der entsprechenden allgemeineren Versionen aus Teil II und III folgen. Die Präsentation in Teil I, welche nicht in [1] und [2] zu finden ist, ist zu empfehlen, um einen Überblick über die relevanten Mechanismen ohne technische Details zu erhalten. In Teil II analysieren wir die *quadratische Vektorgleichung (QVE)*. Diese Gleichung erscheint auf natürliche Weise bei der Anwendung der Methode der Resolventenentwicklung und wird von den Diagonaleinträgen der Resolvente der zugrunde liegenden Zufallsmatrix erfüllt, wenn die Größe der Matrix nach Unendlich strebt. Sehr detaillierte Informationen über die Lösung dieser Gleichung sowie deren Stabilität unter Störungen sind für die in Teil III dieser Arbeit durchgeführten Untersuchungen unentbehrlich. In jenem Teil zeigen wir die Konvegenz der Resolventeneinträge gegen die Lösung der QVE und beweisen die Universalität der lokalen Eigenwertstatistik. Die Resultate in [1] und [2] sind in Zusammenarbeit mit László Erdős und Oskari Ajanki entstanden.

Apart from minor modifications Part II and III coincide both in content and writing with [1] and [2], respectively. Certain paragraphs concerning the background of the problem in Section 1 of Part I can be found in [1] and [2] as well.

D.2 Einführung

In seinem Paper [64] führte Wigner selbstadjungierte Zufallsmatrizen, $\mathbf{H} = \mathbf{H}^*$, mit zentrierten und unabhängig identisch verteilten Einträgen - abgesehen von den durch die Symmetrie vorgegebenen Restriktionen - ein. Er bewies, dass die empirische Dichte der Eigenwerte mit wachsender Größe der Matrix gegen die nach ihm benannte Halbkreisverteilung konvergiert und stellte die Hypothese auf, dass die Verteilung der Abstände zwischen aufeinanderfolgenden Eigenwerten universell sei und daher mit jener übereinstimme, welche durch das Gaußsche Ensemble derselben Symmetrieklasse (GOE/GUE/GSE) vorhergesagt wird.

Die Einträge der Matrizen dieser Ensembles sind - von der Symmetrie abgesehen - unabhängig und folgen der Standard-Gauß-Verteilung. Aufgrund der Invarianz dieser Ensembles unter ihrer großen Symmetriegruppe ist es möglich die gemeinsame Verteilung aller Eigenwerte explizit zu berechnen. Ihre Dichte bezüglich des N-dimensionalen Lebesgue-Maßes ist gegeben durch

$$\rho^{(N)}(\lambda_1, \dots, \lambda_N) \,=\, c_{N,\beta} \prod_{i \neq j} |\lambda_i - \lambda_j|^\beta \mathrm{e}^{-N\frac{\beta}{2}\sum_{i=1}^N \lambda_i^2},$$

wobei β die Werte 1, 2 oder 4 annehmen kann, je nachdem, ob die zugrundeliegende Zufallsmatrix der reel und symmetrischen (GOE), komplex Hermitschen (GUE) oder symplektischen (GSE) Symmetrieklasse angehört, N die Größe der Matrix bezeichnet und $c_{N,\beta}$ eine Normierungskonstante darstellt. Die global Eigenwertdichte (1-Punktfunktion) dieser N -Teilchen-Verteilung ist das Integral der Funktion $\rho^{(N)}$ über die $N-1$ Variablen $\lambda_2, \ldots, \lambda_N$. Für die oben beschriebenen Standard-Gaußschen-Ensembles ist diese Dichte Wigners berühmtes Halbkreisgesetz,

$$\rho_{\mathrm{sc}}(\lambda) := \frac{1}{2\pi}\sqrt{(4-\lambda^2)_+}\,. \tag{D.1}$$

Im Falle dass $\beta = 2$ ist, kann die angemessen normierte k-Punktfunktion, die durch Integration der Dichte $\rho^{(N)}$ über $N-k$ Variablen entsteht, als Determinante geschrieben werden,

$$\rho_k^{(N)}(\lambda_1, \ldots, \lambda_k) = \frac{(N-k)!}{N!}\det(K^{(N)}(\lambda_i, \lambda_j))_{i,j=1}^k\,.$$

Der Kern $K^{(N)}$ ist hierbei explizit durch orthogonale Polynome gegeben. Das asymptotische Verhalten dieser Polynome impliziert, dass im Limes großer N die lokale Eigenwertstatistik einer GUE-Matrix sich einem determinantalen Punktprozess annähert, deren Korrelationen durch den Dyson-Kern,

$$\frac{1}{N\rho(\lambda)}K^{(N)}\Big(\lambda + \frac{x_1}{N\rho(\lambda)}, \lambda + \frac{x_2}{N\rho(\lambda)}\Big) \to \frac{\sin\pi(x_1-x_2)}{\pi(x_1-x_2)}\,, \qquad N \to \infty\,,$$

gegeben sind. Dieser Kern ist universell in dem Sinne, dass er unabhängig von der Position λ im Spektrum ist, sofern die Eigenwertdichte dort nicht verschwindet, also $\rho(\lambda) > 0$ gilt. Ein ähnliches Argument für $\beta = 1$ und $\beta = 4$ zeigt dass auch hier die lokale Eigenwertstatistik durch einen determinantalen Punktprozess beschrieben werden kann. Hier müssen jedoch sowohl die Kernfunktion $K^{(N)}$ als auch ihr Limes durch eine 2×2-matrixwertige Funktion ersetzt werden und eine allgemeinere Klasse spezieller Polynome kommt im Beweis zum Einsatz. Die ersten Resultate dieser Art, welche die Universalität der Gaußsche Ensembles im Inneren des Spektrums mit mathematischer Rigorosität zeigen, wurden von Dyson, Mehta und Gaudin in den 60er Jahren bewiesen.

Wigners revolutionäre Einsicht war es, dass dieses Universalitätsphänomen weit über diese einfachen Modelle hinaus Gültigkeit besitzt und in einer breiten Vielfalt physikalischer Systeme beobachtet werden kann. Wie im Rahmen der Gaußschen Ensembles bestimmt auch dort nur die zugrundeliegende Symmetrieklasse die lokale spektrale Statistik. Obwohl sich diese Hypothese bis jetzt jedem Versuch eines rigorosen Beweises widersetzt hat, wird allgemein davon ausgegangen, dass auch die lokale Statistik der Spektra zufälliger Schrödinger-Operatoren im delokalisieren Regime und der Quantisierungen klassisch chaotischer Systeme durch die Theorie der Zufallsmatritzen (RMT) beschrieben werden kann. Die von RMT vorhergesagten statistischen Eigenschaften können in diversen Systemen beobachtet werden, die von der Verteilung der Nullstellen der Riemannschen-ζ-Funktion [6] und niederenergetischer Vibrationen großer Moleküle [20] über Neutronenresonanzen in schweren Atomkernen [49] bis zu den Eigenwerten des Dirac-operator in der QCD [63] reichen. Obwohl für keines dieser Beispiele das Universalitätsphänomen mathematisch erwiesen ist, gab es im letzten Jahrenzehnt wesentliche Fortschritte im Verständnis jener Mechanismen, die zu diesem Phänomen führen. Diese Fortschritte machen es möglich heutzutage Universalität für eine breite Klasse an Zufallsmatrixmodellen, einschließlich Wigners ursprünglichen Modells, zu beweisen

Wigners Universalitätshypothese folgend, sollte die lokale Eigenwertstatistik von großen selbstadjungierten Zufallsmatrizen mit unabhängig identisch verteilten Einträgen universell sein und somit nicht von den Verteilungen der Einträge abhängen. Diese Hypothese, die auch unter dem Namen Wigner-Dyson-Mehta Hypothese bekannt ist, konnte in den vergangenen Jahren in einer Reihe von Arbeiten gezeigt werden. Das stärkste Resultat für Wigner-Matrizen im Inneren des Spektrums ist Theorem 7.2 in [24], siehe auch [35] und [59] für eine Zusammenfassen der geschichtlichen Abfolge, sowie in engem Zusammenhang stehende Ergebnisse. Tatsächlich lässt sich die Drei-Schritt-Strategie zur Universalität, welche in [33, 36, 26] entwickelt wurde, auch auf *erweiterte Wigner-Matrizen* anwenden, in denen die Einträge unterschiedliche Verteilungen besitzen können, solange die zugehörige Matrix der Varianzen, $s_{ij} := \mathbb{E}|h_{ij}|^2$, stochastisch ist, also solange $\sum_j s_{ij} = 1$ unabhängig von i gilt. Die Stochastizität von $\mathbf{S}$ garantiert, dass die Eigenwertdichte im Limes dem Halbkreisgesetz genügt und die Diagonalelemente G_{ii} der Resolvente $\mathbf{G}(z) = (\mathbf{H} - z)^{-1}$ mit $\operatorname{Im} z > 0$ nicht nur deterministisch, sondern auch von i unabhängig werden. Störungstheorie bis zur zweiten Ordnung legt nahe, dass diese Elemente asymptotisch einem System selbstkonsistenter Gleichungen genügen,

$$-\frac{1}{G_{ii}} \approx z + \sum_{j=1}^{N} s_{ij} G_{jj} \,. \tag{D.2}$$

Falls die Matrix der Varianzen die Stochastizitätsannahme der erweiterten Wigner-Matrizen erfüllt, so vereinfacht sich dieses System zu einer quadratischen Gleichung für eine einzige numerische Größe

$$-\frac{1}{m_{\mathrm{sc}}} = z + m_{\mathrm{sc}} \,. \tag{D.3}$$

Hier ist $m_{\mathrm{sc}} \approx G_{ii}$ der gemeinsame Wert aller Diagonalenträge im Limes $N \to \infty$. Die Lösung der Gleichung (D.3) ist die Stieltjes-Transformation von Wigners Halbkreisverteilung,

$$m_{\mathrm{sc}}(z) = \int_{\mathbb{R}} \frac{\rho_{\mathrm{sc}}(\tau)\mathrm{d}\tau}{\tau - z} \,. \tag{D.4}$$

In dieser Arbeit lassen wir allgemeine Varianzmatrizen zu, die nicht der Stochastizitätsannahme genügen müssen. Die dazugehörigen allgemeinen selbstadjungierten Zafallsmatrizen mit unabhängigen Einträgen nennen wir vom *Wigner-Typ*. Wir zeigen, dass das System selbstkonsistenter Gleichungen (D.2) nach wie vor Gültigkeit besitzt, dass sich dieses jedoch nicht wie oben beschrieben zu einer einzigen Gleichung vereinfachen lässt. In der Tat bleiben für jedes gegebene z in der oberen komplexen Halbebene

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

die Einträge $G_{ii}(z)$ der Resolvente auch im Limes großer N vom Index i abhängig und konvergieren gegen die Lösung $m_i = m_i(z)$ der quadratischen Vektorgleichung (QVE)

$$-\frac{1}{m_i} = z + \sum_{j=1}^{N} s_{ij} m_j \,, \qquad i = 1, \ldots, N \,, \tag{D.5}$$

für N komplexe Zahlen $m_1, \ldots, m_N \in \mathbb{H}$. Die Relevanz dieser Gleichung in dem hier dargestellten Zusammenhang wurde bereits von Girko [41] erkannt, siehe auch die Arbeiten sowohl von Helton, Far und Speicher [43], as auch von Anderson und Zeitouni [5]. Eine tief gehende Analyse der Gleichung jedoch, die ihre Nutzung im Beweis der Universalitätshypothese erst möglich macht, wurde vor der vorliegenden Arbeit nicht durchgeführt.

Hauptaufgabe in Teil II dieser Arbeit wird es sein eine detaillierte Analyse dieses Systems nichtlinearer Gleichungen zu präsentieren. Eine Vielzahl qualitativer und quantitativer Aspekte sind dabei betrachtenswert. Wir werden uns jedoch vor allem mit drei Punkten auseinandersetzen: (i) Regularität der Lösung in der Nähe der reellen Achse mit der Ausnahme einiger weniger singulärer Punkte; (ii) Klassifikation dieser Singularitäten; (iii) Stabilität der

Lösung von (D.5) unter kleinen Störungen. Wir zeigen, dass der Limes von $N^{-1}\operatorname{Tr}\mathbf{G}$, nämlich die Größe $\langle\mathbf{m}\rangle := \frac{1}{N}\sum_i m_i$, die Stieltjes-Transformation einer im inneren Ihres eigenen Trägers analytischen Wahrscheinlichkeitsdichte ρ auf der reellen Achse ist, welche nicht mit der Halbkreisverteilung übereinstimmt. Diese Funktion ρ stellt asymptotisch die Eigenwertdichte der Zufallsmatrix $\mathbf{H}$ dar und wir klassifizieren die Klasse an Funktionen, zu der ρ gehört, indem wir ihr Aussehen in der Nähe des Randes ihres Trägers angeben. Dieses ist in der Form von entweder quadratischen oder auch kubischen (cusp) Singularitäten beschrieben und durch eine einparametrige Familie expliziter Formfunktionen, welche zwischen diesen beiden Verhaltensweisen interpolieren, wenn sich eine Lücke im Träger von ρ schließt.

In Teil III dieser Arbeit beweisen wir die Universalität im Inneren des Spektrums für allgemeine Matrizen vom Wigner-Typ. Dieses Resultat erweitert Wigners Vision des Universalitätsphänomens, indem es dieses für eine erheblich umfangreichere Klasse an Zufallsmatrixensembles nachweist als bislang studiert wurden. Insbesondere zeigen wir, dass die lokale Eigenwertverteilung - wie erwartet - vollständig von der globalen Eigenwertdichte unabhängig ist. Ein solches Resultat war bereits zuvor sowohl im Rahmen allgemeiner β - Ensembles [14] (siehe auch [11] und [54]) als auch additiv deformierter Wigner-Ensembles mit einer globalen Dichte, deren Träger sich auf ein einziges Interval beschränkt, bekannt [48]. Unser Ergebnis gilt für sehr allgemeine Varianzmatrizen und erlaubt auch asymptotische Eigenwertdichten, deren Träger sich über mehrere Intervalle erstreckt (Wir betrachten jedoch keine Zufallsmatrizen mit nichtzentrierten Einträgen, abgesehen von einer kleinen Exkursion in den Bereich von Matrizen mit nichtzentrierten Elementen entlang der Diagonalen in Appendix A.1).

Die Allgemeinheit der zulässigen Dichten ist die wesentliche Neuheit in dem von uns betrachteten Modell. Vorangegangene Methoden (siehe [26] für eine pedagogische Darstellung) nutzen in hohem Maße das explizite Halbkreisgesetz und besonders das Wachstumsverhalten am Rand, das durch eine quadratische Singularität gegeben ist. Der Drei-Schritt-Strategie folgend, zeigen wir zunächst die Konvergenz der Resolvente gegen die Diagonalmatrix $\operatorname{diag}(m_1,\ldots,m_N)$ mit einer optimalen spektralen Auflösung $\eta = \operatorname{Im} z \gg N^{-1}$, also knapp oberhalb der typischen Abstände zwischen den Eigenwerten. Mit der Möglichkeit von kubischen Singularitäten und kleinen Lücken im Träger der Eigenwertdichte ρ - und damit im Spektrum der Zufallsmatrix - erscheint eine zusätzliche charakteristische Länge in der Analyse, die kontrolliert werden muss. Im zweiten Schritt zeigen wir spektrale Universalität für Matrizen vom Wigner-Typ, welche zusätzlich über eine kleine unabhängige Gaußsche Komponente verfügen. Dies geschieht mit Hilfe der *Dyson-Brownschen Bewegung* (DBM). Die zuerst für den Zweck des Beweises der Universalität von Wigner-Matrizen in [32, 33] eingeführte Methode des *lokalen Relaxationflusses* verwendet entscheidend, dass die globale Dichte entlang des Flusses der DBM unverändert dem Halbkreisgesetz folgt. In [34], und unabhängig in [47], wurde daher eine neue Methode entwickelt, in der die DBM lokalisiert wird und damit die lokale Universalität in der Nähe einer festen Energie τ bei nichtverschwindender Teilchendichte gezeigt werden kann, vorausgesetzt die Teilchen haben eine geringe Fluktuation um ihre erwarteten Positionen. Diese Annahme kann unter Benutzung des ersten Schrittes in unserem Drei-Schritt-Programm verifiziert werden. Da Zufallsmatrizen eine der wesentlichen Motivationen für die Entwicklung dieser neuen Methode in [34] waren, sind die Resultate bereits derartig formuliert, dass sie sich direkt auf unsere Situation anwenden lassen. Schließlich wird im dritten Schritt ein störungstheoretisches Argument verwendet, um die Annahme der kleinen Gaußschen Komponente zu entfernen. Dieses benutzt, die *Green-Funktion-Vergleichsmethode*, welche zuerst in [36] Verwendung fand und im Wesentlichen ohne Modifikationen übernommen werden kann.

Am Ende von Teil III wenden wir unsere Resultate an, um Universalität für Gaußsche Zufallsmatrizen mit korrelierten Einträgen zu beweisen. Die meisten mathematischen Arbeiten, die sich mit dem Universalitätsphänomen in Zufallsmatrixensembles auseinandersetzen, betreffen entweder Wigner-Matrizen oder *invariante Ensembles*, in denen die Korrelationsstruktur der Matrixelemente sehr spezifisch ist, da das Wahrscheinlichkeitsmaß auf dem Raum der selbstadjungierten Matrizen die Form

$$\mathrm{P}^{(N)}(\mathrm{d}\mathbf{H}) \,=\, c_N\, \mathrm{e}^{-\operatorname{Tr} V(\mathbf{H})}\mathrm{d}\mathbf{H}\,,$$

hat. Da die allermeisten bis heute etablierten Methoden für die Untersuchung von Wigner-Matrizen entwickelt wurden, nutzen sie zu einem erheblichen Ausmaß die Unabhängigkeit der Matrixeinträge. Nur wenige Resultate, die auch Korrelationen erlauben, sind bekannt, siehe [46, 16, 19, 15] für den Gaußschen Fall. Die globale Eigenwertverteilung wurde in einem bestimmten nichtgaußschen Modell mit (angemessen) schwach korrelierten Einträgen mit Hilfe der Momentenmethode als Wigners Halbkreisverteilung in [53] identifiziert und unter Benutzung der Resolventenmethode in [42]. Ein ähnliches Ergebnis für Kovarianzmatrizen wurde in [51]. gezeigt. Alle diese Arbeiten geben nur die makroskopische Eigenwertdichte an und betrachten Modelle, in denen die Korrelationen genügend schwach sind, so dass diese Dichte mit der des unabhängigen Falls übereinstimmt. Eine allgemeinere Korrelationsstruktur mit einer nichttrivialen Eigenwertdichte wurde in [5] untersucht, doch auch in diesem Fall beschränkt sich das Resultat auf die makroskopische Ebene, siehe auch [50]. Wir erwähnen einen weiteren, sehr neuen Beweis für die die Universalität von Adjazenzmatrizen d-regulärer Graphen [10, 9], welchen eine von den vorher genannten Beispielen völlig unterschiedliche Korrelation zugrunde liegt, da hier in jeder Zeile und Spalte die Zahl der Einsen konstant ist.

In unserer Arbeit nutzen wir die einfache Tatsache, dass die (diskrete) Fourier-Tansformation Gaußscher Zufallsmatrizen mit gewissen translationsinvarianten Korrelationsstrukturen beinahe unabhängige Einträge haben (abgesehen von einer zusätzlichen Symmetrie). Da die Varianzmatrix im Fourier-Raum typischerweise nicht stochastisch ist, sind bisherige Resultate über erweiterte Wigner-Matrizen nicht anwendbar. Die in dieser Arbeit ausgearbeitet Theorie jedoch kann verwendet werden, um Konvergenz der Resolvente und Universalität zu zeigen. Die off-diagonalen Resolventeneiträge sind in diesem Fall nicht vernachlässigbar (anders als im unabhängigen Fall) und diese Einträge erben ihre Abklingeigenschaften von den Korrelationen der Einträge der Zufallsmatrix.

References

[1] O. Ajanki, L. Erdős, and T. Krüger. Quadratic vector equations on the complex upper half plane. *arXiv:1506.05095*, 2015.

[2] O. Ajanki, L. Erdős, and T. Krüger. Universality for general Wigner-type matrices. *arXiv:1506.05098*, 2015.

[3] J. Alt. The Local Semicircle Law for Random Matrices with a Fourfold Symmetry. *arXiv:1506.04683*, 2015.

[4] G. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2010.

[5] G. Anderson and O. Zeitouni. A law of large numbers for finite-range dependent random matrices. *Comm. Pure Appl. Math.*, 61(8):1118–1154, 2008.

[6] B. Aneva. Symmetry of the Riemann operator. *Physics Letters B*, 1999.

[7] Z. Bao and L. Erdős. Delocalization for a class of random block band matrices. *arXiv:1503.07510*, 2015.

[8] R. B. Bapat and T. E. S. Raghavan. *Nonnegative Matrices and Applications*, volume 64 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 1997.

[9] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau. Bulk eigenvalue statistics for random regular graphs. *arXiv:1505.06700*, 2015.

[10] R. Bauerschmidt, A. Knowles, and H.-T. Yau. Local semicircle law for random regular graphs. *arXiv:1503.08702v3*, 2015.

[11] F. Bekerman, A. Figalli, and A. Guionnet. Transport maps for Beta-matrix models and Universality. *Commun. Math. Phys.*, 338:589–619, 2015.

[12] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized wigner matrices. *Elect. J. Probab.*, 19(33):1–53, 2014.

[13] J. Borwein, A. Lewis, and R. Nussbaum. Entropy minimization, dad problems, and doubly stochastic kernels. *J. Funct. Anal.*, 123(2):264 – 307, 1994.

[14] P. Bourgade, L. Erdős, and H.-T. Yau. Universality of general β-ensembles. *Duke Math. J.*, 163(6):1127–1190, 2014.

[15] A. Boutet de Monvel and A. Khorunzhy. On the Norm and Eigenvalue Distribution of Large Random Matrices. *Ann. Probab.*, 27:913–944, 1999.

[16] A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuk. Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Processes and Related Fields*, 2(4):607–636, 1996.

[17] E. Brézin and S. Hikami. Level spacing of random matrices in an external source. *Phys. Rev. E*, 58:7176, 1998.

[18] E. Brézin and S. Hikami. Universal singularity at the closure of a gap in a random matrix theory. *Phys. Rev. E*, 57:4140, 1998.

[19] A. Chakrabarty, R. S. Hazra, and D. Sarkar. From random matrices to long range dependence. *arXiv:1401.0780*, 2014.

[20] S. Cilibert, P. D. L. Rios, and F. Piazza. Glasslike Structure of Globular Proteins and the Boson Peak. *Phys. Rev. Lett.*, 2006.

[21] E. B. Davies. The functional calculus. *J. Lond. Math. Soc. (2)*, 52(1):166–176, 1995.

[22] F. J. Dyson. A Brownian-Motion Model for the Eigenvalues of a Random Matrix. *J. Math. Phys.*, 3:1191, 1962.

[23] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14(8):1837–1926, 2013.

[24] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Commun. Math. Phys.*, 314(3):587–640, 2012.

[25] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Delocalization and Diffusion Profile for Random Band Matrices. *Commun. Math. Phys.*, 323:367–416, 2013.

[26] 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(0):1–58, 2013.

[27] L. Erdős, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau. Bulk universality for Wigner matrices. *Comm. Pure Appl. Math.*, 63(7):895–925, 2010.

[28] L. Erdős, J. A. Ramírez, B. Schlein, and H.-T. Yau. Universality of sine-kernel for Wigner matrices with a small Gaussian perturbation. *Electron. J. Probab.*, 15(18):526–603, 2010.

[29] L. Erdős, B. Schlein, and H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

[30] L. Erdős, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Commun. Math. Phys.*, 287:641–655, 2010.

[31] L. Erdős, B. Schlein, and H.-T. Yau. Wegner estimate and level repulsion for Wigner random matrices. *Int Math Res Notices*, pages 436–479, 2010.

[32] L. Erdős, B. Schlein, and H.-T. Yau. Universality of random matrices and local relaxation flow. *Invent. Math.*, 185:75–119, 2011.

[33] L. Erdős, B. Schlein, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 48(1):1–46, 2012.

[34] L. Erdős and K. Schnelli. Universality for Random Matrix Flows with Time-dependent Density. *arXiv:1504.00650*, 2015.

[35] L. Erdős and H.-T. Yau. Universality of local Spectral statistics of random matrices. *Bull. Amer. Math. Soc*, 49:377–414, 2012.

[36] L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2011.

[37] L. Erdős, H.-T. Yau, and J. Yin. Universality for generalized Wigner matrices with Bernoulli distribution. *J. Comb.*, 2(1):15–82, 2011.

[38] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012.

[39] R. Froese, D. Hasler, and W. Spitzer. Absolutely Continuous Spectrum for the Anderson Model on a Tree: A Geometric Proof of Klein's Theorem. *Commun. Math. Phys.*, 269(1):239–257, 2007.

[40] J. Garnett. *Bounded Analytic Functions*. Springer Science & Business Media, 2007.

[41] V. L. Girko. *Theory of stochastic canonical equations. Vol. I*, volume 535 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 2001.

[42] F. Götze, A. Naumov, and A. Tikhomirov. Semicircle Law for a Class of Random Matrices with Dependent Entries. *arXiv:1211.0389v2*, 2012.

[43] J. W. Helton, R. R. Far, and R. Speicher. Operator-valued Semicircular Elements: Solving A Quadratic Matrix Equation with Positivity Constraints. *International Mathematics Research Notices*, 2007.

[44] T. Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer Science & Business Media, Berlin, Heidelberg, 2 edition, 2012.

[45] M. Keller, D. Lenz, and S. Warzel. An invitation to trees of finite cone type: random and deterministic operators. *arXiv:1403.4426*, 2014.

[46] 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*, pages 97–127. Amer. Math. Soc., Providence, RI, 1994.

[47] B. Landon and H.-T. Yau. Convergence of local statistics of Dyson Brownian motion. *arXiv:1504.03605*, 2015.

[48] J. O. Lee, K. Schnelli, B. Stetler, and H.-T. Yau. Bulk Universality for Deformed Wigner Matrices. *arXiv:1405.6634*, 2014.

[49] M. L. Mehta. *Random Matrices (Second Edition)*. Elsevier, 2014.

[50] F. Merlevede, M. Peligrad, and M. Banna. On the limiting spectral distribution for a large class of random matrices with correlated entries. *arXiv:1312.0037*, 2013.

[51] S. O'Rourke. A note on the marchenko-pastur law for a class of random matrices with dependent entries. *Electron. Commun. Probab.*, 17:no. 28, 1–13, 2012.

[52] C. Sadel. Absolutely Continuous Spectrum for Random Schrödinger Operators on Tree-Strips of Finite Cone Type. *Ann. Henri Poincaré*, 14(4):737–773, 2012.

[53] J. H. Schenker and H. Schulz-Baldes. Semicircle law and freeness for random matrices with symmetries or correlations. *Math. Res. Lett.*, 12(4):531–542, 2005.

[54] M. Shcherbina. On fluctuations of eigenvalues of random band matrices. *arXiv:1504.05762*, 2015.

[55] T. Shcherbina. On the second mixed moment of the characteristic polynomials of 1D band matrices. *Commun. Math. Phys.*, 328(1):45–82, 2014.

[56] T. Shcherbina. Universality of the local regime for the block band matrices with a finite number of blocks. *J. Stat. Phys.*, 155(3):466–499, 2014.

[57] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35:876–879, 1964.

[58] T. Tao and V. Vu. Random matrices: The Four Moment Theorem for Wigner ensembles. *arXiv:1112.1976*, 2011.

[59] T. Tao and V. Vu. Random matrices: the universality phenomenon for Wigner ensembles. In *Modern aspects of random matrix theory*, pages 121–172. Amer. Math. Soc., Providence, RI, 2014.

[60] T. Tao and V. Vu. Random matrices: Universality of local spectral statistics of non-Hermitian matrices. *Ann. Probab.*, 43(2):782–874, 2015.

[61] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*. Amer. Math. Soc., 2012.

[62] C. A. Tracy and H. Widom. The Pearcey Process. *Commun. Math. Phys.*, 263:381–400, 2006.

[63] J. Verbaarschot and T. Wettig. Random Matrix Theory and Chiral Symmetry in QCD. *arXiv:hep-ph/0003017*, 2000.

[64] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.

List of symbols

$\mathbb{P}$	-	underlying probability measure
$\mathbb{E}$	-	expectation with respect to $\mathbb{P}$
N	-	dimension of random matrix
z	-	spectral parameter $= \tau + \mathrm{i}\eta$
$\mathbb{C}$	-	complex plane
$\mathbb{H}$	-	complex upper half plane
$\mathbf{S}$	-	variance matrix with entries s_{ij}
S	-	operator with kernel $s(x,y)$ or S_{xy}
s	-	variance profile function in Part I
$\mathbf{m}$	-	solution of discrete QVE
m	-	solution of QVE
m_{sc}	-	Stieltjes transform of semicircle law
ρ	-	density of states in Parts I and III
ρ_{sc}	-	semicircle law
$\mathbf{H}$	-	random matrix with entries h_{ij}
$\mathbf{H}^{(U)}$	-	$\mathbf{H}$ with indices in U deleted
$\mathbf{G}$	-	resolvent of $\mathbf{H}$ with entries G_{ij}
$\mathbf{G}^{(U)}$	-	resolvent of $\mathbf{H}^{(U)}$
λ_k	-	eigenvalues of $\mathbf{H}$
$\mathbf{u}$	-	eigenvector of $\mathbf{H}$ with entries u_i
$\mathbf{d}$	-	random error of discrete QVE
d	-	perturbation of QVE
F	-	operator with kernel $\lvert m_x\rvert s(s,y)\lvert m_y\rvert$
f	-	Perron-Frobenius eigenvector of F

$(\mathfrak{X}, \mathcal{S}, \pi)$	-	probability space with elements $x, y, \ldots$
$\mathscr{B}$	-	bounded measurable functions on $\mathfrak{X}$
$\mathscr{B}_+$	-	functions in $\mathscr{B}$ with values in $\mathbb{H}$
L^p	-	function space $\mathrm{L}^p(\mathfrak{X}, \pi)$
$\langle \cdot \rangle$	-	average of a vector or a function
$\\|\cdot\\|_{\mathscr{B}}$	-	supremum norm
$\\|\cdot\\|_2$	-	L^2-norm
$\\|\cdot\\|_1$	-	L^1-norm
$\\|\cdot\\|_{\mathscr{B}\to\mathscr{B}}$	-	operator norm from $\mathscr{B}$ to itself
$\\|\cdot\\|_{\mathrm{L}^2\to\mathscr{B}}$	-	operator norm from L^2 to $\mathscr{B}$
v	-	generating measure / density
Ψ_{edge}	-	shape function at an edge
$\Psi_{\min}$	-	shape function at internal local minimum
$\lesssim, \gtrsim, \sim$	-	inequality up to constants
$\vert\vert\vert \cdot \vert\vert\vert_I$	-	norm defined in (6.12)
γ_k	-	classical position in Part I
γ_k	-	location of internal minima of $\langle v \rangle$ in Part II
α_i	-	left edge
β_j	-	right edge
B	-	operator $\vert m\vert^2/m^2 - F$
β	-	eigenvalue of B close to zero
b	-	eigenvector of B with eigenvalue β
K, φ	-	model parameters for block FID in Part II
ϑ, γ	-	model parameters for no outlier row in Part II

$\mathrm{Spec}(\cdot)$	-	L^2-spectrum of an operator
$\mathrm{Gap}(\cdot)$	-	spectral gap of an operator
g	-	solution of perturbed QVE in Part II
Θ	-	scalar from split $g - m = \lvert m\rvert(\Theta b + r)$
μ_1, μ_2, μ_3	-	coefficients of cubic in Part II
μ_k	-	bounds on moments in Part I and Part III
Ω	-	solution of cubic in normal coordinates
$\widehat{\Omega}$	-	roots of reduced cubic
Δ	-	gap size
$\prec$	-	stochastic domination
$\mathbb{M}$	-	local minima of ρ
c, C, c_k, C_k	-	constants depending on model parameters
γ	-	tolerance exponent in Part III
σ	-	scalar quantity $\langle f^3 \operatorname{sign} \mathrm{Re} m\rangle$ in Part II
$\widehat{\sigma}$	-	absolute value of σ in Part II
σ	-	$\widehat{\sigma}$ from Part II in Part III
π_1, π_2	-	rescaled coefficients of the cubic
$\mathbb{T}$	-	discrete torus
$\mathbb{S}$	-	dual discrete torus
L, ρ	-	model parameters from uniform primitivity in Part II
$\mathbf{g}$	-	vector of diagonal resolvent entries
Λ_{d}	-	distance between $\mathbf{g}$ and $\mathbf{m}$
Λ_{o}	-	size of off-diagonal resolvent elements
Λ	-	maximum of Λ_{d} and Λ_{o}
p, P, L	-	model parameters in Part III

$\mathcal{O}(\varphi)$	-	an expression that is bounded by $C\varphi$
$\mathcal{O}_{\mathscr{B}}(\varphi)$	-	a function that is bounded by $C\varphi$ in $\|\cdot\|_{\mathscr{B}}$-norm
$\mathcal{O}_{\mathscr{B}\to\mathscr{B}}(\varphi)$	-	an operator that is bounded by $C\varphi$ in $\|\cdot\|_{\mathscr{B}\to\mathscr{B}}$-norm
$\mathcal{O}_{\mathrm{L}^2\to\mathrm{L}^2}(\varphi)$	-	an operator that is bounded by $C\varphi$ in $\|\cdot\|_{\mathrm{L}^2\to\mathrm{L}^2}$-norm
J_η	-	functional defined in (9.16)

Erklärung über Eigenständigkeit der wissenschaftlichen Arbeit

Der Unterzeichnende (Torben Krüger) erklärt hiermit an Eides statt, dass die in der Dissertation mit dem Titel 'Local spectral universality for random matrices with independent entries' erbrachten wissenschaftlichen Leistungen in eigenständiger Arbeit entstanden sind, dass die Beiträge anderer Personen zu darin enthaltenen Teilergebnissen in der Arbeit selbst wahrheitsgemäß benannt wurden und dass eine Vorbereitung oder Beschaffung von Material zur Anfertigung der Dissertation durch Personen, die sich gewerbsmäßig oder gegen Entgelt hierzu erbieten, nicht stattgefunden hat.

Datum: 29th June 2016 Unterschrift: ______________________