

THE DYSON EQUATION WITH LINEAR SELF-ENERGY:  
SPECTRAL BANDS, EDGES AND CUSPS

JOHANNES ALT, LÁSZLÓ ERDŐS, AND TORBEN KRÜGER

Received: September 13, 2018

Revised: July 16, 2020

Communicated by Friedrich Götze

ABSTRACT. We study the unique solution  $m$  of the Dyson equation

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

on a von Neumann algebra  $\mathcal{A}$  with the constraint  $\operatorname{Im} m \geq 0$ . Here,  $z$  lies in the complex upper half-plane,  $a$  is a self-adjoint element of  $\mathcal{A}$  and  $S$  is a positivity-preserving linear operator on  $\mathcal{A}$ . We show that  $m$  is the Stieltjes transform of a compactly supported  $\mathcal{A}$ -valued measure on  $\mathbb{R}$ . Under suitable assumptions, we establish that this measure has a uniformly  $1/3$ -Hölder continuous density with respect to the Lebesgue measure, which is supported on finitely many intervals, called bands. In fact, the density is analytic inside the bands with a square-root growth at the edges and internal cubic root cusps whenever the gap between two bands vanishes. The shape of these singularities is universal and no other singularity may occur. We give a precise asymptotic description of  $m$  near the singular points. These asymptotics generalize the analysis at the regular edges given in the companion paper on the Tracy-Widom universality for the edge eigenvalue statistics for correlated random matrices [8] and they play a key role in the proof of the Pearcey universality at the cusp for Wigner-type matrices [15, 19]. We also extend the finite dimensional band mass formula from [8] to the von Neumann algebra setting by showing that the spectral mass of the bands is topologically rigid under deformations and we conclude that these masses are quantized in some important cases.

2020 Mathematics Subject Classification: Primary 46T20; Secondary 45G10, 46L10, 60B20

Keywords and Phrases: Dyson equation, positive operator-valued measure, Stieltjes transform, band rigidity

## CONTENTS

1	INTRODUCTION	1423
2	MAIN RESULTS	1427
2.1	Examples	1432
2.2	Main ideas of the proofs	1432
3	THE SOLUTION OF THE DYSON EQUATION	1436
4	REGULARITY OF THE SOLUTION AND THE SELF-CONSISTENT DENSITY OF STATES	1440
4.1	Linear stability of the Dyson equation	1440
4.2	Regularity of the self-consistent density of states – Proof of Proposition 2.3	1444
4.3	Regularity of the solution – Proof of Proposition 2.4	1445
5	SPECTRAL PROPERTIES OF THE STABILITY OPERATOR FOR SMALL SELF-CONSISTENT DENSITY OF STATES	1448
5.1	Spectrum of the stability operator close to zero	1448
5.2	Small eigenvalue of stability operator and corresponding eigenvectors	1451
5.3	Behaviour of $\sigma$ and $\psi$ close to the real axis	1454
5.4	Bound on the derivatives of $m$ when self-consistent density of states is small	1460
6	THE CUBIC EQUATION	1464
6.1	General cubic equation	1466
6.2	Cubic equation associated to Dyson stability equation	1469
6.3	The cubic equation for the shape analysis	1471
7	CUBIC ANALYSIS	1472
7.1	Shape regular points	1477
7.2	Cubic equations in normal form	1482
7.3	Proof of Theorem 7.7	1488
7.4	Proof of Theorem 7.1 and Proposition 7.5	1491
7.5	Proofs of Theorem 7.2 and Proposition 7.6	1497
7.6	Characterisations of a regular edge	1499
8	BAND MASS FORMULA – PROOF OF PROPOSITION 2.6	1501
9	DYSON EQUATION FOR KRONECKER RANDOM MATRICES	1504
9.1	The Kronecker setup	1505
9.2	$N \times N$ -Kronecker random matrices	1507
9.3	Limits of Kronecker random matrices	1508

10 PERTURBATIONS OF THE DATA PAIR 1510

A STIELTJES TRANSFORMS OF POSITIVE OPERATOR-VALUED MEASURES 1521

B POSITIVITY-PRESERVING, SYMMETRIC OPERATORS ON  $\mathcal{A}$  1523

C NON-HERMITIAN PERTURBATION THEORY 1528

D CHARACTERIZATION OF  $\text{supp } \rho$  1531

1 INTRODUCTION

An important task in random matrix theory is to determine the eigenvalue distribution of a random matrix as its size tends to infinity. Similarly, in free probability theory, the scalar-valued distribution of operator-valued semicircular elements is of particular interest. In both cases, the distribution can be obtained from the corresponding *Dyson equation*

$$-m(z)^{-1} = z\mathbb{1} - a + S[m(z)] \tag{1.1}$$

on some von Neumann algebra  $\mathcal{A}$  with a unit  $\mathbb{1}$  and a tracial state  $\langle \cdot \rangle$ . Here,  $z$  lies in  $\mathbb{H} := \{w \in \mathbb{C} : \text{Im } w > 0\}$ , the complex upper half-plane,  $a = a^* \in \mathcal{A}$  and  $S : \mathcal{A} \rightarrow \mathcal{A}$  is a positivity-preserving linear operator. There is a unique solution  $m : \mathbb{H} \rightarrow \mathcal{A}$  of (1.1) under the assumption that  $\text{Im } m(z) := (m(z) - m(z)^*)/(2i)$  is a strictly positive element of  $\mathcal{A}$  for all  $z \in \mathbb{H}$  [30]. For suitably chosen  $a$  and  $S$  as well as  $\mathcal{A}$ , this solution characterizes the distributions in the applications mentioned above. In fact, in both cases, the distribution will be the measure  $\rho$  on  $\mathbb{R}$  whose Stieltjes transform is given by  $z \mapsto \langle m(z) \rangle$ . The measure  $\rho$  is called the *self-consistent density of states* and its support is the *self-consistent spectrum*. This terminology stems from the physics literature on the Dyson equation, where  $z$  is often called *spectral parameter* and  $S[m]$  the self-energy. The linearity of the *self-energy operator*  $S$  is a distinctive feature of our setup. We first explain the connection between the eigenvalue density of a large random matrix and the Dyson equation. Let  $H \in \mathbb{C}^{n \times n}$  be a  $\mathbb{C}^{n \times n}$ -valued random variable,  $n \in \mathbb{N}$ , such that  $H = H^*$ . A central objective is the analysis of the *empirical spectral measure*  $\mu_H := n^{-1} \sum_{i=1}^n \delta_{\lambda_i}$ , or its expectation, the *density of states*, for large  $n$ , where  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of  $H$ . Clearly,  $n^{-1} \text{Tr}(H - z)^{-1}$  is the Stieltjes transform of  $\mu_H$  at  $z \in \mathbb{H}$ . Therefore, the resolvent  $(H - z)^{-1}$  is commonly studied to obtain information about  $\mu_H$ . In fact, for many random matrix ensembles, in particular models with decaying correlations among the entries, the resolvent  $(H - z)^{-1}$  is well-approximated for large  $n$  by the solution  $m(z)$  of the Dyson equation (1.1). Here, we choose  $\mathcal{A} = \mathbb{C}^{n \times n}$  equipped with the operator norm induced by the Euclidean distance on  $\mathbb{C}^n$  and the normalized trace  $\langle \cdot \rangle = n^{-1} \text{Tr}(\cdot)$  as tracial state as well as

$$a := \mathbb{E}H, \quad S[x] := \mathbb{E}[(H - a)x(H - a)], \quad x \in \mathbb{C}^{n \times n}. \tag{1.2}$$

If  $(H - z)^{-1}$  is well-approximated by  $m(z)$  for large  $n$  then  $\mu_H$  will be well-approximated by the deterministic measure  $\rho$ , whose Stieltjes transform is given by  $z \mapsto \langle m(z) \rangle$ . The importance of the Dyson equation (1.1) for random matrix theory has been realized by many authors on various levels of generality [10, 13, 25, 31, 39, 48], see also the monographs [24, 37] and the more recent works [2, 4, 6, 7, 9, 20, 28, 32].

Secondly, we relate the Dyson equation to free probability theory by noticing that the Cauchy transform of a shifted operator-valued semicircular element is given by  $m$ . More precisely, let  $\mathcal{B}$  be a unital  $C^*$ -algebra,  $\mathcal{A} \subset \mathcal{B}$  be a  $C^*$ -subalgebra with the same unit  $\mathbf{1}$  and  $E: \mathcal{B} \rightarrow \mathcal{A}$  is a conditional expectation (we refer to Chapter 9 in [36] for notions from free probability theory). Pick an  $a = a^* \in \mathcal{A}$  and an operator-valued semicircular element  $s = s^* \in \mathcal{B}$ . Then  $G(z) := E[(z - s - a)^{-1}]$  is the *Cauchy-transform* of  $s + a$ . In this case,  $m(z) = -G(z)$  satisfies (1.1) with  $S[x] := E[sxs]$  for all  $x \in \mathcal{A}$  [44]. If  $\mathcal{A}$  is a von Neumann algebra with a tracial state, then our results yield information about the scalar-valued distribution  $\rho = \rho_{s+a}$  of  $s + a$  with respect to this state. The study of qualitative regularity properties for this distribution has a long history in free probability. For example, the question of whether  $\rho$  has atoms or not is intimately related to non-commutative identity testing (see [22, 34] and references therein) and the notions of free entropy and Fischer information (see [43, 45] and the survey [47]). We also refer to the recent preprint [35], where the distribution of rational functions in noncommutative random variables is studied with the help of linearization ideas from [27, 26] and [29]. Under certain assumptions, our results provide extremely detailed information about the regularity properties of  $\rho$ , thus complementing these more general insights. In particular, we show that  $\rho_s$  is absolutely continuous with respect to the Lebesgue measure away from zero for any operator-valued semicircular element  $s$ . For other applications of the Dyson equation (1.1) in free probability theory, we refer to [30, 41, 44, 46] and the recent monograph [36].

In this paper, we analyze the regularity properties of the self-consistent density of states  $\rho$  in detail. More precisely, under suitable assumptions on  $S$ , we show that the boundedness of  $m$  already implies that  $\rho$  has a  $1/3$ -Hölder continuous density  $\rho(\tau)$  with respect to the Lebesgue measure. We provide a broad class of models for which the boundedness of  $m$  is ensured. Furthermore, the set where the density is positive,  $\{\tau : \rho(\tau) > 0\}$ , splits into finitely many connected components, called *bands*. The density is real-analytic inside the bands with a square root growth behavior at the edges. If two bands touch, however, a cubic root cusp emerges. These are the only possible types of singularities. In fact,  $m(z)$  is the Stieltjes transform of a positive operator-valued measure  $\nu$  and we establish the properties mentioned above for  $\nu$  as well. We also extend the *band mass formula* from [8] expressing the masses that  $\rho$  assigns to the bands. We use it to infer a certain quantization of the band masses that we call *band rigidity*, because it is invariant under small perturbations of the data  $a$  and  $S$  of the Dyson equation. In particular, we extend a quantization result from [26] and [40] to cover limits of Kronecker random matrices. We remark that for

the analogous phenomenon in the context of random matrices the term “exact separation of eigenvalues” was coined in [11].

In the commutative setup, the band structure and singularity behavior of the density have been obtained in [3, 1], where a detailed analysis of the regularity of  $\rho$  was initiated. In the special noncommutative situation  $\mathcal{A} = \mathbb{C}^{n \times n}$  and  $\langle \cdot \rangle = n^{-1} \text{Tr}(\cdot)$ , it has been shown that  $\rho$  is Hölder-continuous and real-analytic wherever it is positive [4]. Recently, in the same setup, the precise behavior of  $\rho$  near the spectral edges was obtained in [8], where it was a key ingredient in the proof of the Tracy-Widom universality of the local spectral statistics near the spectral edges for random matrices with general correlation structure. However, this analysis works only at edges that are well separated from each other (so called *regular edges*), i.e. away from the *cusps* where two edges merge and away from the *almost cusps*, i.e. regions with small spectral gaps or small but nonzero minima of the density. The main novelty of the current work is to give an effective regularity analysis for the general noncommutative case with a precise quantitative description of all singularities including the almost cusps. One of the main applications is the proof of the eigenvalue rigidity on optimal scale throughout the entire spectrum. This is a key input for the recent proof of the local spectral universality at the cusp for general Wigner-type matrices, i.e. the Pearcey statistics for the complex hermitian case in [19] and its real symmetric counterpart in [15]. We remark that cusp universality settles the third and last ubiquitous spectral universality regime after the bulk and edge universalities studied extensively earlier, see [21] and references therein.

The key strategy behind the current paper as well as its predecessors [3, 1, 4, 8] is a refined stability analysis of the Dyson equation (1.1) against small perturbations. It turns out that the equation is stable in the bulk regime, i.e., where  $\rho(\text{Re } z)$  is separated away from zero, but is unstable near the points, where the density vanishes. Even the stability in the bulk requires an unconventional idea; it relies on rewriting the stability operator, i.e., the derivative of the Dyson equation with respect to  $m$ , through the use of a positivity-preserving symmetric map, called the *saturated self-energy operator*,  $F$ . We then extract information on the spectral gap of  $F$  by a Perron-Frobenius argument using the positivity of  $\text{Im } m$  [3, 1]. In the non-commutative setup this transformation was based on a novel *balanced polar decomposition* formula [4]. In the small density regime, in particular near the regular edges studied in [8], the stability deteriorates due to an *unstable direction*, which is related to the Perron-Frobenius eigenvector of  $F$ . The analysis boils down to a scalar quantity,  $\Theta$ , the overlap between the solution and the unstable direction. For the commutative case in [3, 1], it is shown that  $\Theta$  approximately satisfies a cubic equation. The structural property of this cubic equation is its *stability*, i.e., that the coefficients of the cubic and quadratic terms do not simultaneously vanish. This guarantees that higher order terms are negligible and the order of any singularity is either cubic root or square root.

Now we synthesize both analyses in the previous works to study the small

density regime in the most general setup. The major obstacle is the noncommutativity that already substantially complicated the bulk analysis [4], but there the saturated self-energy operator,  $F$ , governed all estimates. However, in the regime of small density the unstable direction is identified via the top eigenvector of a non-symmetric operator that coincides with the symmetric  $F$  only in the commutative case. Thus we need to perform a non-symmetric perturbation expansion that requires precise control on the resolvent of the non-selfadjoint stability operator in the entire complex plane. We still work with a cubic equation for  $\Theta$ , but the analysis of its coefficients is considerably more involved than in [3, 1].

The situation is much simpler near the regular edges, where the cubic equation simplifies to a quadratic equation; this analysis was performed in [8] at least in the finite dimensional non-commutative case. The main novelty of the present paper lies in handling the most complicated case, the cusps and almost cusps, where we need to deal with a genuine cubic equation. The second goal of the paper is to give a unified treatment of all spectral regimes in the general von Neumann algebraic setup. A few arguments pertaining the regular edges are relatively simple extensions from [8] to the infinite dimensional case. We will indicate these instances but for the reader's convenience we chose to include these proofs since in the current paper we work under weaker conditions and in a more general setup than in [8].

We stress that along all estimates, the noncommutativity is a permanent enemy; in some cases it can be treated perturbatively, but for the most critical parts new non-perturbative proofs are needed. Most critically, the stability of the cubic equation is proven with a new method.

Another novelty of the current paper, in addition to handling the noncommutativity and lack of symmetry, is that we present the cubic analysis in a conceptually clean way that will be used in future works. Our analysis strongly suggests that our cubic equation for  $\Theta$  is the key to any detailed singularity analysis of Dyson-type equations and its remarkable structure is responsible for the universal behavior of the singularities in the density.

As a final remark we compare our self-consistent density of states  $\rho$ , obtained from the Dyson equation, with the equilibrium density  $\rho_V$  considered in invariant matrix ensembles defined via a Gibbs factor  $\exp(-\frac{1}{n} \text{Tr } V(H))$ , where  $V$  is an external potential. Recall that  $\rho_V$  is the solution of a variational principle [18]. Both densities approximate the empirical density of states of a prominent class of random matrix ensembles, but they have quite different singularity structures at the vanishing points. Our classification theorem shows that  $\rho$  has only square root and cusp singularities. On the other hand, if  $V \in C^2$  then  $\rho_V$  is  $1/2$ -Hölder continuous, in particular it cannot have any cusp singularity. Moreover  $\rho_V$  may vanish at the edges of its support not necessarily as a square root, see e.g. a behaviour  $\rho_V(x) \approx (x_+)^{5/2}$  in Example 1.2 of [16]. In general, only powers  $\alpha = 2k$  and  $\alpha = 2k + \frac{1}{2}$ ,  $k \in \mathbb{N}$  are possible for the vanishing behavior  $\rho_V(x) \approx (x_+)^{\alpha}$ . These patterns persist under small additive perturbations with an independent GUE matrix, moreover, at critical coupling, a cusp singu-

larity similar to our case appears as well [17]. A summary of known behaviours of  $\rho_V$  near its vanishing points in relation with  $V$  is found in Section 1.3 of [12]. The complexity of these patterns indicates that a concise classification theorem of singularities, similar to our result on  $\rho$  with merely two types of singularities, does not hold for  $\rho_V$ .

ACKNOWLEDGEMENT The first and second author were partially funded by ERC Advanced Grant RANMAT No. 338804. The third author was partially supported by the Hausdorff Center for Mathematics.

## 2 MAIN RESULTS

Let  $\mathcal{A}$  be a finite von Neumann algebra over the complex numbers with unit  $\mathbf{1}$  and norm  $\|\cdot\|$ . We recall that a von Neumann algebra  $\mathcal{A}$  is called *finite* if there is a state  $\langle \cdot \rangle: \mathcal{A} \rightarrow \mathbb{C}$  which is (i) *tracial*, i.e.,  $\langle xy \rangle = \langle yx \rangle$  for all  $x, y \in \mathcal{A}$ , (ii) *faithful*, i.e.,  $\langle x^*x \rangle = 0$  for some  $x \in \mathcal{A}$  implies  $x = 0$ , and (iii) *normal*, i.e., continuous with respect to the weak\* topology. In the following,  $\langle \cdot \rangle$  will always denote such state. The tracial state defines a scalar product  $\mathcal{A} \times \mathcal{A} \rightarrow \mathbb{C}$  through

$$\langle x, y \rangle := \langle x^*y \rangle \tag{2.1}$$

for  $x, y \in \mathcal{A}$ . The induced norm is denoted by  $\|x\|_2 := \langle x, x \rangle^{1/2}$  for  $x \in \mathcal{A}$ . Clearly,  $|\langle x \rangle| \leq \|x\|_2 \leq \|x\|$  for all  $x \in \mathcal{A}$ . We note that if  $\mathcal{A}$  is represented as algebra of operators on a Hilbert space then  $\|\cdot\|$  coincides with the operator norm induced by the scalar product on such Hilbert space. We follow the convention that small letters are elements of  $\mathcal{A}$  while capital letters denote linear operators on  $\mathcal{A}$ . The spectrum of  $x \in \mathcal{A}$  is denoted by  $\text{Spec } x$ , i.e.,  $\text{Spec } x = \mathbb{C} \setminus \{z \in \mathbb{C}: (x - z)^{-1} \in \mathcal{A}\}$ .

For an operator  $T: \mathcal{A} \rightarrow \mathcal{A}$ , we will work with three norms. We denote these norms by  $\|T\|$ ,  $\|T\|_2$  and  $\|T\|_{2 \rightarrow \|\cdot\|}$  if  $T$  is considered as an operator  $(\mathcal{A}, \|\cdot\|) \rightarrow (\mathcal{A}, \|\cdot\|)$ ,  $(\mathcal{A}, \|\cdot\|_2) \rightarrow (\mathcal{A}, \|\cdot\|_2)$  or  $(\mathcal{A}, \|\cdot\|_2) \rightarrow (\mathcal{A}, \|\cdot\|)$ , respectively.

We denote by  $\mathcal{A}_{\text{sa}}$  the self-adjoint elements of  $\mathcal{A}$ , by  $\mathcal{A}_+$  the cone of positive definite elements of  $\mathcal{A}$ , i.e.,

$$\mathcal{A}_{\text{sa}} := \{x \in \mathcal{A}: x^* = x\}, \quad \mathcal{A}_+ := \{x \in \mathcal{A}_{\text{sa}}: x > 0\},$$

and by  $\overline{\mathcal{A}}_+$ , the  $\|\cdot\|$ -closure of  $\mathcal{A}_+$ , the cone of positive semidefinite elements (or positive elements). We now introduce two classes of linear operators on  $\mathcal{A}$  that preserve the cone  $\overline{\mathcal{A}}_+$ . Such operators are called *positivity-preserving* (or *positive maps*). We define

$$\Sigma := \{S: \mathcal{A} \rightarrow \mathcal{A}: S \text{ is linear, symmetric wrt. (2.1)} \\ \text{and preserves the cone } \overline{\mathcal{A}}_+\}, \tag{2.2a}$$

$$\Sigma_{\text{flat}} := \left\{ S \in \Sigma: \varepsilon \mathbf{1} \leq \inf_{x \in \mathcal{A}_+} \frac{S[x]}{\langle x \rangle} \leq \sup_{x \in \mathcal{A}_+} \frac{S[x]}{\langle x \rangle} \leq \varepsilon^{-1} \mathbf{1} \text{ for some } \varepsilon > 0 \right\}. \tag{2.2b}$$

Moreover, if  $S: \mathcal{A} \rightarrow \mathcal{A}$  is a positivity-preserving operator, then  $S$  is bounded, i.e.,  $\|S\|$  is finite (see e.g. Proposition 2.1 in [38]).

Let  $a \in \mathcal{A}_{\text{sa}}$  be a self-adjoint element and  $S \in \Sigma$ . For the *data pair*  $(a, S)$ , we consider the associated *Dyson equation*

$$-m(z)^{-1} = z\mathbb{1} - a + S[m(z)], \quad (2.3)$$

with spectral parameter  $z \in \mathbb{H} := \{w \in \mathbb{C} : \text{Im } w > 0\}$ , for a function  $m: \mathbb{H} \rightarrow \mathcal{A}$  such that its imaginary part is positive definite,

$$\text{Im } m(z) = \frac{1}{2i}(m(z) - m(z)^*) \in \mathcal{A}_+.$$

There always exists a unique solution  $m$  to the Dyson equation (2.3) satisfying  $\text{Im } m(z) \in \mathcal{A}_+$  [30]. Moreover, this solution is holomorphic in  $z$  [30]. For Dyson equations in the context of renormalization theory,  $a$  is called the *bare matrix* and  $S$  the *self-energy (operator)*. In applications to free probability theory,  $S$  is usually denoted by  $\eta$  and called the *covariance mapping* or *covariance matrix* [36].

We now introduce positive operator-valued measures with values in  $\overline{\mathcal{A}}_+$ . If  $v$  maps Borel sets on  $\mathbb{R}$  to elements of  $\overline{\mathcal{A}}_+$  such that  $\langle x, v(\cdot)x \rangle$  is a positive measure for all  $x \in \mathcal{A}$  then we say that  $v$  is a *measure on  $\mathbb{R}$  with values in  $\overline{\mathcal{A}}_+$*  or an  *$\overline{\mathcal{A}}_+$ -valued measure on  $\mathbb{R}$* .

First, we list a few propositions that are necessary to state our main theorem. They will be proven in Section 3, Section 4.2 and Section 4.3, respectively.

**PROPOSITION 2.1** (Stieltjes transform representation). *Let  $(a, S) \in \mathcal{A}_{\text{sa}} \times \Sigma$  be a data pair and  $m$  the solution to the associated Dyson equation, (2.3). Then there exists a measure  $v$  on  $\mathbb{R}$  with values in  $\overline{\mathcal{A}}_+$  such that  $v(\mathbb{R}) = \mathbb{1}$  and*

$$m(z) = \int_{\mathbb{R}} \frac{v(d\tau)}{\tau - z} \quad (2.4)$$

for all  $z \in \mathbb{H}$ . The support of  $v$  and the spectrum of  $a$  satisfy the following inclusions

$$\text{supp } v \subset \text{Spec } a + [-2\|S\|^{1/2}, 2\|S\|^{1/2}], \quad (2.5a)$$

$$\text{Spec } a \subset \text{supp } v + [-\|S\|^{1/2}, \|S\|^{1/2}], \quad (2.5b)$$

where for the sum of two subsets  $I$  and  $J$  of  $\mathbb{R}$  we used the notation  $I + J := \{\tau_1 + \tau_2 : \tau_1 \in I, \tau_2 \in J\}$ .

Furthermore, for any  $z \in \mathbb{H}$ ,  $m(z)$  satisfies the bound

$$\|m(z)\|_2 \leq \frac{2}{\text{dist}(z, \text{Conv Spec } a)}, \quad (2.6)$$

where  $\text{Conv Spec } a$  denotes the convex hull of  $\text{Spec } a$ .



Our goal is to obtain regularity results for the measure  $v$ . We first present some regularity results on the self-consistent density of states introduced in the following definition.

DEFINITION 2.2 (Density of states). Let  $(a, S) \in \mathcal{A}_{\text{sa}} \times \Sigma$  be a data pair,  $m$  the solution to the associated Dyson equation, (2.3), and  $v$  the  $\overline{\mathcal{A}}_+$ -valued measure of Proposition 2.1. The positive measure  $\rho = \langle v \rangle$  on  $\mathbb{R}$  is called the *self-consistent density of states* or short *density of states*.

We have  $\text{supp } \rho = \text{supp } v$  due to the faithfulness of  $\langle \cdot \rangle$ . Moreover, the Stieltjes transform of  $\rho$  is given by  $\langle m \rangle$  since, by (2.3), for any  $z \in \mathbb{H}$ , we have

$$\langle m(z) \rangle = \int_{\mathbb{R}} \frac{\rho(d\tau)}{\tau - z}.$$

PROPOSITION 2.3 (Regularity of density of states). *Let  $(a, S)$  be a data pair with  $S \in \Sigma_{\text{flat}}$  and  $\rho_{a,S}$  the corresponding density of states from Definition 2.2. Then  $\rho_{a,S}$  has a uniformly Hölder-continuous, compactly supported density with respect to the Lebesgue measure,*

$$\rho_{a,S}(d\tau) = \rho_{a,S}(\tau)d\tau.$$

Furthermore, there exists a universal constant<sup>1</sup>  $c > 0$  such that the function  $\rho: \mathcal{A}_{\text{sa}} \times \Sigma_{\text{flat}} \times \mathbb{R} \rightarrow [0, \infty)$ ,  $(a, S, \tau) \mapsto \rho_{a,S}(\tau)$  is locally Hölder-continuous with Hölder exponent  $c$  and real-analytic whenever it is positive, i.e., for any  $(a, S, \tau) \in \mathcal{A}_{\text{sa}} \times \Sigma_{\text{flat}} \times \mathbb{R}$  such that  $\rho_{a,S}(\tau) > 0$  the function  $\rho$  is real-analytic in a neighbourhood of  $(a, S, \tau)$  in  $\mathcal{A}_{\text{sa}} \times \Sigma_{\text{flat}} \times \mathbb{R}$ . Here,  $\mathcal{A}_{\text{sa}}$  and  $\Sigma_{\text{flat}}$  are equipped with the metrics induced by  $\| \cdot \|$  on  $\mathcal{A}$  and its operator norm on  $\mathcal{A} \rightarrow \mathcal{A}$ , respectively.

In addition to the flatness condition  $S \in \Sigma_{\text{flat}}$ , the other essential input of our analysis is a boundedness assumption on  $m$ , the solution to (2.3) (see (2.7) below). These two assumptions imply several key facts about  $m$ . The most important ones are (i) 1/3-Hölder continuity of  $m$  (Proposition 2.4), (ii)  $\text{Im } m$  is comparable (as self-adjoint elements of  $\mathcal{A}$ ) with its average  $\langle \text{Im } m \rangle \mathbb{1}$  (Proposition 3.5) and, finally, (iii) the specific quadratic or cubic singularity structure of  $\text{Im } m$  (Theorem 2.5).

While the flatness condition  $S \in \Sigma_{\text{flat}}$  can be directly checked from the data pair, the boundedness of  $m$  is more subtle. However, in the random matrix context, in Section 9, we provide, for a large class of models, a sufficient condition for (2.7) below that is purely expressed in terms of  $a$  and  $S$ .

We start with the Hölder-continuity of  $m$  stated in the following Proposition 2.4. We remark that in the finite dimensional case, where  $\mathcal{A} = \mathbb{C}^{N \times N}$  and  $\langle \cdot \rangle = \frac{1}{N} \text{Tr}(\cdot)$ , this proposition has already been established in [8, Corollary 4.5] and the arguments there remain valid in our more general setup. Nevertheless, we will present its proof to keep the current work self-contained.

<sup>1</sup>By carefully following the proof, we see that choosing  $c = 1/200$  is sufficient.

PROPOSITION 2.4 (Regularity of  $m$ ). *Let  $(a, S)$  be a data pair with  $S \in \Sigma_{\text{flat}}$  and  $m$  the solution to the associated Dyson equation, (2.3). Suppose that for a nonempty open interval  $I \subset \mathbb{R}$  we have*

$$\limsup_{\eta \downarrow 0} \sup_{\tau \in I} \|m(\tau + i\eta)\| < \infty. \quad (2.7)$$

*Then  $m$  has a  $1/3$ -Hölder continuous extension (also denoted by  $m$ ) to any closed interval  $I' \subset I$ , i.e.,*

$$\sup_{z_1, z_2 \in I' \times i[0, \infty)} \frac{\|m(z_1) - m(z_2)\|}{|z_1 - z_2|^{1/3}} < \infty \quad (2.8)$$

*and  $\sup\{\|m(z)\| : z \in I' \times i[0, \infty)\} < \infty$ . Moreover,  $m$  is real-analytic in  $I$  wherever  $\rho$  is positive.*

The purpose of the interval  $I$  in Proposition 2.4 (see also Theorem 2.5 below) is to demonstrate the local nature of these statements and their proofs; if  $m$  is bounded on  $I$  in the sense of (2.7) then we will prove regularity of  $m$  and later its behaviour close to singularities on a genuine subinterval  $I' \subset I$ . At first reading, the reader may ignore this subtlety and assume  $I' = I = \mathbb{R}$ .

In Proposition 4.7 below, we provide a quantitative version of (2.8) under slightly weaker conditions than those of Proposition 2.4.

In the following main theorem, we require that the solution  $m$  to (2.3), originally defined for all  $z \in \mathbb{H}$ , has a continuous extension to some interval  $I \subset \mathbb{R}$ . This is apparently a stronger condition than the boundedness condition (2.7), but they are in fact equivalent as Proposition 2.4 shows. Furthermore, if  $m$  has a continuous extension to  $I$  then the restriction of the measure  $v$  from (2.4) to  $I$  has a density with respect to the Lebesgue measure, i.e.,

$$v(A) = \frac{1}{\pi} \int_A \text{Im } m(\tau) d\tau \quad (2.9)$$

for each Borel set  $A \subset I$ .

THEOREM 2.5 (Im  $m$  close to its singularities). *Let  $(a, S)$  be a data pair with  $S \in \Sigma_{\text{flat}}$  and  $m$  the solution to the associated Dyson equation, (2.3). Suppose  $m$  has a continuous extension to a nonempty open interval  $I \subset \mathbb{R}$ , i.e.,  $m$  is continuous on  $I \times i[0, \infty)$ . Then any  $\tau_0 \in \text{supp } \rho \cap I$  with  $\rho(\tau_0) = 0$  belongs to exactly one of the following cases:*

*Edge: The point  $\tau_0$  is a left/right edge of the support of the density of states  $\rho$ , i.e., there is some  $\varepsilon > 0$  such that  $\text{Im } m(\tau_0 \mp \omega) = 0$  for  $\omega \in [0, \varepsilon]$  and for some  $v_0 \in \mathcal{A}_+$  we have*

$$\text{Im } m(\tau_0 \pm \omega) = v_0 \omega^{1/2} + \mathcal{O}(\omega), \quad \omega \downarrow 0.$$

*Cusp: The point  $\tau_0$  lies in the interior of  $\text{supp } \rho$  and for some  $v_0 \in \mathcal{A}_+$  we have*

$$\text{Im } m(\tau_0 + \omega) = v_0 |\omega|^{1/3} + \mathcal{O}(|\omega|^{2/3}), \quad \omega \rightarrow 0.$$

Moreover,  $\text{supp } \rho \cap I = \text{supp } v \cap I$  is a finite union of closed intervals with nonempty interior.

Theorem 2.5 is a simplified version of our more detailed and quantitative Theorem 7.1 below. In fact, we can treat all small local minima of  $\rho$  on  $\text{supp } \rho \cap I$  – not only those ones, where  $\rho$  vanishes – and provide precise expansions corresponding to those in Theorem 2.5 which are valid in some neighbourhood of  $\tau_0$ . In these expansions in Theorem 7.1, we also track the exact dependence on  $\rho(\tau_0)$ . In particular, we completely control the transition between a small gap in  $\text{supp } \rho$  and a small local minimum of  $\rho$  through the cusp regime that arises when a data pair with a cusp is subject to a small variation. By applying  $\langle \cdot \rangle$  to the results of Theorem 2.5 and Theorem 7.1, we also obtain an expansion of the self-consistent density of states  $\rho$  near small local minima in Theorem 7.2 below.

Finally, we present our quantization result. This result has appeared in [8, Proposition 5.1] for the simpler setting  $\mathcal{A} = \mathbb{C}^{N \times N}$  and under the flatness condition  $S \in \Sigma_{\text{flat}}$ . In the current work we will follow the same strategy of proof when  $\mathcal{A}$  is a general von Neumann algebra with certain adjustments to treat the possibly infinite dimension and the lack of flatness.

PROPOSITION 2.6 (Band mass formula). *Let  $(a, S) \in \mathcal{A}_{\text{sa}} \times \Sigma$  be a data pair and  $m$  the solution to the associated Dyson equation, (2.3). We assume that there is a constant  $C > 0$  such that  $S[x] \leq C \langle x \rangle \mathbb{1}$  for all  $x \in \overline{\mathcal{A}}_+$ . Then we have*

- (i) *For each  $\tau \in \mathbb{R} \setminus \text{supp } \rho$ , there is  $m(\tau) \in \mathcal{A}_{\text{sa}}$  such that  $\lim_{\eta \downarrow 0} \|m(\tau + i\eta) - m(\tau)\| = 0$ . Moreover,  $m(\tau)$  determines the mass of  $(-\infty, \tau)$  and  $(\tau, \infty)$  with respect to  $\rho$  in the sense that*

$$\rho((-\infty, \tau)) = \langle \mathbf{1}_{(-\infty, 0)}(m(\tau)) \rangle, \tag{2.10}$$

where  $\mathbf{1}_{(-\infty, 0)}$  denotes the characteristic function of the interval  $(-\infty, 0)$  and  $\mathbf{1}_{(-\infty, 0)}(m(\tau))$  is defined via the spectral calculus of the self-adjoint element  $m(\tau)$ .

- (ii) *If  $\pi: \mathcal{A} \rightarrow \mathbb{C}^{n \times n}$  is a faithful representation such that  $\langle x \rangle = n^{-1} \text{Tr}(\pi(x))$  for all  $x \in \mathcal{A}$  and  $J \subset \text{supp } \rho$  is a connected component of  $\text{supp } \rho$  then we have*

$$n\rho(J) \in \{1, \dots, n\}.$$

*In particular,  $\text{supp } \rho$  has at most  $n$  connected components.*

We will prove Proposition 2.6 in Section 8 below. A result similar to part (ii) has been obtained by a different method in [26], see also [40]. In fact, we will use the band mass formula, (2.10), in Corollary 9.4 below to strengthen the quantization result in (ii) for a large class of random matrices (Kronecker matrices, see Section 9). In Section 10, we study the stability of the Dyson equation, (2.3), under small general perturbations of the data pair  $(a, S)$ .

## 2.1 EXAMPLES

We now present some examples that show the different types of singularities described by Theorem 2.5. These examples are obtained by considering the Dyson equation, (2.3), on  $\mathbb{C}^{n \times n}$  with  $\langle \cdot \rangle = n^{-1} \text{Tr}$  for large  $n$  and choosing  $a = 0$  as well as  $S = S_\alpha$ , where

$$S_\alpha[x] := \frac{1}{n} D(r_\alpha \text{diag}(x))$$

for any  $x \in \mathbb{C}^{n \times n}$ . Here, for  $x \in \mathbb{C}^{n \times n}$ ,  $\text{diag}(x) \in \mathbb{C}^n$  denotes the column vector of diagonal entries. We write  $D(v)$  with  $v \in \mathbb{C}^n$  for the diagonal matrix in  $\mathbb{C}^{n \times n}$  with  $v$  on its diagonal. Moreover,  $r_\alpha \in \mathbb{C}^{n \times n}$  is the symmetric block matrix from Figure 1 with  $\alpha \in (0, \infty)$ . All entries in each block are the indicated constants. Therefore,  $S_\alpha[x]$  is a diagonal matrix with the vector

$$r_\alpha = \begin{array}{|c|c|} \hline \alpha & 1 \\ \hline 1 & \alpha \\ \hline \end{array}$$

Figure 1: Structure of  $r_\alpha \in \mathbb{C}^{n \times n}$ .

$r_\alpha \text{diag}(x)/n$  on its diagonal. In fact, this example can also be realized on  $\mathbb{C}^2$  with entrywise multiplication. Here, we choose  $\langle (x_1, x_2) \rangle = \delta x_1 + (1 - \delta)x_2$ , where  $\delta$  is the relative block size of the small block in the definition of  $r_\alpha$ . Indeed, the solution  $m \in \mathbb{C}^{n \times n}$  of the Dyson equation on  $\mathbb{C}^{n \times n}$  has the structure  $m = D(m_1, \dots, m_1, m_2, \dots, m_2)$ , where  $m_1$  appears  $n\delta$  times and  $m_2$  appears  $(1 - \delta)n$  times. Thus, in the setup on  $\mathbb{C}^2$ , the Dyson equation for  $(m_1, m_2) \in \mathbb{C}^2$  can be written as

$$-\begin{pmatrix} m_1^{-1} \\ m_2^{-1} \end{pmatrix} = z \begin{pmatrix} 1 \\ 1 \end{pmatrix} + R_\alpha \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \quad R_\alpha = \begin{pmatrix} \alpha\delta & 1 - \delta \\ \delta & \alpha(1 - \delta) \end{pmatrix} \quad (2.11)$$

for  $(m_1, m_2) \in \mathbb{C}^2$ . We remark that  $R_\alpha$  is symmetric with respect to the scalar product (2.1) induced by  $\langle \cdot \rangle$ . Figure 2 contains the graphs of some self-consistent densities of states  $\rho$  obtained from (2.11) for  $\delta = 0.1$  and different values of  $\alpha$ . As the self-consistent density of states is symmetric around zero in these cases, only the part of the density on  $[0, \infty)$  is shown. The density in Figure 2 (a) has a small internal gap with square root edges on both sides of this gap. Figure 2 (b) contains a cusp which is transformed, by increasing  $\alpha$ , into an internal nonzero local minimum in Figure 2 (c). This nonzero local minimum is covered by Theorem 7.1 (d) below.

## 2.2 MAIN IDEAS OF THE PROOFS

In this subsection, we informally summarize several key ideas in the proofs of Proposition 2.4 and Theorem 2.5.

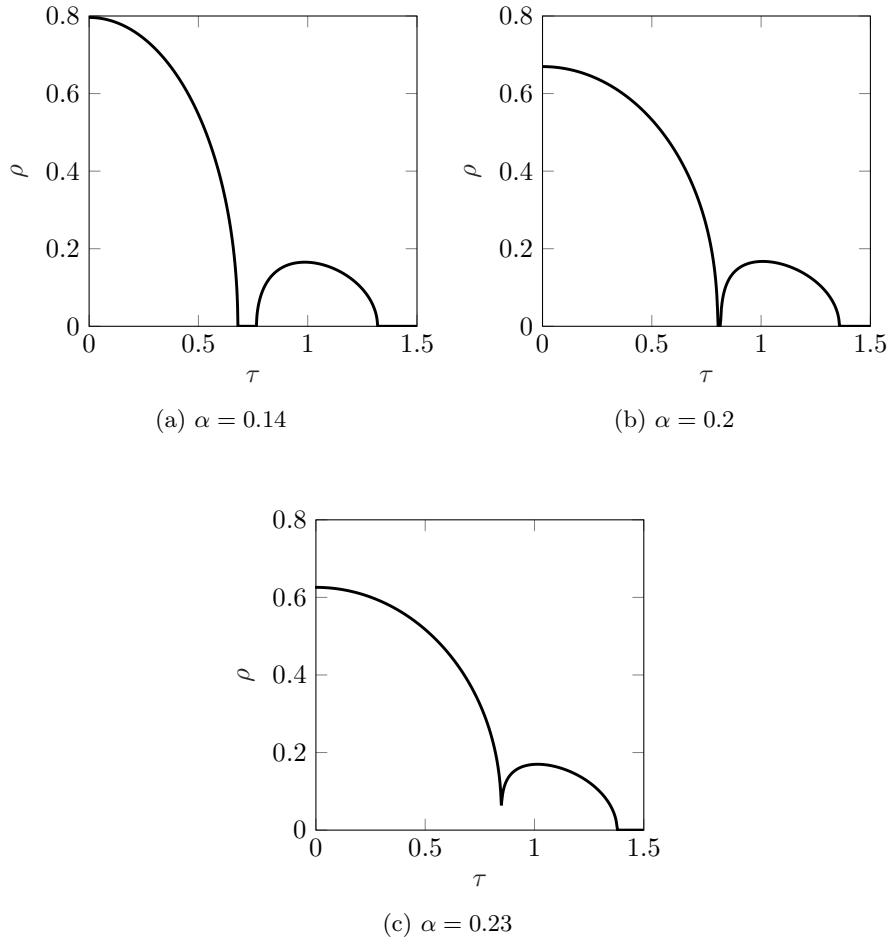


Figure 2: Examples of the self-consistent density of states  $\rho$  from (2.11) for  $\delta = 0.1$  and several values of  $\alpha$ .

HÖLDER-CONTINUITY OF  $m$ . To simplify the notation, we assume in this outline that<sup>2</sup>  $\|m(z)\| \lesssim 1$  for all  $z \in \mathbb{H}$ . We first show that  $\text{Im } m(z)$  is  $1/3$ -Hölder continuous and then conclude the same regularity for  $m = m(z)$ . To that end, we now control  $\partial_z \text{Im } m(z)$  by differentiating the Dyson equation, (2.3), with respect to  $z$ . Since  $2i\partial_z \text{Im } m = \partial_z m$  by the holomorphicity of  $m$  (as

---

<sup>2</sup>In this subsection, we write  $x \lesssim y$  if the two nonnegative scalars  $x, y \geq 0$  satisfy  $x \leq Cy$  for some constant  $C > 0$ . For a more general definition specifying the dependence of  $C$  on other constants, we refer to Convention 3.4 below.

stated below (2.3)), this yields

$$2i\partial_z \operatorname{Im} m = (\operatorname{Id} - C_m S)^{-1}[m^2]. \tag{2.12}$$

Here,  $\operatorname{Id}$  denotes the identity map on  $\mathcal{A}$  and  $C_m : \mathcal{A} \rightarrow \mathcal{A}$  is defined by  $C_m[x] := mxm$  for any  $x \in \mathcal{A}$ .

We first analyse the derivative of  $\operatorname{Im} m$  and prove that  $\operatorname{Im} m$  is  $1/3$ -Hölder continuous since, as we show next, the right-hand side of (2.12) is bounded in terms of  $\operatorname{Im} m$ , namely by  $\|\operatorname{Im} m\|^{-2}$ . In order to control the norm of the inverse  $(\operatorname{Id} - C_m S)^{-1}$  of the stability operator, we rewrite it in a more symmetric form. We find an invertible  $V$  with  $\|V\|, \|V^{-1}\| \lesssim 1$ , a unitary operator  $U$  and a self-adjoint operator  $T$  acting on  $\mathcal{A}$  such that

$$\operatorname{Id} - C_m S = V^{-1}(U - T)V.$$

The Rotation-Inversion Lemma from [1] (see Lemma 4.4 below) is designed to control  $(U - T)^{-1}$  for a unitary operator  $U$  and a self-adjoint operator  $T$  with  $\|T\|_2 \leq 1$ . Applying this lemma in our setup yields  $\|(\operatorname{Id} - C_m S)^{-1}\| \lesssim \|\operatorname{Im} m\|^{-2}$ .

Since  $\|m\| \lesssim 1$ , we thus obtain

$$\|\partial_z \operatorname{Im} m\| \lesssim \|\operatorname{Im} m\|^{-2}. \tag{2.13}$$

This bound implies that  $(\operatorname{Im} m)^3 : \mathbb{H} \rightarrow \mathcal{A}_+$  is uniformly Lipschitz-continuous. Hence, we can extend  $\operatorname{Im} m$  to a  $1/3$ -Hölder continuous function on  $\mathbb{R} \cup \mathbb{H}$  and we obtain

$$m(z) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\operatorname{Im} m(\tau) d\tau}{\tau - z}.$$

This also implies that  $m$  is uniformly  $1/3$ -Hölder continuous on  $\mathbb{R} \cup \mathbb{H}$ . Furthermore,  $m(\tau)$  and  $\operatorname{Im} m(\tau)$  are real-analytic in  $\tau$  around  $\tau_0 \in \mathbb{R}$ , wherever  $\rho(\tau_0)$  is positive.

**BEHAVIOUR OF  $\operatorname{Im} m$  WHERE IT IS NOT ANALYTIC.** Owing to (2.13), some unstable behaviour of the Dyson equation is expected close to points  $\tau_0 \in \mathbb{R}$ , where  $\operatorname{Im} m(\tau_0)$  is zero or small. In order to analyze this behaviour of  $\operatorname{Im} m(\tau)$ , we compute  $\Delta := m(\tau_0 + \omega) - m(\tau_0)$  from the Dyson equation, (2.3), with  $\omega \in \mathbb{R}$ . Since  $m$  has a continuous extension to  $\mathbb{R}$ , (2.3) holds true for  $z \in \mathbb{R}$  as well. We evaluate (2.3) at  $z = \tau_0$  and  $z = \tau_0 + \omega$  and obtain the quadratic  $\mathcal{A}$ -valued equation

$$B[\Delta] = mS[\Delta]\Delta + \omega m\Delta + \omega m^2, \quad B := \operatorname{Id} - C_m S, \quad m := m(\tau_0). \tag{2.14}$$

The blow-up of the inverse  $B^{-1}$  of the stability operator  $B$  close to  $\tau_0$  requires analyzing the contributions of  $\Delta$  in the unstable direction of  $B^{-1}$  separately. In fact,  $B$  possesses precisely one unstable direction denoted by  $b$  since we will show that  $\|T\|_2$  is a non-degenerate eigenvalue of  $T$ . We decompose  $\Delta$  into

$\Delta = \Theta b + r$ , where  $\Theta$  is the scalar contribution of  $\Delta$  in the direction  $b$  and  $r$  lies in the spectral subspace of  $B$  complementary to  $b$ .

We view  $\tau_0$  as fixed and consider  $\omega$  as the main variable with  $|\omega| \ll 1$ . Projecting (2.14) onto  $b$  and its complement yield the scalar-valued cubic equation

$$\psi\Theta(\omega)^3 + \sigma\Theta(\omega)^2 + \pi\omega = \mathcal{O}(|\omega||\Theta(\omega)| + |\Theta(\omega)|^4) \tag{2.15}$$

with two parameters  $\psi \geq 0$  and  $\sigma \in \mathbb{R}$ . In fact, the 1/3-Hölder continuity of  $m$  implies  $\Theta = \mathcal{O}(|\omega|^{1/3})$  and, hence, the right-hand side of (2.15) is indeed of lower order than the terms on the left-hand side. Analyzing (2.15) instead of (2.14) is a more tractable problem since we have reduced a quadratic  $\mathcal{A}$ -valued equation, (2.14), to the scalar-valued cubic equation, (2.15).

The essential feature of the cubic equation (2.15) is its stability. By this, we mean that there exists a constant  $c > 0$  such that

$$\psi + \sigma^2 \geq c.$$

This bound will follow from the structure of the Dyson equation and prevents any singularities of higher order than  $\omega^{1/2}$  or  $\omega^{1/3}$ . Obtaining more detailed information about  $\Theta$  from (2.15) requires applying Cardano’s formula with an error term. Therefore, we switch to *normal coordinates*,  $(\omega, \Theta(\omega)) \rightarrow (\lambda, \Omega(\lambda))$ , in (2.15). We will study four normal forms, one quadratic  $\Omega(\lambda)^2 + \Lambda(\lambda) = 0$ , and three cubics,  $\Omega(\lambda)^3 + \Lambda(\lambda) = 0$  and  $\Omega(\lambda)^3 \pm 3\Omega(\lambda) + 2\Lambda(\lambda) = 0$ , where  $\Lambda(\lambda)$  is a perturbation of the identity map  $\lambda \mapsto \lambda$ . The first case corresponds to the square root singularity of the isolated edge, the second is the cusp. The last two cases describe the situation of *almost cusps*, see later.

The correct branches in Cardano’s formula are identified with the help of four *selection principles* for the solution  $\Omega(\lambda)$  corresponding to  $\Theta$  of the cubic equation in normal form (see SP1 to SP4’ at the beginning of Section 7.2 below). These selection principles are special properties of  $\Omega$  which originate from the continuity of  $m$ ,  $\text{Im } m \geq 0$  and the Stieltjes transform representation, (2.4), of  $m$ . Once the correct branch is chosen, we obtain the precise behaviour of  $\text{Im } m$  around  $\tau_0$ , where  $\tau_0 \in \text{supp } \rho$  satisfies  $\rho(\tau_0) = 0$  or even  $\rho(\tau_0) \ll 1$ , from Cardano’s formula and careful estimates of  $r$  in the decomposition  $\Delta = \Theta b + r$  (see Theorem 7.1 below).

STRUCTURE OF THE REMAINDER OF THE PAPER

We conclude this section with an overview of the structure of the remaining sections of the present work. In Section 3, we prove Proposition 2.1 and a few basic estimates on the solution  $m$  of the Dyson equation, (2.3). Section 4 is devoted to the analysis of the regularity of  $m$  and the self-consistent density of states  $\rho$  and contains the proofs of Proposition 2.3 and Proposition 2.4. Section 5 and Section 6 prepare the proof of Theorem 2.5 in Section 7. In Section 5, we analyse the stability operator in the regime where the self-consistent density of states  $\rho$  is small. In particular, we obtain that it has one unstable direction

which we determine through the use of non-Hermitian perturbation theory. Section 6 contains the derivation of the cubic equation (cf. (2.15)) describing the component of  $m$  in this unstable direction as a function of the spectral parameter  $z$ . In Section 7, we then obtain Theorem 2.5 as a consequence of its extended version, Theorem 7.1. The latter result is also proved in Section 7 by a detailed analysis of the cubic equation from Section 6. Proposition 2.6 is shown in Section 8. Section 9 provides some sufficient checkable conditions in terms of  $a$  and  $S$  ensuring the flatness of  $S$  and the boundedness of  $m$  (see (2.7)) in the setup of Kronecker random matrices. In Section 10, we study how the solution  $m$  of the Dyson equation changes under small perturbations of the data pair  $(a, S)$ . The appendices collect a few auxiliary results applied throughout the paper as well as their proofs.

### 3 THE SOLUTION OF THE DYSON EQUATION

In this section, we first introduce some notations used in the proof of Proposition 2.1, then prove the proposition and finally give a few further properties of  $m$ .

For  $x, y \in \mathcal{A}$ , we introduce the bounded operator  $C_{x,y}: \mathcal{A} \rightarrow \mathcal{A}$  defined through  $C_{x,y}[h] := xhy$  for  $h \in \mathcal{A}$ . We set  $C_x := C_{x,x}$ . For  $x, y \in \mathcal{A}$ , the operator  $C_{x,y}$  satisfies the simple relations

$$C_{x,y}^* = C_{x^*,y^*}, \quad C_{x,y}^{-1} = C_{x^{-1},y^{-1}},$$

where  $C_{x,y}^*$  is the adjoint with respect to the scalar product defined in (2.1). Here, the second identity holds if  $x$  and  $y$  are invertible in  $\mathcal{A}$ . In fact,  $C_{x,y}$  is invertible if and only if  $x$  and  $y$  are invertible in  $\mathcal{A}$ .

In the following, we will often use the functional calculus for normal elements of  $\mathcal{A}$ . As we will explain now, our setup allows for a direct way to represent  $\mathcal{A}$  as a subalgebra of the bounded operators on a Hilbert space. Therefore, one can think of the functional calculus being performed on this Hilbert space. The Hilbert space is the completion of  $\mathcal{A}$  equipped with the scalar product defined in (2.1) and denoted by  $L^2$ . In order to represent  $\mathcal{A}$  as subalgebra of the bounded operators  $B(L^2)$  on  $L^2$ , we denote by  $\ell_x$  for  $x \in \mathcal{A}$  the left-multiplication on  $L^2$  by  $x$ , i.e.,  $\ell_x: L^2 \rightarrow L^2$ ,  $\ell_x(y) = xy$  for  $y \in L^2$ . The inclusion  $\mathcal{A} \subset L^2$  and the Cauchy-Schwarz inequality yield the well-definedness of  $\ell_x$  and  $\ell_x \in B(L^2)$ , the bounded linear operators on  $L^2$ . In fact,

$$\mathcal{A} \rightarrow B(L^2), \quad x \mapsto \ell_x$$

defines a faithful representation of  $\mathcal{A}$  as a von Neumann algebra in  $B(L^2)$  [42, Theorem 2.22].

We now introduce the *balanced polar decomposition* of  $m$ . For any  $z \in \mathbb{H}$ , we define  $w = w(z) \in \mathcal{A}$ ,  $q = q(z) \in \mathcal{A}$  and  $u = u(z) \in \mathcal{A}$  through

$$\begin{aligned} w &:= (\operatorname{Im} m)^{-1/2} (\operatorname{Re} m) (\operatorname{Im} m)^{-1/2} + i\mathbb{1}, & q &:= |w|^{1/2} (\operatorname{Im} m)^{1/2}, \\ u &:= \frac{w}{|w|}, \end{aligned} \tag{3.1}$$



where  $|w|$  and  $w/|w|$  are determined by the spectral calculus of the self-adjoint operator  $(\operatorname{Im} m)^{-1/2}(\operatorname{Re} m)(\operatorname{Im} m)^{-1/2}$ . Thus, we have

$$m(z) = \operatorname{Re} m(z) + i\operatorname{Im} m(z) = q^*uq. \tag{3.2}$$

Here,  $u$  is unitary and commutes with  $w$ . The decomposition  $m = q^*uq$  was already introduced and also called balanced polar decomposition in [4] in the special setting of matrix algebras. The operators  $|w|^{1/2}$ ,  $q$  and  $u$  correspond to  $\mathbf{W}$ ,  $\mathbf{W}\sqrt{\operatorname{Im} \mathbf{M}}$  and  $\mathbf{U}^*$  in the notation of [4], respectively. With the definitions in (3.1), (2.3) reads as

$$-u^* = q(z - a)q^* + F[u], \tag{3.3}$$

where we introduced the *saturated self-energy operator*

$$F := C_{q,q^*}SC_{q^*,q}. \tag{3.4}$$

It is positivity-preserving as well as symmetric,  $F = F^*$ , and corresponds to the saturated self-energy operator  $\mathcal{F}$  in [4].

*Proof of Proposition 2.1.* The existence of  $v$  will be a consequence of the following lemma which will be proven in Appendix A below.

LEMMA 3.1. *Let  $\mathcal{A}$  be a von Neumann algebra with unit  $\mathbf{1}$  and a tracial, faithful, normal state  $\langle \cdot \rangle: \mathcal{A} \rightarrow \mathbb{C}$ . If  $h: \mathbb{H} \rightarrow \mathcal{A}$  is a holomorphic function satisfying  $\operatorname{Im} h(z) \in \mathcal{A}_+$  for all  $z \in \mathbb{H}$  and*

$$\lim_{\eta \rightarrow \infty} i\eta h(i\eta) = -\mathbf{1} \tag{3.5}$$

*then there exists a unique measure  $v: \mathcal{B} \rightarrow \mathcal{A}$  on the Borel sets  $\mathcal{B}$  of  $\mathbb{R}$  with values in  $\overline{\mathcal{A}}_+$  such that*

$$h(z) = \int_{\mathbb{R}} \frac{v(d\tau)}{\tau - z} \tag{3.6}$$

*for all  $z \in \mathbb{H}$  and  $v(\mathbb{R}) = \mathbf{1}$ .*

In order to apply Lemma 3.1, we have to verify (3.5) for  $h = m$ . To that end, we take the imaginary part of (2.3) and use  $\operatorname{Im} m \geq 0$  as well as  $S \in \Sigma$  to conclude

$$-\operatorname{Im} m^{-1}(z) = \operatorname{Im} z\mathbf{1} + S[\operatorname{Im} m] \geq \operatorname{Im} z\mathbf{1}.$$

Hence,  $\|m(z)\| \leq (\operatorname{Im} z)^{-1}$  as for any  $x \in \mathcal{A}$  we have  $\|x\| \leq 1$  if  $x$  is invertible and  $\operatorname{Im} x^{-1} \geq \mathbf{1}$ . Therefore, evaluating (2.3) at  $z = i\eta$ ,  $\eta > 0$ , and multiplying the result by  $m$  from the left yield

$$i\eta m(i\eta) = -\mathbf{1} + m(i\eta)a - m(i\eta)S[m(i\eta)] \rightarrow -\mathbf{1}$$

for  $\eta \rightarrow \infty$  as  $S$  is bounded. Hence, Lemma 3.1 implies the existence of  $v$ , i.e., the Stieltjes transform representation of  $m$  in (2.4).

This representation has the following well-known bounds as a direct consequence (e.g. [3, 4, 7]).

LEMMA 3.2. *Let  $v$  be the measure in Proposition 2.1 and  $\rho = \langle v \rangle$ . Then, for any  $z \in \mathbb{H}$ , we have*

$$\|m(z)\| \leq \frac{1}{\text{dist}(z, \text{supp } \rho)}, \quad \text{Im } m(z) \leq \frac{\text{Im } z}{\text{dist}(z, \text{supp } \rho)^2} \mathbf{1}. \quad (3.7)$$

□

For the proofs of (2.5a) and (2.5b), we refer to the proofs of Proposition 2.1 in [4] and (3.4) in [7] in the matrix setup, the same argument works for our general setup as well.

We now prove (2.6). Taking the imaginary part of the Dyson equation, (3.3), yields

$$\text{Im } u = (\text{Im } z)qq^* + F[\text{Im } u] \geq \max\{(\text{Im } z)qq^*, F[\text{Im } u]\}.$$

Thus,  $\text{Im } u \geq (\text{Im } z)\|(qq^*)^{-1}\|^{-1}\mathbf{1}$ . We remark that  $qq^*$  is invertible since  $\text{Im } m(z) > 0$  for  $z \in \mathbb{H}$ . Therefore, the following Lemma 3.3 with  $h = \text{Im } u/\|\text{Im } u\|_2$  implies  $\|F\|_2 \leq 1$ .

LEMMA 3.3. *Let  $T: \mathcal{A} \rightarrow \mathcal{A}$  be a positivity-preserving operator which is symmetric with respect to (2.1). If there are  $h \in \mathcal{A}$  and  $\varepsilon > 0$  such that  $h \geq \varepsilon\mathbf{1}$  and  $Th \leq h$  then  $\|T\|_2 \leq 1$ .*

*Proof.* The argument in the proof of Lemma 4.6 in [3] also yields this lemma in our current setup. □

We rewrite the Dyson equation (3.3) in the form

$$q(a - z)q^* = u^* + F[u]. \quad (3.8)$$

We take the  $\|\cdot\|_2$ -norm on both sides of (3.8) and use that  $\|u\|_2 = 1$  (since it is unitary) and  $\|F\|_2 \leq 1$  to find

$$\|q(a - z)q^*\|_2 \leq 2. \quad (3.9)$$

Then we use the polar decomposition  $m = q^*uq$  again and with  $z = \tau + i\eta$  find

$$\begin{aligned} \langle m, (C_{a-\tau} + \eta^2)m \rangle &= \text{Re} \langle m, C_{a-z, (a-z)^*}m \rangle \\ &\leq |\langle m, C_{a-z, (a-z)^*}m \rangle| \\ &= |\langle q(a - z)q^*, C_{u^*, u}[q(a - z)q^*] \rangle| \leq 4, \end{aligned}$$

where the last step holds because of (3.9). Recall that  $a = a^*$ . Since  $\text{Spec}(C_{a-\tau}) = \{\lambda\mu : \lambda, \mu \in \text{Spec}(a - \tau)\}$  we have  $\inf \text{Spec}(C_{a-\tau}) \geq \text{dist}(\tau, \text{Conv Spec } a)^2$ , provided  $\tau \notin \text{Conv Spec } a$ . Thus in this case (2.6) follows. In case  $\tau \in \text{Conv Spec } a$  we simply use the trivial bound  $\|m\|_2 \leq \|m\| \leq \eta^{-1}$  from the first inequality of (3.7) and (2.6) still holds. This completes the proof of Proposition 2.1. □

From now on until the end of Section 4.2, we will always assume that  $S$  is *flat*, i.e.,  $S \in \Sigma_{\text{flat}}$  (cf. (2.2b)). In fact, all of our estimates will be uniform in all data pairs  $(a, S)$  that satisfy

$$c_1 \langle x \rangle \mathbf{1} \leq S[x] \leq c_2 \langle x \rangle \mathbf{1}, \quad \|a\| \leq c_3 \tag{3.10}$$

for all  $x \in \mathcal{A}_+$  with the some fixed constants  $c_1, c_2, c_3 > 0$ . Therefore, the constants  $c_1, c_2, c_3$  from (3.10) are called *model parameters* and we introduce the following convention.

CONVENTION 3.4 (Comparison relation, universal constant). *Let  $x, y \in \mathcal{A}_{\text{sa}}$ . We write  $x \lesssim y$ , or, equivalently,  $y \gtrsim x$ , if there is  $c > 0$  depending only on the model parameters  $c_1, c_2, c_3$  from (3.10) such that  $cy - x$  is positive definite, i.e.,  $cy - x \in \overline{\mathcal{A}}_+$ . Moreover, we write  $x \sim y$  if  $x \lesssim y$  and  $x \gtrsim y$ . We also use this notation for scalars  $x, y$ . Moreover, we write  $x = y + \mathcal{O}(\alpha)$  for  $x, y \in \mathcal{A}$  and  $\alpha > 0$  if  $\|x - y\| \lesssim \alpha$ . Constants that do not depend on any parameter are called universal constants. Their numerical values can easily be obtained from our proofs, but they are irrelevant and we do not follow them.*

We remark that we will choose a different set of model parameters starting from Section 4.3 below and redefine  $\lesssim$  as well as  $\sim$  accordingly (cf. Convention 4.6 below).

PROPOSITION 3.5 (Properties of the solution). *Let  $(a, S)$  be a data pair satisfying (3.10) and  $m$  be the solution to the associated Dyson equation, (2.3). We have*

$$\|m(z)\|_2 \lesssim 1, \tag{3.11}$$

$$\|m(z)\| \lesssim \frac{1}{\langle \text{Im } m(z) \rangle + \text{dist}(z, \text{supp } \rho)}, \tag{3.12}$$

$$\|m(z)^{-1}\| \lesssim 1 + |z|, \tag{3.13}$$

$$\langle \text{Im } m(z) \rangle \mathbf{1} \lesssim \text{Im } m(z) \lesssim (1 + |z|^2) \|m(z)\|^2 \langle \text{Im } m(z) \rangle \mathbf{1}, \tag{3.14}$$

$$\|m(z)\| \gtrsim (1 + |z|)^{-1} \tag{3.15}$$

*uniformly for  $z \in \mathbb{H}$ .*

The following proof of Proposition 3.5 resembles the one of Proposition 4.2 in [4].

*Proof.* We take the imaginary part of the Dyson equation, (2.3), multiply the result by  $m^*$  from the left and  $m$  from the right and obtain

$$\text{Im } m = m^* (\langle \text{Im } z \rangle \mathbf{1} + S[\text{Im } m]) m \geq m^* S[\text{Im } m] m \gtrsim \langle \text{Im } m \rangle m^* m \tag{3.16}$$

due to the lower bound on  $S$  from (3.10). We apply  $\langle \cdot \rangle$  on both sides and conclude (3.11).

We now set  $K := \|a\| + 2\|S\|^{1/2}$ . Note that  $K \lesssim 1$  since  $\|a\| \lesssim 1$  and  $\|S\| \leq \|S\|_{2 \rightarrow \cdot, \cdot} \lesssim 1$  by the upper bounds on  $a$  and  $S$  in (3.10) and Lemma B.2 (i). The Stieltjes transform representation (2.4) implies that  $\|m(z)\| \leq \text{dist}(z, \text{supp } v)^{-1}$ . Since  $\text{supp } \rho = \text{supp } v$  by the faithfulness of  $\langle \cdot, \cdot \rangle$ , for the proof of (3.12), it thus suffices to show that  $\|m(z)\| \lesssim \langle \text{Im } m(z) \rangle^{-1}$  for  $|z| \leq 1 + K$  as  $\text{dist}(z, \text{supp } \rho) \gtrsim 1 \gtrsim \langle \text{Im } m(z) \rangle$  for  $|z| \geq 1 + K$  by (3.11). Applying the lower bound on  $S$  from (3.10) to the imaginary part of the Dyson equation, (2.3), yields

$$-\text{Im } m(z)^{-1} \geq S[\text{Im } m(z)] \gtrsim \langle \text{Im } m(z) \rangle \mathbf{1}.$$

Since  $\text{Im } x^{-1} \geq 1$  implies  $\|x\| \leq 1$  for any invertible  $x \in \mathcal{A}$ , we conclude  $\|m(z)\| \lesssim \langle \text{Im } m(z) \rangle^{-1}$ , thus, completing the proof of (3.12).

For the proof of (3.13), we start from (2.3) and estimate

$$\|m(z)^{-1}\| \leq |z| + \|a\| + \|S\|_{2 \rightarrow \cdot, \cdot} \|m\|_2 \lesssim 1 + |z|,$$

where we used  $\|a\| \lesssim 1$ ,  $\|m\|_2 \lesssim 1$  and  $\|S\|_{2 \rightarrow \cdot, \cdot} \lesssim 1$  in the last step. These bounds follows from (3.10), (3.11) and Lemma B.2 (i). This proves (3.13) and, thus, (3.15) since  $\|m\| \|m^{-1}\| \geq 1$ .

What remains is showing (3.14). By (2.5a), we know that  $\text{supp } v \subset [-K, K]$ . First, we assume that  $|z| \geq 1 + K$ . The Stieltjes transform representation (2.4) implies that  $z \mapsto \langle x, m(z)x \rangle$  is the Stieltjes transform of a probability measure supported in  $[-K, K]$  for any  $x \in \mathcal{A}$  such that  $\|x\|_2 = 1$ . Hence,  $\text{Im } m(z) \sim \langle \text{Im } m(z) \rangle \sim |z|^{-2} \text{Im } z$  if  $|z| \geq 1 + K$ . Moreover,  $\|m(z)\| \sim |z|^{-1}$  in this regime. Thus, we proved (3.14) if  $|z| \geq 1 + K$ .

We now assume that  $|z| \leq 1 + K$ . From (3.16) we conclude that  $\text{Im } m \gtrsim \langle \text{Im } m \rangle \|m^{-1}\|^{-2} \mathbf{1}$ , and, thus, the lower bound on  $\text{Im } m$  in (3.14) as  $\|m^{-1}\| \lesssim 1$  by (3.13). For the upper bound, we take the imaginary part of the Dyson equation and estimate

$$\text{Im } m = m^*(\text{Im } z \mathbf{1} + S[\text{Im } m])m \lesssim (\text{Im } z + \langle \text{Im } m \rangle) m^* M \lesssim \langle \text{Im } m \rangle \|m\|^2 \mathbf{1},$$

where we used in the last step that  $\text{Im } m \gtrsim (\text{Im } z) \mathbf{1}$  by the Stieltjes transform representation (2.4) and  $|z| \lesssim 1$ . This proves the missing upper bound on  $\text{Im } m$  in (3.14) and, thus, completes the proof of Proposition 3.5.  $\square$

#### 4 REGULARITY OF THE SOLUTION AND THE SELF-CONSISTENT DENSITY OF STATES

In this section, we will prove Proposition 2.3 and Proposition 2.4. Their proofs are based on a bound on the inverse of the stability operator  $\text{Id} - C_m S$  of the Dyson equation, (2.3), which will be given in Proposition 4.1 below. These results are the basis of the analysis of the singularities in the subsequent sections.

##### 4.1 LINEAR STABILITY OF THE DYSON EQUATION

For the formulation of the following proposition, we introduce the harmonic extension of the density of states  $\rho$  defined in Definition 2.2 to  $\mathbb{H}$ . The harmonic

extension at  $z \in \mathbb{H}$  is denoted by  $\rho(z)$  and given by

$$\rho(z) := \frac{1}{\pi} \langle \text{Im } m(z) \rangle.$$

PROPOSITION 4.1 (Linear Stability). *There is a universal constant  $C > 0$  such that, for the solution  $m$  to (2.3) associated to any  $a \in \mathcal{A}_{\text{sa}}$  and  $S \in \Sigma$  satisfying (3.10), we have*

$$\|(\text{Id} - C_{m(z)}S)^{-1}\|_2 \lesssim 1 + \frac{1}{(\rho(z) + \text{dist}(z, \text{supp } \rho))^C} \tag{4.1}$$

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

Before proving Proposition 4.1, we will explain how the linear stability yields the Hölder-continuity and analyticity of  $\rho$  in Proposition 2.3. Indeed, assuming that  $m \equiv m(a, S, z)$  viewed as a function of  $(a, S, z) \in \mathcal{A}_{\text{sa}} \times \Sigma_{\text{flat}} \times \mathbb{H}$  depends differentiably on  $(a, S, z)$ , we can compute the directional derivative  $\nabla_{(d,D,\delta)}$  at  $(a, S, z)$  of both sides in (2.3). The result of this computation is

$$(\text{Id} - C_m S)[\nabla_{(d,D,\delta)} m] = m(\delta - d + D[m])m.$$

Using the bound in Proposition 4.1 and  $\rho(z) = \pi^{-1} \langle \text{Im } m(z) \rangle$ , we conclude from (3.12) that

$$|\nabla_{(d,D,\delta)} \rho| \leq \frac{1}{\rho^C} (|\delta| + \|d\| + \|D\|) \tag{4.2}$$

with a possibly larger  $C$ . Therefore, it is clear that the control on  $(\text{Id} - C_m S)^{-1}$  will be the key input in the proof of Proposition 2.3.

In order to prove Proposition 4.1, we will use the representation

$$\text{Id} - C_m S = C_{q^*,q} C_u (C_u^* - F) C_{q^*,q}^{-1}, \tag{4.3}$$

where  $q$ ,  $u$  and  $F$  were defined in (3.1) and (3.4), respectively. This representation has the advantage that  $C_u^*$  is unitary and  $F$  is symmetric. Hence, it is much easier to obtain some spectral properties for  $C_u^* - F$  compared to  $\text{Id} - C_m S$ . Now, we will first analyze  $q$  and  $F$  in the following two lemmas and then use this knowledge to verify Proposition 4.1.

LEMMA 4.2. *If (3.10) holds true then we have*

$$\|q(z)\| \lesssim (1 + |z|)^{1/2} \|m(z)\|, \quad \|q(z)^{-1}\| \lesssim (1 + |z|) \|m(z)\|^{1/2}$$

uniformly for  $z \in \mathbb{H}$ .

*Proof.* For  $q = q(z)$ , we will show below that

$$\frac{A^{1/2}}{B^{1/2}} \|m(z)^{-1}\|^{-1} \mathbf{1} \leq q^* q \leq \frac{B^{1/2}}{A^{1/2}} \|m(z)\| \mathbf{1} \tag{4.4}$$

if  $A\mathbb{1} \leq \operatorname{Im} m(z) \leq B\mathbb{1}$  for some  $A, B \in (0, \infty)$  and  $z \in \mathbb{H}$ . Choosing  $A$  and  $B$  according to (3.14), using the  $C^*$ -property of  $\|\cdot\|$ ,  $\|q^*q\| = \|q\|^2$ , and (3.13), we immediately obtain Lemma 4.2.

For the proof of (4.4), we set  $g := \operatorname{Re} m$  and  $h := \operatorname{Im} m$ . Using the monotonicity of the square root, we compute

$$\begin{aligned} q^*q &= h^{1/2}(\mathbb{1} + h^{-1/2}gh^{-1}gh^{-1/2})^{1/2}h^{1/2} \\ &\leq A^{-1/2}h^{1/2}(h^{-1/2}(h^2 + g^2)h^{-1/2})^{1/2}h^{1/2} \\ &\leq \|m\|A^{-1/2}h^{1/2}. \end{aligned}$$

Here, we employed  $h^{-1} \leq A^{-1}\mathbb{1}$  as well as  $\mathbb{1} \leq A^{-1}h$  in the first step and  $(\operatorname{Re} m)^2 + (\operatorname{Im} m)^2 = (m^*m + mm^*)/2 \leq \|m\|^2$  in the second step. Thus,  $h \leq B\mathbb{1}$  yields the upper bound in (4.4). Similar estimates using  $\mathbb{1} \geq B^{-1}h$  and  $\|m^{-1}\|^{-2} \leq (m^*m + mm^*)/2$  prove the lower bound in (4.4) which completes the proof of the lemma.  $\square$

LEMMA 4.3 (Properties of  $F$ ). *If the bounds in (3.10) are satisfied then  $\|F\|_2$  is a simple eigenvalue of  $F: \mathcal{A} \rightarrow \mathcal{A}$  defined in (3.4). Moreover, there is a unique eigenvector  $f \in \mathcal{A}_+$  such that  $F[f] = \|F\|_2 f$  and  $\|f\|_2 = 1$ . This eigenvector satisfies*

$$1 - \|F\|_2 = (\operatorname{Im} z) \frac{\langle f, qq^* \rangle}{\langle f, \operatorname{Im} u \rangle}. \tag{4.5}$$

*In particular,  $\|F\|_2 \leq 1$ . Furthermore, the following properties hold true uniformly for  $z \in \mathbb{H}$  satisfying  $|z| \leq 3(1 + \|a\| + \|S\|^{1/2})$  and<sup>3</sup>  $\|F(z)\|_2 \geq 1/2$ :*

(i) *The eigenvector  $f$  has upper and lower bounds*

$$\|m\|^{-4}\mathbb{1} \lesssim f \lesssim \|m\|^4\mathbb{1}. \tag{4.6}$$

(ii) *The operator  $F$  has a spectral gap  $\vartheta \in (0, 1]$  satisfying  $\vartheta \gtrsim \|m\|^{-28}$  and*

$$\operatorname{Spec}(F/\|F\|_2) \subset [-1 + \vartheta, 1 - \vartheta] \cup \{1\}. \tag{4.7}$$

*Proof.* The definition of  $F$  in (3.4), (3.10) and Lemma 4.2 imply

$$(1 + |z|)^{-4}\|m(z)\|^{-2}\langle a \rangle \mathbb{1} \lesssim F[a] \lesssim (1 + |z|)^2\|m(z)\|^4\langle a \rangle \mathbb{1} \tag{4.8}$$

for all  $a \in \mathcal{A}_+$  and all  $z \in \mathbb{H}$ . We will use Lemma B.1 (ii) from Appendix B. The condition (B.1) with  $T = F$  is satisfied by (4.8) with constants depending on  $\|m\|$  and  $|z|$ . Hence, Lemma B.1 (ii) implies the existence and uniqueness of the eigenvector  $f$ . We compute the scalar product of  $f$  with the imaginary part of (3.3). Since  $F$  is symmetric, this immediately yields (4.5).

We now assume that  $z \in \mathbb{H}$  satisfies  $|z| \leq 3(1 + \|a\| + \|S\|^{1/2})$  and  $\|F(z)\|_2 \geq 1/2$ . Then  $|z| \lesssim 1$  and, by using this in (4.8), we thus obtain (4.6) and (4.7) from Lemma B.1 (ii) since  $\|m\| \gtrsim 1$  by (3.13).  $\square$

<sup>3</sup>The lower bound  $1/2$  is chosen solely for convenience. Any other constant in  $(0, 1)$  could be used instead.

The following proof of Proposition 4.1 proceeds similarly to the one of Proposition 4.4 in [4].

*Proof of Proposition 4.1.* We will distinguish several cases. If  $|z| \geq 3(1 + \kappa)$  with  $\kappa := \|a\| + 2\|S\|^{1/2}$  then we conclude from (2.4) and  $\text{supp } \rho \subset [-\kappa, \kappa]$  by (2.5a) that  $\|m(z)\| \leq (|z| - \kappa)^{-1}$ . Thus,

$$\|C_{m(z)}S\|_2 \leq \frac{\|S\|_2}{(|z| - \kappa)^2} \leq \frac{\|S\|}{4(1 + \kappa)^2} \leq \frac{1}{4}.$$

Here, we used  $\|S\|_2 \leq \|S\|$  since  $S$  is symmetric and  $\kappa \geq \|S\|^{1/2}$ . This shows (4.1) for large  $|z|$ .

Next, we assume  $|z| \leq 3(1 + \kappa)$ . In this regime, we use the alternative representation of  $\text{Id} - C_mS$  in (4.3) and the spectral properties of  $F$  from Lemma 4.3. Indeed, from (4.3) and Lemma 4.2, we conclude

$$\begin{aligned} \|(\text{Id} - C_mS)^{-1}\|_2 &\lesssim \|m\|^3 \|(C_u^* - F)^{-1}\|_2 \\ &\lesssim \frac{1}{(\rho(z) + \text{dist}(z, \text{supp } \rho))^3} \|(C_u^* - F)^{-1}\|_2 \end{aligned} \tag{4.9}$$

as  $u \in \mathcal{A}$  is unitary. Here, we used (3.12) in the last step. If  $\|F(z)\|_2 \leq 1/2$  then this immediately yields (4.1) as  $\|C_u\|_2 = 1$ . We now assume  $\|F(z)\|_2 \geq 1/2$ . In this case, we will use the following lemma.

LEMMA 4.4 (Rotation-Inversion Lemma). *Let  $U$  be a unitary operator on  $L^2$  and  $T$  a symmetric operator on  $L^2$ . We assume that  $\|T\|_2 \leq 1$  is an eigenvalue of  $T$  whose eigenspace is one-dimensional. Moreover, suppose that  $T$  has a spectral gap, i.e., there is a constant  $\theta > 0$  such that*

$$\text{Spec } T \subset [-\|T\|_2 + \theta, \|T\|_2 - \theta] \cup \{\|T\|_2\}.$$

*Then there is a universal constant  $C > 0$  such that*

$$\|(U - T)^{-1}\|_2 \leq \frac{C}{\theta|1 - \|T\|_2 \langle t, U[t] \rangle|},$$

*where  $t \in L^2$  is the normalized,  $\|t\|_2 = 1$ , eigenvector of  $T$  corresponding to  $\|T\|_2$ .*

The proof of this lemma is identical to the proof of Lemma 5.6 in [1], where a result of this type was first applied in the context of vector Dyson equations. We start from the estimate (4.9), use the Rotation-Inversion Lemma, Lemma 4.4, with  $U = C_u^*$  and  $T = F$  as well as (4.7) and (3.12) and obtain

$$\begin{aligned} \|(\text{Id} - C_mS)^{-1}\|_2 &\lesssim \frac{(\rho(z) + \text{dist}(z, \text{supp } \rho))^{-31}}{|1 - \|F\|_2 \langle f, C_u^*[f] \rangle|} \\ &\leq \frac{(\rho(z) + \text{dist}(z, \text{supp } \rho))^{-31}}{\max\{1 - \|F\|_2, |1 - \langle fC_u^*[f] \rangle|\}}. \end{aligned}$$

In order to complete the proof of (4.1), we now show that

$$\max\{1 - \|F\|_2, |1 - \langle fC_u^*[f] \rangle|\} \gtrsim (\rho(z) + \text{dist}(z, \text{supp } \rho))^6. \tag{4.10}$$

We first prove auxiliary upper and lower bounds on  $\text{Im } u = (q^*)^{-1}(\text{Im } m)q^{-1}$ . We have

$$\rho(z)(\rho(z) + \text{dist}(z, \text{supp } \rho))^2 \mathbf{1} \lesssim \text{Im } u \lesssim \frac{\text{Im } z \|m\|}{\text{dist}(z, \text{supp } \rho)^2} \mathbf{1}. \tag{4.11}$$

For the lower bound, we used the lower bound in (3.14), Lemma 4.2 and (3.12). The upper bound is a direct consequence of (3.7) as well as Lemma 4.2. Since  $\langle f, qq^* \rangle \geq \|(qq^*)^{-1}\|^{-1} \langle f \rangle \gtrsim \|m\| \langle f \rangle$  by Lemma 4.2, the relation (4.5) and the upper bound in (4.11) yield

$$1 - \|F\|_2 \gtrsim \text{dist}(z, \text{supp } \rho)^2.$$

As  $1 - \langle fC_{\text{Re } u}[f] \rangle \geq 0$  and  $\langle f^2 \rangle = 1$ , we obtain from the lower bound in (4.11) that

$$\begin{aligned} |1 - \langle fC_u^*[f] \rangle| &\geq \text{Re}[1 - \langle fC_u^*[f] \rangle] = 1 - \langle fC_{\text{Re } u}[f] \rangle + \langle fC_{\text{Im } u}[f] \rangle \\ &\gtrsim \rho(z)^2(\rho(z) + \text{dist}(z, \text{supp } \rho))^4. \end{aligned} \tag{4.12}$$

This completes the proof of (4.10) and hence of Proposition 4.1. □

#### 4.2 REGULARITY OF THE SELF-CONSISTENT DENSITY OF STATES – PROOF OF PROPOSITION 2.3

The following proof of Proposition 2.3 is similar to the one of Proposition 2.2 in [4].

*Proof of Proposition 2.3.* We first show that  $\rho: \mathbb{H} \rightarrow (0, \infty)$  has a uniformly Hölder-continuous extension to  $\overline{\mathbb{H}}$ , which we will also denote by  $\rho$ . This extension restricted to  $\mathbb{R}$  will be the density of the measure  $\rho$  from Definition 2.2. Since  $\text{Id} - C_m S$  is invertible for each  $z \in \mathbb{H}$  by (4.1), the implicit function theorem allows us to differentiate (2.3) with respect to  $z$ . This yields

$$(\text{Id} - C_m S)[\partial_z m] = m^2. \tag{4.13}$$

Since  $z \mapsto \langle m(z) \rangle$  is holomorphic on  $\mathbb{H}$  as remarked below (2.3), we have  $2\pi i \partial_z \rho(z) = 2i \partial_z \text{Im} \langle m(z) \rangle = \partial_z \langle m(z) \rangle$ . Thus, we obtain from (4.13) that

$$|\nabla_z \rho| \lesssim |\partial_z \rho| + |\partial_z \rho| \lesssim \|\partial_z m\|_2 \leq \|(\text{Id} - C_m S)^{-1}\|_2 \|m\|^2 \lesssim \rho^{-(C+2)} \tag{4.14}$$

Here, we used (4.1),  $\rho(z) \lesssim \|m(z)\|_2 \lesssim 1$  by (3.11) and (3.12) in the last step. Hence,  $\rho^{C+3}$  is a uniformly Lipschitz-continuous function on  $\mathbb{H}$ . Therefore,  $\rho$  defines uniquely a uniformly  $1/(C+3)$ -Hölder continuous function on  $\mathbb{R}$  which is a density of the measure  $\rho$  from Definition 2.2 with respect to the Lebesgue measure on  $\mathbb{R}$ .



Next, we show the Hölder-continuity with respect to  $a$  and  $S$ . As before in (4.2), we compute the derivatives and use (3.12) and (4.1) to obtain

$$|\nabla_{(d,D)}\rho_{(a,S)}(z)| \lesssim |\langle \nabla_{(d,D)}m \rangle| \lesssim \frac{\|d\| + \|D\|}{\rho^{C+3}}.$$

Since the constants in (4.1) and (3.12) depend on the constants in (3.10), we conclude that  $\rho$  is also a locally  $1/(C + 4)$ -Hölder continuous function of  $a$  and  $S$ .

We are left with showing that  $\rho$  is real-analytic in a neighbourhood of  $(a, S, \tau_0)$  if  $\rho_{a,S}(\tau_0) > 0$ . Since  $\rho(\tau_0) > 0$ , we can extend  $m$  to  $\tau_0$  by (4.14). Moreover,  $m(\tau_0)$  is invertible as  $\text{Im } m(\tau_0) > 0$  and, thus, solves (2.3) with  $z = \tau_0$ . Since (2.3) depends analytically on  $a, S$  and  $z = \tau$  in a small neighbourhood of  $(a, S, \tau_0)$ , the solution  $m$  and thus  $\rho$  will depend analytically on  $(a, S, \tau)$  in this neighbourhood by the implicit function theorem. This completes the proof of Proposition 2.3.  $\square$

### 4.3 REGULARITY OF THE SOLUTION – PROOF OF PROPOSITION 2.4

In this section, we prove Proposition 2.4. In order to modularize the following arguments, we first collect in Assumptions 4.5 below the essential conditions on  $m$  and  $F$  implying the conclusion of Proposition 2.4. Finally, the proof of Proposition 2.4 will consist of checking that Assumptions 4.5 are satisfied if  $S \in \Sigma_{\text{flat}}$  and (2.7) holds.

To formulate Assumptions 4.5, for any  $I \subset \mathbb{R}$  and  $\eta_* > 0$ , we introduce the set

$$\mathbb{H}_{I,\eta_*} := \{z \in \mathbb{H} : \text{Re } z \in I, \text{Im } z \in (0, \eta_*]\} \tag{4.15}$$

and its closure  $\overline{\mathbb{H}}_{I,\eta_*}$ .

ASSUMPTIONS 4.5. Let  $m$  be the solution of (2.3) for  $a = a^* \in \mathcal{A}$  satisfying  $\|a\| \leq k_1$  with a positive constant  $k_1$  and  $S \in \Sigma$  satisfying  $\|S\|_{2 \rightarrow \cdot} \leq k_2$  for some positive constant  $k_2$ . For an interval  $I \subset \mathbb{R}$  and some  $\eta_* \in (0, 1]$ , we assume that

- (i) There are positive constants  $k_3, k_4$  and  $k_5$  such that

$$\|m(z)\| \leq k_3, \tag{4.16}$$

$$k_4 \langle \text{Im } m(z) \rangle \mathbf{1} \leq \text{Im } m(z) \leq k_5 \langle \text{Im } m(z) \rangle \mathbf{1}, \tag{4.17}$$

uniformly for all  $z \in \mathbb{H}_{I,\eta_*}$ .

- (ii) The operator  $F := C_{q,q^*} S C_{q^*,q}$  has a simple eigenvalue  $\|F\|_2$  with eigenvector  $f \in \mathcal{A}_+$  that satisfies (4.5) for all  $z \in \mathbb{H}_{I,\eta_*}$ . Moreover, (4.7) holds true and there are positive constants  $k_6, k_7$  and  $k_8$  such that

$$k_6 \mathbf{1} \leq f \leq k_7 \mathbf{1}, \quad \vartheta \geq k_8. \tag{4.18}$$

uniformly for all  $z \in \mathbb{H}_{I,\eta_*}$ .

We remark that  $S \in \Sigma_{\text{flat}}$  is not necessarily required in Assumptions 4.5. In fact, we will show in Lemma 4.8 below that  $S \in \Sigma_{\text{flat}}$  and (4.16) imply all other conditions in Assumptions 4.5.

CONVENTION 4.6 (Model parameters, Comparison relation). *For the remainder of the Section 4 as well as Section 5 and Section 6, we will only consider  $k_1, \dots, k_8$  as model parameters and understand the comparison relation  $\sim$  from Convention 3.4 with respect to this set of model parameters.*

We remark that all of our estimates will be uniform in  $\eta_* \in (0, 1]$ . Therefore,  $\eta_*$  is not considered a model parameter. At the end of this section, we will directly conclude Proposition 2.4 from the following proposition.

PROPOSITION 4.7 (Regularity of  $m$ ). *Let Assumptions 4.5 hold true on an interval  $I \subset \mathbb{R}$  for some  $\eta_* \in (0, 1]$ .*

*Then, for any  $\theta \in (0, 1]$ ,  $m$  can be uniquely extended to  $I_\theta := \{\tau \in I : \text{dist}(\tau, \partial I) \geq \theta\}$  such that it is uniformly  $1/3$ -Hölder continuous, indeed,*

$$\|m(z_1) - m(z_2)\| \lesssim \theta^{-4/3} |z_1 - z_2|^{1/3} \tag{4.19}$$

*for all  $z_1, z_2 \in I_\theta \times i[0, \infty)$ . Moreover, we have  $\|m(z)\| \lesssim 1 + \theta^{-4/3}$  for all  $z \in I_\theta \times i[0, \infty)$ .*

*Furthermore, if  $\rho(\tau_0) > 0$ ,  $\tau_0 \in I$ , then  $m$  is real-analytic in a neighbourhood of  $\tau_0$  and*

$$\|\partial_\tau m(\tau_0)\| \lesssim \rho(\tau_0)^{-2}. \tag{4.20}$$

We remark that the bound in (4.20) will be extended to higher derivatives in Lemma 5.7 below.

In the following lemma, we establish a very helpful consequence of (i) in Assumptions 4.5. Moreover, part (ii) of the following lemma shows that all conditions in Assumptions 4.5 are satisfied if we assume (4.16) and the flatness of  $S$ .

LEMMA 4.8. *Let  $m$  be the solution to (2.3) for some data pair  $(a, S) \in \mathcal{A}_{\text{sa}} \times \Sigma$ . We have*

(i) *Let  $\|a\| \lesssim 1$ ,  $\|S\| \lesssim 1$  and  $U \subset \mathbb{H}$  such that  $\sup\{|z| : z \in U\} \lesssim 1$ . If (4.16) and (4.17) hold true uniformly for  $z \in U$  then, uniformly for  $z \in U$ , we have*

$$\|q\|, \|q^{-1}\| \sim 1, \quad \text{Im } u \sim \langle \text{Im } u \rangle \mathbb{1} \sim \rho \mathbb{1}. \tag{4.21}$$

(ii) *Let  $I \subset [-C, C]$  for some  $C \sim 1$  and (4.16) hold true uniformly for all  $z \in \mathbb{H}_{I, \eta_*}$ . If  $S \in \Sigma_{\text{flat}}$  and  $\|a\| \lesssim 1$  then  $\|S\|_{2 \rightarrow \cdot} \lesssim 1$ , (4.17) holds true uniformly for all  $z \in \mathbb{H}_{I, \eta_*}$  and part (ii) of Assumptions 4.5 is satisfied.*

(iii) *If Assumptions 4.5 hold true then, uniformly for  $z \in \mathbb{H}_{I, \eta_*}$ , we have*

$$\|(\text{Id} - C_{m(z)} S)^{-1}\|_2 + \|(\text{Id} - C_{m(z)} S)^{-1}\| \lesssim \rho(z)^{-2}. \tag{4.22}$$

*Proof of Lemma 4.8.* For the proof of (i), we use  $\|a\| \lesssim 1$ ,  $\|S\| \lesssim 1$  and (2.3) to show  $\|m(z)^{-1}\| \lesssim 1$  uniformly for all  $z \in U$ . Thus, following the proof of Lemma 4.2 immediately yields the estimates on  $q$  and  $q^{-1}$  in (4.21) due to (4.16) and (4.17). Thus, as  $\|q\|, \|q^{-1}\| \sim 1$ , we obtain the missing relations in (4.21) from (4.17) since

$$\operatorname{Im} u = (q^*)^{-1}(\operatorname{Im} m)q^{-1} \sim \operatorname{Im} m \sim \langle \operatorname{Im} m \rangle \sim \langle \operatorname{Im} u \rangle.$$

We now show (ii). By Lemma B.2 (i), the upper bound in the definition of flatness, (3.10), implies  $\|S\|_{2 \rightarrow \|\cdot\|} \lesssim 1$ . Owing to (4.16) and (3.13), we have  $\|m(z)\| \sim 1$  for all  $z \in \mathbb{H}_{I, \eta_*}$ . Hence, (4.17) follows from (3.14) since  $|z| \leq C + 1$  for  $z \in \mathbb{H}_{I, \eta_*}$ . Moreover, (ii) in Assumptions 4.5 is a consequence of Lemma 4.3.

To prove (4.22), we follow the proof of Proposition 4.1 and replace the use of (3.12) as well as (4.6) and (4.7) from Lemma 4.3 by (4.16) and (4.18), respectively. This yields

$$\|(\operatorname{Id} - C_m S)^{-1}\|_2 \lesssim 1 + |1 - \|F\|_2 \langle f C_u^*[f] \rangle|^{-1} \lesssim |1 - \|F\|_2 \langle f C_u^*[f] \rangle|^{-1}, \tag{4.23}$$

where we used in the last step that (4.16) implies  $\rho(z) \lesssim 1$  on  $\mathbb{H}_{I, \eta_*}$ . Since  $\operatorname{Im} u \sim \rho$  by (4.21) and  $\|F\|_2 \leq 1$  by (4.5) that holds under Assumptions 4.5 (ii), we conclude

$$|1 - \|F\|_2 \langle f C_u^*[f] \rangle|^{-1} \lesssim |1 - \langle f C_u^*[f] \rangle|^{-1} \lesssim \rho^{-2}$$

as in (4.12) in the proof of Proposition 4.1. This shows  $\|(\operatorname{Id} - C_m S)^{-1}\|_2 \lesssim \rho(z)^{-2}$ . Using  $\|S\|_{2 \rightarrow \|\cdot\|} \lesssim 1$  and Lemma B.2 (ii), we obtain the missing  $\|\cdot\|$ -bound in (4.22). This completes the proof of Lemma 4.8.  $\square$

*Proof of Proposition 4.7.* Similarly to the proof of Proposition 2.3, we obtain

$$\|\partial_z \operatorname{Im} m(z)\| \lesssim \|\partial_z m(z)\| \leq \|(\operatorname{Id} - C_m S)^{-1}\| \|m(z)\|^2 \lesssim \rho(z)^{-2} \sim \|\operatorname{Im} m(z)\|^{-2} \tag{4.24}$$

for  $z \in \mathbb{H}_{I, \eta_*}$  from (4.16), (4.22) and (4.17). By the submultiplicativity of  $\|\cdot\|$ ,  $(\operatorname{Im} m(z))^3: \mathbb{H}_{I, \eta_*} \rightarrow (\mathcal{A}, \|\cdot\|)$  is a uniformly Lipschitz-continuous function. Hence,  $\operatorname{Im} m(z)$  is uniformly 1/3-Hölder continuous on  $\mathbb{H}_{I, \eta_*}$  (see e.g. Theorem X.1.1 in [14]) and, thus, has a uniformly 1/3-Hölder continuous extension to  $\overline{\mathbb{H}}_{I, \eta_*}$ . We conclude that the measure  $\nu$  restricted to  $I$  has a density with respect to the Lebesgue measure on  $I$ , i.e., (2.9) holds true for all measurable  $A \subset I$ . Now, (A.3) in Lemma A.1 implies the uniform 1/3-Hölder continuity of  $m$  on  $I_\theta \times i(0, \infty)$ . In particular,  $m$  can be uniquely extended to a uniformly 1/3-Hölder continuous function on  $I_\theta \times i[0, \infty)$  such that (4.19) holds true. By (3.7), we have that  $\|m(z)\| \leq 1$  if  $\operatorname{Im} z \geq 1$ . Thus,  $\|m(z)\| \lesssim 1 + \theta^{-4/3}$  for all  $z \in I_\theta \times i[0, \infty)$  follows directly from (4.19).

To prove the analyticity of  $m$ , we refer to the proof of the analyticity of  $\rho$  in Proposition 2.3. The bound (4.20) can be read off from (4.24). This completes the proof of the proposition.  $\square$

*Proof of Proposition 2.4.* By (2.7), there are  $C_0 > 0$  and  $\eta_* \in (0, 1]$  such that  $\|m(\tau + i\eta)\| \leq C_0$  for all  $\tau \in I$  and  $\eta \in (0, \eta_*]$ . Hence, by Lemma 4.8 (ii), the flatness of  $S$  implies Assumptions 4.5 on  $I \cap [-C, C]$  for  $C := 3(1 + \|a\| + \|S\|^{1/2})$ , i.e.,  $C \sim 1$ . Therefore, Proposition 4.7 yields Proposition 2.4 on  $I \cap [-C, C]$ . Owing to (3.7) and  $\text{supp } v = \text{supp } \rho$ , we have  $\text{dist}(\tau, \text{supp } v) \geq 1$  for  $\tau \in I$  satisfying  $\tau \notin [-C + 1, C - 1]$ . Hence, for these  $\tau$ , the Hölder-continuity follows immediately from (A.4) in Lemma A.1. Moreover, we have  $\|m(z)\| \leq 1$  if  $z \in \mathbb{H}$  with  $\text{Re } z \in I \setminus [-C + 1, C - 1]$  by (3.7) as  $\text{dist}(z, \text{supp } v) \geq 1$ . By (2.5a), we have  $\text{Im } m(\tau) = 0$  for  $\tau \in I$  satisfying  $\tau \notin [-C, C]$ . Therefore, the statement about the analyticity is trivial outside of  $[-C, C]$ . This completes the proof of Proposition 2.4.  $\square$

5 SPECTRAL PROPERTIES OF THE STABILITY OPERATOR FOR SMALL SELF-CONSISTENT DENSITY OF STATES

In this section, we study the stability operator  $B = B(z) := \text{Id} - C_{m(z)}S$ , when  $\rho = \rho(z)$  is small and Assumptions 4.5 hold true. Note that we do not require  $S$  to be flat, i.e., to satisfy (3.10). We will first show that  $B$  has a single eigenvalue close to zero while all other eigenvalues of  $B$  are well separated from zero. In a second step, we then obtain precise expansions for this eigenvalue as well as the corresponding left and right eigenvectors expressing this unstable direction in terms of basic quantities derived from  $m$ . The obtained spectral information will be employed in Section 6 to deduce a cubic equation for the component of  $m$  in the unstable direction as a function of the spectral parameter  $z$ . In Subsections 5.3 and 5.4, we will use the precise knowledge about the small isolated eigenvalue of  $B$  and its associated eigenvectors to extend the quantities derived from  $m$  to the real axis and establish a bound on the derivatives of  $m$ .

5.1 SPECTRUM OF THE STABILITY OPERATOR CLOSE TO ZERO

This subsection is devoted to the proof of Lemma 5.1 below, a basic result about the structure of the spectrum of  $B$  close to zero. The key ingredient in the proof is viewing  $B = \text{Id} - C_m S$  as a perturbation of the operator  $B_0$ , which we introduce now. We define

$$\begin{aligned} s &:= \text{sign } \text{Re } u, & B_0 &:= C_{q^*,q}(\text{Id} - C_s F)C_{q^*,q}^{-1}, \\ E &:= (C_{q^*sq} - C_m)S = C_{q^*,q}(C_s - C_u)FC_{q^*,q}^{-1}, \end{aligned} \tag{5.1}$$

with  $u$  and  $q$  defined in (3.1) and  $F$  defined in (3.4). Note  $B_0 = \text{Id} - C_{q^*sq}S$ , i.e., in the definition of  $B$ ,  $u$  in  $m = q^*uq$  is replaced by  $s$ . Thus, we have  $B = B_0 + E$ . Apart from the base change  $C_{q^*,q}$ , the basic constituents of  $B_0$  are the two self-adjoint operators  $C_s$  and  $F$ . Their self-adjointness simplifies the spectral analysis which is the main advantage of working with  $B_0$  instead of  $B$ . Under Assumptions 4.5, (4.21) holds true which we will often use in the following. Since  $1 - |\text{Re } u| = 1 - \sqrt{1 - (\text{Im } u)^2} \leq (\text{Im } u)^2 \lesssim \rho^2$ , we also obtain

$$\text{Re } u = s + \mathcal{O}(\rho^2), \quad \text{Im } u = \mathcal{O}(\rho), \quad \text{Re } m = q^*sq + \mathcal{O}(\rho^2) \tag{5.2}$$

and with  $C_s - C_u = \mathcal{O}(\|s - u\|) = \mathcal{O}(\rho)$  we get

$$E = \mathcal{O}(\rho). \tag{5.3}$$

Here, we use the notation  $R = T + \mathcal{O}(\alpha)$  for operators  $T$  and  $R$  on  $\mathcal{A}$  and  $\alpha > 0$  if  $\|R - T\| \lesssim \alpha$ . We introduce

$$f_u := \rho^{-1} \operatorname{Im} u. \tag{5.4}$$

By the functional calculus for the normal operator  $u$ ,  $\operatorname{Re} u$ ,  $s$  and  $f_u$  commute. Hence,  $C_s[f_u] = f_u$ . From the imaginary part of (3.3) and (4.21), we conclude that

$$(\operatorname{Id} - F)[f_u] = \rho^{-1} \operatorname{Im} z q q^* = \mathcal{O}(\rho^{-1} \operatorname{Im} z). \tag{5.5}$$

The following technical lemma provides control on the resolvent of the stability operator  $B$  and its relatives. In particular, it implies that these operators have one eigenvalue close to zero while all other eigenvalues are separated away from zero by a controlled amount. It has been stated for the finite dimensional situation  $\mathcal{A} = \mathbb{C}^{N \times N}$  in [8, Corollary 4.8]. For the reader’s convenience we present its proof following the same line of reasoning as in [8]. For  $z \in \mathbb{C}$  and  $\varepsilon > 0$ , we denote by  $D_\varepsilon(z) := \{w \in \mathbb{C}: |z - w| < \varepsilon\}$  the disk in  $\mathbb{C}$  of radius  $\varepsilon$  around  $z$ .

LEMMA 5.1 (Spectral properties of stability operator). *Let  $T \in \{\operatorname{Id} - F, \operatorname{Id} - C_s F, B_0, B, \operatorname{Id} - C_{m^*, m} S\}$ . If Assumptions 4.5 are satisfied on an interval  $I \subset \mathbb{R}$  for some  $\eta_* \in (0, 1]$ , then there are  $\rho_* \sim 1$  and  $\varepsilon \sim 1$  such that*

$$\|(T - \omega \operatorname{Id})^{-1}\|_2 + \|(T - \omega \operatorname{Id})^{-1}\| + \|(T^* - \omega \operatorname{Id})^{-1}\| \lesssim 1 \tag{5.6}$$

uniformly for all  $z \in \mathbb{H}_{I, \eta_*}$  satisfying  $\rho(z) + \rho(z)^{-1} \operatorname{Im} z \leq \rho_*$  and for all  $\omega \in \mathbb{C}$  with  $\omega \notin D_\varepsilon(0) \cup D_{1-2\varepsilon}(1)$ . Furthermore, there is a single simple (algebraic multiplicity 1) eigenvalue  $\lambda$  in the disk around 0, i.e.,

$$\begin{aligned} \operatorname{Spec}(T) \cap D_\varepsilon(0) &= \{\lambda\} \quad \text{and} \quad \operatorname{rank} P_T = 1, \\ \text{where } P_T &:= -\frac{1}{2\pi i} \int_{\partial D_\varepsilon(0)} (T - \omega \operatorname{Id})^{-1} d\omega. \end{aligned} \tag{5.7}$$

Before proving Lemma 5.1, we explain the origin of its assumption  $\rho(z) + \rho(z)^{-1} \operatorname{Im} z \leq \rho_*$ . The condition  $\rho(z) \leq \rho_*$  ensures the effectivity of perturbation theory as  $B - B_0 = E = \mathcal{O}(\rho)$  by (5.3). The self-adjoint operator  $f_u$  is an approximate eigenvector of  $C_s F$  due to  $C_s[f_u] = f_u$  and (5.5) if  $\rho(z)^{-1} \operatorname{Im} z \leq \rho_*$ .

If Assumptions 4.5 are satisfied on  $I$  for some  $\eta_* \in (0, 1]$  then we have

$$f_u = \rho^{-1} \operatorname{Im} u \sim 1. \tag{5.8}$$

uniformly for  $z \in \mathbb{H}_{I, \eta_*}$  due to (4.21). This fact will often be used in the following without mentioning it.

*Proof.* First, we notice that for each choice of the operator  $T$  from the lemma, the bound  $\|\text{Id} - T\|_{2 \rightarrow \|\cdot\|} \lesssim 1$  holds because of  $\|S\|_{2 \rightarrow \|\cdot\|} \lesssim 1$ , (4.16) and (4.21). Therefore invertibility of  $T - \omega \text{Id}$  as an operator on  $L^2$  implies invertibility as an operator on  $\mathcal{A}$ , as long as  $\omega$  stays away from 1, due to Lemma B.2 (ii). It suffices thus to show the bound on the  $\|\cdot\|_2$ -norm from (5.6) and (5.7). For  $T = \text{Id} - F$  both assertions hold by Lemma 4.3. In particular, we find

$$f = \|f_u\|_2^{-1} f_u + \mathcal{O}(\rho^{-1} \text{Im } z), \tag{5.9}$$

where  $f$  is the single top eigenvector of  $F$ ,  $Ff = \|F\|_2 f$  (see Lemma 4.3). The proof of (5.9) follows from (5.5) and  $\|F\|_2 = 1 + \mathcal{O}(\rho^{-1} \text{Im } z)$  (cf. (4.5)) by straightforward perturbation theory of the simple isolated eigenvalue  $\|F\|_2$ . We will now prove (5.7) and the  $\|\cdot\|_2$ -norm bound

$$\|(T - \omega \text{Id})^{-1}\|_2 \lesssim 1, \quad \omega \notin D_\varepsilon(0) \cup D_{1-2\varepsilon}(1) \tag{5.10}$$

for the choices  $T = \text{Id} - C_s F, B_0, B, \text{Id} - C_{m^*, m} S$  in this order. We start with  $T = \text{Id} - C_s F$ . We introduce the interpolation  $T_t := \text{Id} - V_t F$  between  $T_0 = \text{Id} - F$  and  $T_1 = \text{Id} - C_s F$  by setting

$$V_t := (1 - t)\text{Id} + tC_s, \quad t \in [0, 1].$$

Once we have established (5.10) with  $T = T_t$  for all  $t \in [0, 1]$ , the assertion about the single isolated eigenvalue (5.7) also follows for  $T = T_t$ . Indeed, the rank of the spectral projection  $P_{T_t}$  is a continuous function of  $t$  and thus  $\text{rank } P_{T_t} = \text{rank } P_{T_0} = 1$  by what we have already shown.

In order to show (5.10) we consider two regimes. On the one hand, for  $|\omega| \geq 3$  we simply use  $\|F\|_2 \leq 1$  and  $\|V_t\|_2 \leq 1$ . On the other hand, for  $|\omega| \leq 3$  we estimate the norm of  $((1 - \omega)\text{Id} - V_t F)[x]$  from below for any  $x \in L^2$ . For this purpose we decompose  $x = \alpha f + y$  according to the top eigenvector  $f$  of  $F$ , with  $y \perp f$  and  $\alpha \in \mathbb{C}$ . Then we find

$$\begin{aligned} \|((1 - \omega)\text{Id} - V_t F)[x]\|_2^2 &= |\alpha|^2 |\omega|^2 + \|((1 - \omega)\text{Id} - V_t F)[y]\|_2^2 \\ &\quad + \mathcal{O}(\rho^{-1} \text{Im } z \|x\|_2^2) \\ &\geq |\alpha|^2 \varepsilon^2 + (\vartheta - 2\varepsilon)^2 (\|x\|_2^2 - |\alpha|^2) \\ &\quad + \mathcal{O}(\rho^{-1} \text{Im } z \|x\|_2^2), \end{aligned} \tag{5.11}$$

where  $\vartheta \sim 1$  is the spectral gap of  $F$  from (4.7). In the equality of (5.11) we used that  $V_t F[f] = f + \mathcal{O}(\rho^{-1} \text{Im } z)$  and  $FV_t[f] = f + \mathcal{O}(\rho^{-1} \text{Im } z)$  due to (5.9),  $V_t[f_u] = f_u$  and  $\|F\|_2 = 1 + \mathcal{O}(\rho^{-1} \text{Im } z)$ , as well as the orthogonality of  $y$  and  $f$ . For the inequality in (5.11) we estimated  $|\omega| \geq \varepsilon$  and used

$$\|((1 - \omega)\text{Id} - V_t F)[y]\|_2^2 \geq (1 - |\omega| - \|F\|_2(1 - \vartheta))^2 \|y\|_2^2 \geq (\vartheta - 2\varepsilon)^2 (\|x\|_2^2 - |\alpha|^2).$$

From (5.11) we now conclude  $\|((1 - \omega)\text{Id} - V_t F)[x]\|_2^2 \gtrsim \|x\|_2^2$  by choosing  $\varepsilon$  and  $\rho_*$  small enough.

Since we have established the claim of the lemma for  $T = \text{Id} - C_s F$  it also follows for  $T = B_0$  because of the definition of  $B_0$  in (5.1) and (4.21). Thus  $B_0$  has a simple isolated eigenvalue in  $D_\varepsilon(0)$  and we can use analytic perturbation theory to establish the lemma for the choices  $T = B, \text{Id} - C_{m^*,m} S$ . Note that in either case  $T = B_0 + \mathcal{O}(\rho)$  due to  $\|s - u\| \lesssim \rho$  (cf. (5.2)).  $\square$

5.2 SMALL EIGENVALUE OF STABILITY OPERATOR AND CORRESPONDING EIGENVECTORS

Lemma 5.1 asserts that the stability operator  $B$  has a simple small eigenvalue well separated from the rest of its spectrum. In the main result of this section, Corollary 5.2, we give a precise expansion of this eigenvalue and the corresponding left and right eigenvectors. First we introduce some notation.

If  $z \in \mathbb{H}_{I,\eta^*}$  satisfies  $\rho(z) + \rho(z)^{-1} \text{Im} z \leq \rho_*$  for  $\rho_* \sim 1$  from Lemma 5.1 then we denote by  $P_{s,F}$  the spectral projection corresponding to the isolated eigenvalue of  $\text{Id} - C_s F$ , i.e.,  $P_{s,F}$  equals  $P_T$  in (5.7) with  $T = \text{Id} - C_s F$ . We also set  $Q_{s,F} := \text{Id} - P_{s,F}$ . Moreover, for such  $z$ , we define  $\psi$  and  $\sigma$  by

$$\psi(z) := \langle s f_u^2, (\text{Id} + F)(\text{Id} - C_s F)^{-1} Q_{s,F} [s f_u^2] \rangle, \quad \sigma(z) := \langle s f_u^3 \rangle. \tag{5.12}$$

In the following corollary of Lemma 5.1, we consider  $B$  as a perturbation of  $B_0$  and correspondingly expand its isolated eigenvalue and eigenvectors. In [8, Corollary 4.8] a simpler expansion has been performed in the vicinity of an edge point, i.e., where  $\text{Im} m$  follows the square root behaviour from Theorem 2.5. However, here we have to expand to higher order because we cover the neighbourhood of any cubic root cusp from Theorem 2.5 as well.

**COROLLARY 5.2.** *Let  $z \in \mathbb{H}_{I,\eta^*}$  satisfy  $\rho(z) + \rho(z)^{-1} \text{Im} z \leq \rho_*$  for  $\rho_* \sim 1$  from Lemma 5.1. Let  $\beta_0$  and  $\beta$  be the isolated eigenvalues in  $D_\varepsilon(0)$  of  $B_0$  and  $B$ , respectively (cf. Lemma 5.1). We denote by  $P_0$  and  $P$  the spectral projections corresponding to  $\beta_0$  and  $\beta$ , respectively. Then with  $Q_0 := \text{Id} - P_0$  and  $Q := \text{Id} - P$  we have*

$$\|B^{-1}Q\| + \|B^{-1}Q\|_2 + \|B_0^{-1}Q_0\| \lesssim 1. \tag{5.13}$$

Furthermore, we set  $b_0 := P_0 C_{q^*,q} [f_u]$  and  $l_0 := P_0^* C_{q,q^*}^{-1} [f_u]$ . Then  $b_0$  and  $l_0$  are right and left eigenvectors of  $B_0$  associated to  $\beta_0$  and we have

$$b_0 = C_{q^*,q} [f_u] + \mathcal{O}(\rho^{-1} \text{Im} z), \quad l_0 = C_{q,q^*}^{-1} [f_u] + \mathcal{O}(\rho^{-1} \text{Im} z), \tag{5.14a}$$

$$\beta_0 = \frac{\text{Im} z}{\rho} \frac{\pi}{\langle f_u^2 \rangle} + \mathcal{O}(\rho^{-2} (\text{Im} z)^2) = \mathcal{O}(\rho^{-1} \text{Im} z). \tag{5.14b}$$

The definitions  $b := P[b_0]$  and  $l := P^*[l_0]$  yield right and left eigenvectors of  $B$

associated to  $\beta$  which satisfy

$$b = b_0 + 2i\rho C_{q^*,q}(\text{Id} - C_s F)^{-1} Q_{s,F}[s f_u^2] + \mathcal{O}(\rho^2 + \text{Im } z), \tag{5.15a}$$

$$l = l_0 - 2i\rho C_{q,q^*}^{-1}(\text{Id} - F C_s)^{-1} Q_{s,F}^* F[s f_u^2] + \mathcal{O}(\rho^2 + \text{Im } z), \tag{5.15b}$$

$$\beta \langle l, b \rangle = \pi \rho^{-1} \text{Im } z - 2i\rho\sigma + 2\rho^2 \left( \psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^3 + \text{Im } z + \rho^{-2}(\text{Im } z)^2). \tag{5.15c}$$

Moreover, we have

$$\|b\| \lesssim 1, \quad \|l\| \lesssim 1. \tag{5.16}$$

For later use, we record some identities here. From (5.9) in the proof of Lemma 5.1 with  $C_s[f_u] = f_u$ , we obtain the first relation in

$$P_{s,F} = \frac{\langle f_u, \cdot \rangle}{\langle f_u^2 \rangle} f_u + \mathcal{O}(\rho^{-1} \text{Im } z), \quad P_{s,F}^* = P_{s,F} + \mathcal{O}(\rho^{-1} \text{Im } z), \tag{5.17}$$

$$Q_{s,F}^* = Q_{s,F} + \mathcal{O}(\rho^{-1} \text{Im } z).$$

This first relation together with  $f_u = f_u^*$  implies the second and third one. Moreover, the definitions of  $B_0$  and  $Q_0$  yield

$$B_0^{-1} Q_0 = C_{q^*,q}(\text{Id} - C_s F)^{-1} Q_{s,F} C_{q^*,q}^{-1}. \tag{5.18}$$

By a direct computation starting from the definition of  $f_u$  in (5.4) and the balanced polar decomposition,  $m = q^* u q$ , we obtain

$$\langle f_u q q^* \rangle = \rho^{-1} \langle \text{Im } m \rangle = \pi. \tag{5.19}$$

*Proof.* The bounds in (5.13) follow directly from the analytic functional calculus and Lemma 5.1. The expressions (5.14a) for the right and left eigenvectors,  $b_0$  and  $l_0$ , corresponding to the simple isolated eigenvalue  $\beta_0$ , follow by simple perturbation theory from

$$B_0^* C_{q,q^*}^{-1}[f_u] = \rho^{-1}(\text{Im } z) \mathbf{1}, \quad B_0 C_{q^*,q}[f_u] = \mathcal{O}(\rho^{-1} \text{Im } z), \tag{5.20}$$

which in turn is a consequence of (5.5) and  $C_s[f_u] = f_u$ . For (5.14b) we take the scalar product with  $b_0$  on both sides of the first equation in (5.20). Then we use (5.14a) and (5.19).

Now we show (5.15a) and (5.15b). By analytic perturbation theory of  $B$  around  $B_0$  we find  $b = b_0 + b_1 + \mathcal{O}(\rho^2)$  and  $l = l_0 + l_1 + \mathcal{O}(\rho^2)$  with  $b_1 := -(B_0 - \beta_0 \text{Id})^{-1} Q_0 E[b_0]$  and  $l_1 := -(B_0^* - \bar{\beta}_0 \text{Id})^{-1} Q_0^* E^*[l_0]$  (cf. Lemma C.1 with  $E$  satisfying (5.3)). Here the invertibility of  $B_0 - \beta_0 \text{Id}$  on the range of  $Q_0$  is seen from the second part of Lemma 5.1 with  $T = B_0$ . In fact,

$$(B_0 - \beta_0 \text{Id})^{-1} Q_0 = B_0^{-1} Q_0 + \mathcal{O}(\beta_0). \tag{5.21}$$



Furthermore, we use (5.14a) and obtain the first equalities below:

$$\begin{aligned} E[b_0] &= C_{q^*,q}(C_s - C_u)F[f_u] + \mathcal{O}(\text{Im } z) \\ &= -2i\rho C_{q^*,q}[sf_u^2] + 2\rho^2 C_{q^*,q}[f_u^3] + \mathcal{O}(\rho^3 + \text{Im } z), \end{aligned} \tag{5.22a}$$

$$\begin{aligned} E^*[l_0] &= C_{q,q^*}^{-1}F(C_s - C_u^*)[f_u] + \mathcal{O}(\text{Im } z) \\ &= 2i\rho C_{q,q^*}^{-1}F[sf_u^2] + 2\rho^2 C_{q,q^*}^{-1}F[f_u^3] + \mathcal{O}(\rho^3 + \text{Im } z). \end{aligned} \tag{5.22b}$$

In the second equality of (5.22a), we applied  $(C_s - C_u)[f_u] = 2(\text{Im } u - i\text{Re } u)(\text{Im } u)f_u = -2i\rho sf_u^2 + 2\rho^2 f_u^3 + \mathcal{O}(\rho^3)$ ,  $\|C_s - C_u\| = \mathcal{O}(\rho)$  (cf. (5.2) and (5.5)). For the second equality in (5.22b), we applied  $(C_s - C_u^*)[f_u] = 2i\rho sf_u^2 + 2\rho^2 f_u^3 + \mathcal{O}(\rho^3)$ .

For the proof of (5.15c), we start from (C.4), use  $E = \mathcal{O}(\rho)$  and obtain

$$\beta\langle l, b \rangle = \beta_0\langle l_0, b_0 \rangle + \langle l_0, E[b_0] \rangle - \langle l_0, EB_0(B_0 - \beta_0\text{Id})^{-2}Q_0E[b_0] \rangle + \mathcal{O}(\rho^3). \tag{5.23}$$

Each of the terms on the right-hand side is computed individually. For the first term, we use  $\langle l_0, b_0 \rangle = \langle f_u^2 \rangle + \mathcal{O}(\rho^{-1}\text{Im } z)$  due to (5.14a) and thus obtain from (5.14b) that

$$\beta_0\langle l_0, b_0 \rangle = \pi\rho^{-1}\text{Im } z + \mathcal{O}(\rho^{-2}(\text{Im } z)^2).$$

Using (5.14a) and (5.22) yields for the second term

$$\begin{aligned} \langle l_0, E[b_0] \rangle &= -2i\rho\langle sf_u^3 \rangle + 2\rho^2\langle f_u^4 \rangle + \mathcal{O}(\rho^3 + \text{Im } z) \\ &= -2i\rho\sigma + 2\rho^2\left(\frac{\sigma^2}{\langle f_u^2 \rangle} + \langle sf_u^2, Q_{s,F}[sf_u^2] \rangle\right) + \mathcal{O}(\rho^3 + \text{Im } z), \end{aligned}$$

where we used  $\text{Id} = P_{s,F} + Q_{s,F}$  and  $\langle sf_u^2, P_{s,F}[sf_u^2] \rangle = \sigma^2/\langle f_u^2 \rangle + \mathcal{O}(\rho^{-1}\text{Im } z)$  by (5.17) in the last step.

For the third term, we use (5.14b) and  $E = \mathcal{O}(\rho)$  which yields

$$\begin{aligned} \langle l_0, EB_0(B_0 - \beta_0\text{Id})^{-2}Q_0E[b_0] \rangle &= \langle E^*[l_0], (B_0 - \beta_0\text{Id})^{-1}Q_0E[b_0] \rangle \\ &\quad + \mathcal{O}(\beta_0\|E\|^2) \\ &= \langle E^*[l_0], B_0^{-1}Q_0E[b_0] \rangle + \mathcal{O}(\rho\text{Im } z) \\ &= -4\rho^2\langle sf_u^2, F(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] \rangle \\ &\quad + \mathcal{O}(\rho\text{Im } z + (\text{Im } z)^2 + \rho^3). \end{aligned}$$

Here, we used (5.21) in the second step and (5.22) as well as (5.18) in the last step. By collecting the results for the three terms in (5.23), we obtain

$$\begin{aligned} \beta\langle l, b \rangle &= \frac{\pi\text{Im } z}{\rho} - 2i\rho\sigma + 2\rho^2\left(\langle sf_u^2, (\text{Id} - F + 2F)(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] \rangle \right. \\ &\quad \left. + \frac{\sigma^2}{\langle f_u^2 \rangle}\right) + \mathcal{O}(\rho^3 + \text{Im } z + \rho^{-2}(\text{Im } z)^2), \end{aligned}$$

where we used  $\langle sf_u^2, Q_{s,F}[sf_u^2] \rangle = \langle sf_u^2, (\text{Id} - F)(\text{Id} - C_s F)^{-1} Q_{s,F}[sf_u^2] \rangle$ , which follows directly from  $C_s = C_s^*$  and  $C_s[sf_u^2] = sf_u^2$ . This yields (5.15c). The bounds in (5.16) are directly implied by (5.15a) and (5.15b), respectively. This finishes the proof of the corollary.  $\square$

The following corollary has appeared prior to this work in [8, Proposition 4.4]. We include its short proof for the reader’s convenience.

**COROLLARY 5.3** (Improved bound on  $B^{-1}$ ). *Let Assumptions 4.5 hold true on an interval  $I \subset \mathbb{R}$  for some  $\eta_* \in (0, 1]$ . Then, uniformly for all  $z \in \mathbb{H}_{I, \eta_*}$ , we have*

$$\|B^{-1}(z)\|_2 + \|B^{-1}(z)\| \lesssim \frac{1}{\rho(z)(\rho(z) + |\sigma(z)|) + \rho(z)^{-1}\text{Im} z} \tag{5.24}$$

with  $\sigma$  from (5.12).

*Proof.* If  $\rho \geq \rho_*$  for some  $\rho_* \sim 1$  then (5.24) have been shown in (4.22) as  $|\sigma| \lesssim 1$ . Therefore, we prove (5.24) for  $\rho \leq \rho_*$  and a sufficiently small  $\rho_* \sim 1$ . By  $\|S\|_{2 \rightarrow \|\cdot\|} \lesssim 1$  and Lemma B.2 (ii), it suffices to show the bound for  $\|\cdot\|_2$ . We follow the proof of (4.22) until (4.23). Hence, for the improved bound, we have to show that

$$|1 - \|F\|_2 \langle fC_u^*[f] \rangle| \gtrsim \rho(\rho + |\sigma|) + \rho^{-1}\text{Im} z. \tag{5.25}$$

We have  $|1 - \|F\|_2 \langle fC_u^*[f] \rangle| \gtrsim \max\{1 - \|F\|_2, |1 - \langle fC_u^*[f] \rangle|\} \gtrsim \rho^{-1}\text{Im} z + |1 - \langle fC_u^*[f] \rangle|$  by (4.5). We continue

$$\begin{aligned} |1 - \langle fC_u^*[f] \rangle| &= |1 - \langle fu^*fu^* \rangle| \gtrsim \langle f\text{Im} uf\text{Im} u \rangle + |\langle f\text{Im} uf\text{Re} u \rangle| \\ &\gtrsim \rho^2 + \rho|\sigma| + \mathcal{O}(\rho^3 + \text{Im} z). \end{aligned}$$

Here, we used  $1 \geq \langle f\text{Re} uf\text{Re} u \rangle$  due to  $\|f\|_2 = 1$ , (4.21) as well as  $\langle f\text{Im} uf\text{Re} u \rangle = \rho\|f_u\|_2^{-2} \langle f_u^3 s \rangle + \mathcal{O}(\rho^3 + \text{Im} z)$  by (5.9) and (5.2). By possibly shrinking  $\rho_* \sim 1$ , we thus obtain (5.25). This completes the proof of (5.24).  $\square$

### 5.3 BEHAVIOUR OF $\sigma$ AND $\psi$ CLOSE TO THE REAL AXIS

This subsection is devoted to several results about the behaviour of  $\sigma(z)$  and  $\psi(z)$  close to the real axis. They are summarised in Lemma 5.5 below in the form as they will be applied in the next section. We now prepare these results by extending  $q, u, f_u$  and  $s$  to the real axis.

**LEMMA 5.4** (Extensions of  $q, u, f_u$  and  $s$ ). *Let  $I \subset \mathbb{R}$  be an interval,  $\theta \in (0, 1]$  and Assumptions 4.5 hold true on  $I$  for some  $\eta_* \in (0, 1]$ . We set  $I_\theta := \{\tau \in I : \text{dist}(\tau, \partial I) \geq \theta\}$ . Then we have*

- (i) *The functions  $q, u$  and  $f_u$  have unique uniformly 1/3-Hölder continuous extensions to  $\overline{\mathbb{H}}_{I_\theta, \eta_*}$ .*

(ii) The function  $z \mapsto \rho(z)^{-1}\text{Im } z$  has a unique uniformly 1/3-Hölder continuous extension to  $\overline{\mathbb{H}}_{I_\theta, \eta_*}$ . In particular, we have

$$\lim_{z \rightarrow \tau_0} \rho(z)^{-1}\text{Im } z = 0 \tag{5.26}$$

for all  $\tau_0 \in \text{supp } \rho \cap I_\theta$ . Moreover, for  $z \in \overline{\mathbb{H}}_{I_\theta, \eta_*}$ , we have

$$\text{dist}(z, \text{supp } \rho) \gtrsim 1 \iff \rho(z)^{-1}\text{Im } z \gtrsim 1.$$

(iii) There is a threshold  $\rho_* \sim 1$  such that  $s = \text{sign}(\text{Re } u)$  has a unique uniformly 1/3-Hölder continuous extension to  $\{w \in \overline{\mathbb{H}}_{I_\theta, \eta_*} : \rho(w) \leq \rho_*\}$ .

*Proof.* For the proof of (i), we will show below that

$$f_m(z) := \rho(z)^{-1}\text{Im } m(z)$$

is uniformly 1/3-Hölder continuous on  $\mathbb{H}_{I_\theta, \eta_*}$ . Indeed, this suffices to obtain the Hölder-continuity of  $q$  and  $u$  since their definitions in (3.1) can be rewritten as

$$\begin{aligned} q &= |h^{-1/2}gh^{-1/2} + i\mathbb{1}|^{1/2}h^{1/2} = (\rho(z)^2\mathbb{1} + f_m^{-1/2}gf_m^{-1}gf_m^{-1/2})^{1/4}f_m^{1/2}, \\ u &= \frac{\rho(z)w}{|\rho(z)w|} = \frac{i\rho(z)\mathbb{1} + f_m^{-1/2}gf_m^{-1/2}}{|i\rho(z)\mathbb{1} + f_m^{-1/2}gf_m^{-1/2}|}, \end{aligned} \tag{5.27}$$

where  $g = \text{Re } m$ ,  $h = \text{Im } m$ ,  $w$  is defined in (3.1) and  $z \in \mathbb{H}$  is arbitrary. Since  $|\rho(z)w| \sim 1$  and  $f_m \sim 1$  on  $\mathbb{H}_{I_\theta, \eta_*}$  by (4.21) as well as (4.17) and  $m$ , hence  $\rho$  and  $\text{Re } m$  are Hölder-continuous on  $I_\theta \times i[0, \infty)$  (Proposition 4.7), it thus suffices to show that  $f_m$  is uniformly Hölder-continuous to conclude from (5.27) that  $q$  and  $u$  are Hölder-continuous. As  $f_u = \rho^{-1}\text{Im } u = (q^*)^{-1}f_mq^{-1}$ , the Hölder-continuity of  $f_m$ , the Hölder-continuity of  $q$  and the upper and lower bounds on  $q$  from (4.21) imply that  $f_u$  can be extended to a 1/3-Hölder continuous function on  $\overline{\mathbb{H}}_{I_\theta, \eta_*}$ .

Therefore, we now complete the proof of (i) by showing the 1/3-Hölder continuity of  $f_m$ . To that end, we distinguish three subsets of  $\mathbb{H}_{I_\theta, \eta_*}$ .

*Case 1:* On the set  $\{z \in \mathbb{H}_{I_\theta, \eta_*} : \rho(z) \geq \rho_*\}$  for any  $\rho_* \sim 1$ , the uniform 1/3-Hölder continuity of  $f_m$  follows from  $\rho(z) \gtrsim 1$  and the 1/3-Hölder continuity of  $m$  from Proposition 4.7.

*Case 2:* In order to analyze  $f_m$  on the set  $\{z \in \mathbb{H}_{I_\theta, \eta_*} : \rho(z) \leq \rho_*\}$  for some  $\rho_* \sim 1$  to be chosen later, we take the imaginary part of the Dyson equation, (2.3), at  $z \in \mathbb{H}$  and obtain

$$B_*[\text{Im } m] = (\text{Im } z)m^*m, \quad B_* := \text{Id} - C_{m^*, m}S, \tag{5.28}$$

where  $m = m(z)$ . From  $m = q^*uq$ , we obtain the representation

$$\text{Id} - C_{m^*, m}S = C_{q^*, q}(\text{Id} - C_{u^*, u}F)C_{q^*, q}^{-1}.$$

Hence, (4.5), Lemma 4.8 (ii) and Lemma B.2 (ii) yield the invertibility of  $B_*$  for each  $z \in \mathbb{H}_{I, \eta_*}$  as well as

$$\|B_*^{-1}(z)\|_2 + \|B_*^{-1}(z)\| \lesssim \frac{1}{1 - \|F\|_2} \lesssim \frac{\rho(z)}{\operatorname{Im} z} \tag{5.29}$$

for all  $z \in \mathbb{H}_{I, \eta_*}$  (compare the proof of (4.22)). Owing to the invertibility of  $B_*$ , we conclude from (5.28) that

$$f_m(z) = \pi \frac{\operatorname{Im} m(z)}{\langle \operatorname{Im} m(z) \rangle} = \pi \frac{B_*^{-1}[m^*m]}{\langle B_*^{-1}[m^*m] \rangle} \tag{5.30}$$

for all  $z \in \mathbb{H}_{I_\theta, \eta_*}$ .

On the set  $\{z \in \mathbb{H}_{I_\theta, \eta_*} : \rho(z)^{-1} \operatorname{Im} z \geq \rho_*\}$  for any  $\rho_* \sim 1$ ,  $B_*^{-1}[m^*m]$  is uniformly 1/3-Hölder continuous due to (5.29) and the 1/3-Hölder continuity of  $m$ . Moreover, from (4.5) and  $\operatorname{Im} u \sim \rho \mathbf{1}$ , we see that  $1 - \|F\|_2 \sim 1$  if  $\rho(z)^{-1} \operatorname{Im} z \gtrsim 1$ . Hence, by Lemma B.3 in Appendix B below,  $(\operatorname{Id} - C_{u^*, u}F)^{-1}$  is positivity-preserving and satisfies

$$(\operatorname{Id} - C_{u^*, u}F)^{-1}[xx^*] \geq xx^* \tag{5.31}$$

for any  $x \in \mathcal{A}$ . We conclude that  $B_*^{-1} = C_{q^*, q}(\operatorname{Id} - C_{u^*, u}F)^{-1}C_{q^*, q}^{-1}$  is positivity-preserving. Together with (4.21), (5.31) implies  $\langle B_*^{-1}[m^*m] \rangle \gtrsim 1$  as  $\|m(z)^{-1}\| \lesssim 1$  by  $\|a\| \lesssim 1$ ,  $\|S\| \lesssim 1$  and (2.3). Thus, (5.30) yields the uniform 1/3-Hölder continuity of  $f_m$  on  $\{z \in \mathbb{H}_{I_\theta, \eta_*} : \rho(z)^{-1} \operatorname{Im} z \geq \rho_*\}$  for any  $\rho_* \sim 1$ .

*Case 3:* We now show that  $f_m$  is Hölder-continuous on  $\{z \in \mathbb{H}_{I_\theta, \eta_*} : \rho(z) + \rho(z)^{-1} \operatorname{Im} z \leq \rho_*\}$  for some sufficiently small  $\rho_* \sim 1$ . In fact, Lemma 5.1 applied to  $T = B_*$  yields the existence of a unique eigenvalue  $\beta_*$  of  $B_*$  of smallest modulus. Inspecting the proof of Corollary 5.2 for  $B$  reveals that this proof only used  $B = B_0 + \mathcal{O}(\rho)$  about  $B$ . Therefore, the same argument works if  $B$  is replaced by  $B_*$  since  $B_* = B_0 + \mathcal{O}(\rho)$  (compare the proof of Lemma 5.1). We thus find a right eigenvector  $b_*$  and a left eigenvector  $l_*$  of  $B_*$  associated to  $\beta_*$ , i.e.,

$$B_*[b_*] = \beta_* b_*, \quad (B_*)^*[l_*] = \overline{\beta_*} l_*$$

which satisfy

$$b_* = b_0 + \mathcal{O}(\rho) = q^* f_u q + \mathcal{O}(\rho + \rho^{-1} \operatorname{Im} z), \tag{5.32a}$$

$$l_* = l_0 + \mathcal{O}(\rho) = q^{-1} f_u (q^*)^{-1} + \mathcal{O}(\rho + \rho^{-1} \operatorname{Im} z), \tag{5.32b}$$

$$\beta_* \langle l_*, b_* \rangle = \pi \rho^{-1} \operatorname{Im} z + \mathcal{O}(\rho + \rho^{-2} (\operatorname{Im} z)^2). \tag{5.32c}$$

Moreover, we have

$$\|B_*^{-1}Q_*\| + \|B_*^{-1}Q_*\|_2 \lesssim 1, \tag{5.33}$$

where  $Q_*$  denotes the spectral projection of  $B_*$  to the complement of the spectral subspace of  $\beta_*$ .

Therefore, as  $\beta_* \neq 0$  (cf. (5.29)) if  $\text{Im } z > 0$ , we obtain

$$\text{Im } m = (\text{Im } z) B_*^{-1}[m^*m] = (\text{Im } z) \left( \beta_*^{-1} \frac{\langle l_*, m^*m \rangle}{\langle l_*, b_* \rangle} b_* + B_*^{-1} Q_*[m^*m] \right).$$

Consequently, as  $\text{Im } m > 0$ , we have

$$\frac{\text{Im } m}{\langle \text{Im } m \rangle} = \frac{\langle l_*, m^*m \rangle b_* + \beta_* \langle l_*, b_* \rangle B_*^{-1} Q_*[m^*m]}{\langle l_*, m^*m \rangle \langle b_* \rangle + \beta_* \langle l_*, b_* \rangle \langle B_*^{-1} Q_*[m^*m] \rangle}, \tag{5.34}$$

which together with (5.30) shows that  $f_m$  is uniformly  $1/3$ -Hölder continuous on  $\{z \in \mathbb{H}_{I_\theta, \eta_*} : \rho(z) + \rho(z)^{-1} \text{Im } z \leq \rho_*\}$ . Here, we used that  $B_*$  and, thus,  $\beta_*$ ,  $l_*$ ,  $b_*$  and  $B_*^{-1} Q_*$  are  $1/3$ -Hölder continuous and the denominator in (5.34) is  $\gtrsim 1$  due to

$$\begin{aligned} \langle l_*, m^*m \rangle &= \langle q^{-1} f_u(q^*)^{-1} q^* u^* q q^* u q \rangle + \mathcal{O}(\rho + \rho^{-1} \text{Im } z) \\ &= \rho^{-1} \text{Im } \langle q^* u u^* q \rangle + \mathcal{O}(\rho + \rho^{-1} \text{Im } z) = \pi + \mathcal{O}(\rho + \rho^{-1} \text{Im } z) \end{aligned}$$

by (5.32a) and (5.32b) as well as  $\langle b_* \rangle = \pi + \mathcal{O}(\rho + \rho^{-1} \text{Im } z)$  by (5.19). Here, we also used (5.32c) and (5.33). This completes the proof of (i).

For the proof of (ii), we multiply (5.28) by  $\rho(z)^{-1} (m^*m)^{-1}$  which yields

$$\rho(z)^{-1} \text{Im } z = (m^*m)^{-1} B_*[f_m].$$

Owing to  $m^*m \geq \|m^{-1}\|^{-2} \gtrsim 1$  as well as the  $1/3$ -Hölder continuity of  $m$ ,  $B_*$  and  $f_m$ , we obtain the same regularity for  $z \mapsto \rho(z)^{-1} \text{Im } z$ . Since  $\lim_{\eta \downarrow 0} \rho(\tau + i\eta)^{-1} \eta = 0$  for  $\tau \in \text{supp } \rho \cap I_\theta$  satisfying  $\rho(\tau) > 0$ , the continuity of  $\rho(z)^{-1} \text{Im } z$  directly implies (5.26). If  $\text{dist}(z, \text{supp } \rho) \gtrsim 1$  then  $\rho(z)^{-1} \text{Im } z \gtrsim 1$  as  $\rho(z) \leq \text{Im } z / \text{dist}(z, \text{supp } \rho)^2$  which can be seen by applying  $\langle \cdot \rangle$  to the second bound in (3.7). Conversely, if  $\text{dist}(z, \text{supp } \rho) \lesssim 1$  then the Hölder-continuity of  $\rho(z)^{-1} \text{Im } z$  and (5.26) imply  $\rho(z)^{-1} \text{Im } z \lesssim 1$ .

We now turn to the proof of (iii). Owing to the first relation in (5.2), there is  $\rho_* \sim 1$  such that  $|\text{Re } u| \geq \frac{1}{2} \mathbf{1}$  if  $z \in \mathbb{H}_{I_\theta, \eta_*}$  satisfies  $\rho(z) \leq \rho_*$ . Therefore, we find a smooth function  $\varphi: \mathbb{R} \rightarrow [-1, 1]$  such that  $\varphi(t) = 1$  for all  $t \in [1/2, \infty)$ ,  $\varphi(t) = -1$  for all  $t \in (-\infty, -1/2]$  and  $s(z) = \text{sign}(\text{Re } u(z)) = \varphi(\text{Re } u(z))$  for all  $z \in \mathbb{H}_{I_\theta, \eta_*}$  satisfying  $\rho(z) \leq \rho_*$ . Since  $\varphi$  is smooth, we conclude that  $\varphi$  is an operator Lipschitz function [5, Theorem 1.6.1], i.e.,  $\|\varphi(x) - \varphi(y)\| \leq C \|x - y\|$  for all self-adjoint  $x, y \in \mathcal{A}$ . Hence, we conclude

$$\|s(z_1) - s(z_2)\| = \|\varphi(\text{Re } u(z_1)) - \varphi(\text{Re } u(z_2))\| \lesssim \|z_1 - z_2\|^{1/3},$$

where we used that  $\varphi$  is operator Lipschitz and  $u$  is  $1/3$ -Hölder continuous in the last step. This completes the proof of Lemma 5.4.  $\square$

LEMMA 5.5 (Properties of  $\psi$  and  $\sigma$ ). *Let  $I \subset \mathbb{R}$  be an interval and  $\theta \in (0, 1]$ . If  $m$  satisfies Assumptions 4.5 on  $I$  for some  $\eta_* \in (0, 1]$  then there is a threshold  $\rho_* \sim 1$  such that, with*

$$\mathbb{H}_{\text{small}} := \{z \in \mathbb{H}_{I_\theta, \eta_*} : \rho(z) + \rho(z)^{-1} \text{Im } z \leq \rho_*\},$$

we have

(i) The functions  $\sigma$  and  $\psi$  defined in (5.12) have unique uniformly  $1/3$ -Hölder continuous extensions to  $\{z \in \overline{\mathbb{H}}_{I_\theta, \eta_*} : \rho(z) \leq \rho_*\}$  and  $\overline{\mathbb{H}}_{\text{small}}$ , respectively.

(ii) Uniformly for all  $z \in \overline{\mathbb{H}}_{\text{small}}$ , we have

$$\psi(z) + \sigma(z)^2 \sim 1. \tag{5.35}$$

*Proof.* For the proof of (i), we choose  $\rho_* \sim 1$  so small that all parts of Lemma 5.4 are applicable. Thus, Lemma 5.4 and  $\sigma = \langle sf_u^3 \rangle$  yield (i) for  $\sigma$ . Similarly, since  $q$  is now defined on  $\overline{\mathbb{H}}_{I_\theta, \eta_*}$ , we can define  $F$  via (3.4) on this set as well. Moreover, owing to the uniform  $1/3$ -Hölder continuity of  $q$  from Lemma 5.4,  $F$  is uniformly  $1/3$ -Hölder continuous on  $\overline{\mathbb{H}}_{I_\theta, \eta_*}$ . Hence, using Lemma 5.1 for  $T = \text{Id} - C_s F$ , the Hölder-continuity of  $s$  and  $f_u$ , the function  $\psi$  has a unique  $1/3$ -Hölder continuous extension to  $\overline{\mathbb{H}}_{\text{small}}$ . This completes the proof of (i) for  $\psi$ .

We now turn to the proof of (ii). In fact, we will show (5.35) only on  $\{w \in \mathbb{H}_{I_\theta, \eta_*} : \rho(w) + \rho(w)^{-1} \text{Im } w \leq \rho_*\}$ , where  $\rho_* \sim 1$  is chosen small enough such that Lemma 5.1 is applicable. By the continuity of  $\sigma$  and  $\psi$ , the bound (5.35) immediately extends to the closure of this set. Instead of (5.35), we will prove that

$$\langle x, (\text{Id} + F)(\text{Id} - C_s F)^{-1} Q_{s,F}[x] \rangle + \langle f_u, x \rangle^2 \sim \|x\|_2^2 \tag{5.36}$$

for all  $x \in \mathcal{A}$  satisfying  $C_s[x] = x$  and  $x = x^*$ . Since these conditions are satisfied by  $x = sf_u^2$ , (5.36) immediately implies (5.35). In fact, the upper bound in (5.36) follows from  $\|(\text{Id} - C_s F)^{-1} Q_{s,F}\|_2 \lesssim 1$  by Lemma 5.1,  $\|F\|_2 \leq 1$  and  $f_u \sim 1$  due to (5.8).

From  $C_s[x] = x$ , we conclude

$$\begin{aligned} \langle x, (\text{Id} + F)(\text{Id} - C_s F)^{-1} Q_{s,F}[x] \rangle &= \langle x, (\text{Id} + C_s F)(\text{Id} - C_s F)^{-1} Q_{s,F}[x] \rangle \\ &= \langle x, ((C_s F - \text{Id}) + 2\text{Id})(\text{Id} - C_s F)^{-1} Q_{s,F}[x] \rangle \\ &= \langle x, (-\text{Id} + 2(\text{Id} - C_s F)^{-1}) Q_{s,F}[x] \rangle. \end{aligned} \tag{5.37}$$

Using (5.17) and  $C_s[f_u] = f_u$ , we see that

$$\begin{aligned} C_s P_{s,F}[x] &= P_{s,F}[x] + \mathcal{O}(\rho^{-1} \text{Im } z), \\ C_s Q_{s,F}[x] &= Q_{s,F}[x] + \mathcal{O}(\rho^{-1} \text{Im } z) \end{aligned} \tag{5.38}$$

for  $x \in \mathcal{A}$  satisfying  $C_s[x] = x$ . When applied to (5.37), the expansion (5.38)

and  $(\text{Id} - FC_s)^{-1} = C_s(\text{Id} - C_sF)^{-1}C_s$  yield

$$\begin{aligned}
 & \langle x, (\text{Id} + F)(\text{Id} - C_sF)^{-1}Q_{s,F}[x] \rangle \\
 &= \langle Q_{s,F}[x], (-\text{Id} + (\text{Id} - C_sF)^{-1} + (\text{Id} - FC_s)^{-1})Q_{s,F}[x] \rangle \\
 &\quad + \mathcal{O}(\|x\|_2^2 \rho^{-1} \text{Im } z) \\
 &= \langle Q_{s,F}[x], (\text{Id} - FC_s)^{-1}(\text{Id} - F^2)(\text{Id} - C_sF)^{-1}Q_{s,F}[x] \rangle \\
 &\quad + \mathcal{O}(\|x\|_2^2 \rho^{-1} \text{Im } z) \\
 &= \langle (\text{Id} - C_sF)^{-1}Q_{s,F}[x], Q_f(\text{Id} - F^2)Q_f(\text{Id} - C_sF)^{-1}Q_{s,F}[x] \rangle \\
 &\quad + \mathcal{O}(\|x\|_2^2 \rho^{-1} \text{Im } z) \\
 &\gtrsim \|Q_f(\text{Id} - C_sF)^{-1}Q_{s,F}[x]\|_2^2 + \mathcal{O}(\|x\|_2^2 \rho^{-1} \text{Im } z) \\
 &\gtrsim \|Q_{s,F}[x]\|_2^2 + \mathcal{O}(\|x\|_2^2 \rho^{-1} \text{Im } z).
 \end{aligned} \tag{5.39}$$

Here, in the first step, we also used the second and third relation in (5.17). In the third step, we then defined the orthogonal projections  $P_f := \langle f, \cdot \rangle f$  and  $Q_f := \text{Id} - P_f$ , where  $Ff = \|F\|_2 f$  (cf. Assumptions 4.5 (ii)), and inserted  $Q_f$  using

$$P_f Q_{s,F} = \mathcal{O}(\rho^{-1} \text{Im } z) \tag{5.40}$$

which follows from (5.9) and (5.17). We also used that  $Q_{s,F}$  commutes with  $(\text{Id} - C_sF)^{-1}$ . The fourth step is a consequence of (4.7) and (4.18). In the last step, we employed  $Q_f Q_{s,F} = Q_{s,F} + \mathcal{O}(\rho^{-1} \text{Im } z)$  by (5.40) and  $\|\text{Id} - C_sF\|_2 \leq 2$ . By (5.17), we have  $\|P_{s,F}[x]\|_2^2 = \langle f_u, x \rangle^2 + \mathcal{O}(\|x\|_2^2 \rho^{-1} \text{Im } z)$  if  $x = x^*$ . Combining this observation with (5.39) proves (5.36) up to terms of order  $\mathcal{O}(\|x\|_2^2 \rho^{-1} \text{Im } z)$ . Hence, possibly shrinking  $\rho_* \sim 1$  and requiring  $\rho(z)^{-1} \text{Im } z \leq \rho_*$  complete the proof of the lemma.  $\square$

REMARK 5.6 (Auxiliary quantities as functions of  $m$ ). Inspecting the proofs of Lemma 5.4 and Lemma 5.5 reveals that  $q, u, f_u$  and  $s$  as well as  $\sigma$  and  $\psi$  are Lipschitz-continuous functions of  $m$ . More precisely, we have the following statements:

- (i) Let  $c_1, c_2, c_3 > 0$  satisfy  $c_1 < c_2$  and  $\mathcal{M}^{(1)} = \mathcal{M}^{(1)}(c_1, c_2, c_3) \subset \mathcal{A}$  be a nonempty subset of  $\mathcal{A}$  satisfying that

$$\begin{aligned}
 & \text{Im } m_1 \in \mathcal{A}_+, \quad c_1 \langle \text{Im } m_1 \rangle \mathbf{1} \leq \text{Im } m_1 \leq c_2 \langle \text{Im } m_1 \rangle \mathbf{1}, \\
 & \left\| \frac{\text{Im } m_1}{\langle \text{Im } m_1 \rangle} - \frac{\text{Im } m_2}{\langle \text{Im } m_2 \rangle} \right\| \leq c_3 \|m_1 - m_2\|
 \end{aligned} \tag{5.41}$$

hold true for all  $m_1, m_2 \in \mathcal{M}^{(1)}$ . Then  $q, u$  and  $f_u$  are uniformly Lipschitz-continuous functions of  $m$  on  $\mathcal{M}^{(1)}$ .

- (ii) For some  $\rho_* > 0$ , let  $\mathcal{M}^{(2)} = \mathcal{M}^{(2)}(c_1, c_2, c_3, \rho_*) \subset \mathcal{A}$  be a subset of  $\mathcal{A}$  satisfying (5.41) for all  $m_1, m_2 \in \mathcal{M}^{(2)}$  and  $\langle \text{Im } m \rangle \leq \pi \rho_*$  for all  $m \in \mathcal{M}^{(2)}$ . Then there is a (small)  $\rho_* \sim 1$  such that  $s$  and  $\sigma$  are uniformly Lipschitz-continuous functions of  $m$  on  $\mathcal{M}^{(2)} \subset \mathcal{A}$ .

(iii) Fix  $c_4 > 0$ . Let  $\mathcal{M}^{(3)} = \mathcal{M}^{(3)}(c_1, c_2, c_3, c_4, \rho_*)$  be a subset of a set  $\mathcal{M}^{(2)}$  from (ii) with  $\rho_* \sim 1$  chosen as in (ii) such that, for any  $m \in \mathcal{M}^{(3)}$ , the operator  $\text{Id} - C_{s(m)}F(m)$  has a unique eigenvalue of smallest modulus and this eigenvalue is simple (recall that  $F = C_{q,q^*}SC_{q^*,q}$  is a function of  $m$  via  $q = q(m)$ ). Let  $Q_m$  denote the spectral projection of  $\text{Id} - C_{s(m)}F(m)$  onto the complement of this eigenvalue. Moreover, we require that

$$\begin{aligned} \|(\text{Id} - C_{s(m_1)}F(m_1))^{-1}Q_{m_1} - (\text{Id} - C_{s(m_2)}F(m_2))^{-1}Q_{m_2}\| \\ \leq c_4\|m_1 - m_2\| \end{aligned} \tag{5.42}$$

holds true for any  $m_1, m_2 \in \mathcal{M}^{(3)}$ . Then  $\psi$  is a uniformly Lipschitz-continuous function of  $m$  on  $\mathcal{M}^{(3)}$ .

We always consider  $\mathcal{M}^{(i)}$ ,  $i = 1, 2, 3$ , with the metric induced by the norm  $\|\cdot\|$  on  $\mathcal{A}$ . The constants in the Lipschitz-continuity estimates as well as  $\rho_*$  given in (ii) only depend on the control parameters  $c_1, c_2, c_3$  and  $c_4$ .

5.4 BOUND ON THE DERIVATIVES OF  $m$  WHEN SELF-CONSISTENT DENSITY OF STATES IS SMALL

The careful analysis of the operator  $B$  and its inverse allows for the precise bounds on the derivatives of  $m$  in the following lemma.

LEMMA 5.7 (Derivatives of  $m$ ). *Let  $I \subset \mathbb{R}$  be an open interval and  $\theta \in (0, 1]$ . If Assumptions 4.5 hold true on  $I$  for some  $\eta_* \in (0, 1]$  then there is  $C \sim 1$  such that*

$$\|\partial_z^k m(\tau)\| \lesssim \frac{C^k k!}{\rho(\tau)^{2k-1}(\rho(\tau) + |\sigma(\tau)|)^k}$$

uniformly for all  $\tau \in I_\theta$  satisfying  $\rho(\tau) > 0$  and all  $k \in \mathbb{N}$  satisfying  $k \geq 1$ . Here, we set  $|\sigma(\tau)| := 0$  if  $\rho(\tau) > \rho_*$  with  $\rho_*$  as in Lemma 5.5.

*Proof.* To indicate the mechanism, we first prove that, for all  $\tau \in I_\theta$  satisfying  $\rho(\tau) > 0$ , we have

$$\begin{aligned} \|\partial_z m(\tau)\| &\lesssim \rho^{-1}(\rho + |\sigma|)^{-1}, & \|\partial_z^2 m(\tau)\| &\lesssim \rho^{-3}(\rho + |\sigma|)^{-2}, \\ \|\partial_z^3 m(\tau)\| &\lesssim \rho^{-5}(\rho + |\sigma|)^{-3}, \end{aligned} \tag{5.43}$$

where  $\rho := \rho(\tau)$  and  $\sigma := \sigma(\tau)$ .

Since  $\rho(\tau) > 0$ ,  $m$  is real analytic around  $\tau$  by Proposition 4.7 and we can differentiate the Dyson equation, (2.3), with respect to  $z$  and evaluate at  $z = \tau$ . Differentiating (2.3) iteratively yields

$$\begin{aligned} B[\partial_z m] &= m^2, & B[\partial_z^2 m] &= 2(\partial_z m)m^{-1}(\partial_z m), \\ B[\partial_z^3 m] &= -6(\partial_z m)m^{-1}(\partial_z m)m^{-1}(\partial_z m) \\ &\quad + 3(\partial_z^2 m)m^{-1}(\partial_z m) + 3(\partial_z m)m^{-1}(\partial_z^2 m) \end{aligned} \tag{5.44}$$



where  $B = \text{Id} - C_m S$  and  $m := m(\tau)$ . Since  $\rho(\tau) > 0$ ,  $B$  is invertible by (5.24), (5.26) and the  $1/3$ -Hölder continuity of  $m$  by Proposition 4.7.

We set  $\rho := \rho(\tau)$ . If  $\rho > \rho_*$  for some  $\rho_* \sim 1$  then (5.43) follows trivially from (5.44),  $\|B^{-1}\| \lesssim 1$  by (5.24) and  $\|m\| + \|m^{-1}\| \lesssim 1$ .

We now prove (5.43) for  $\rho \leq \rho_*$  and some sufficiently small  $\rho_* \sim 1$ . Under this assumption, Lemma 5.1 and Corollary 5.2 are applicable. In the remainder of this proof, the eigenvalue  $\beta$ , the eigenvectors  $l$  and  $b$  as well as the spectral projections  $P$  and  $Q$  are understood to be evaluated at  $\tau$ . We will now estimate the image of  $B^{-1}$  applied to the right-hand sides of (5.44) in order to prove (5.43).

Inserting  $P + Q = \text{Id}$  on the right-hand side of the first identity in (5.44), inverting  $B$  and using

$$P = \frac{\langle l, \cdot \rangle}{\langle l, b \rangle} b$$

as well as  $B^{-1}[b] = \beta^{-1}b$  yield

$$\partial_z m = \frac{\langle l, m^2 \rangle}{\beta \langle l, b \rangle} b + B^{-1}Q[m^2]. \tag{5.45}$$

We will now estimate  $\langle l, m^2 \rangle$  and  $\beta \langle l, b \rangle$ . From  $m = q^* s q + \mathcal{O}(\rho)$  by (5.2), (5.14a), (5.15b) and (5.26), we obtain

$$\langle l, m^2 \rangle = \langle f_u s q q^* s \rangle + \mathcal{O}(\rho) = \pi + \mathcal{O}(\rho), \tag{5.46}$$

where we used  $s f_u s = f_u s^2 = f_u$  and (5.19) in the last step. From (5.15c) and (5.26), we conclude

$$\beta \langle l, b \rangle = -2i\rho\sigma + \rho^2 \left( \psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^3). \tag{5.47}$$

Here and in the remainder of the proof,  $\sigma, \psi, f_u, q$  and  $s$  are understood to be evaluated at  $\tau$ .

Since  $\sigma$  and  $\psi$  are real, we conclude  $|\beta \langle l, b \rangle| \sim \rho(\rho + |\sigma|)$  for  $\rho_* \sim 1$  sufficiently small. As  $\|B^{-1}Q\| \lesssim 1$  and  $\|b\| \lesssim 1$ , we thus obtain  $\|\partial_z m\| \lesssim \rho^{-1}(\rho + |\sigma|)^{-1}$  from (5.45).

Using (5.44), (5.45),  $\|\partial_z m\| \lesssim \rho^{-1}(\rho + |\sigma|)^{-1}$  and  $\|B^{-1}\| \lesssim \rho^{-1}(\rho + |\sigma|)^{-1}$  by Corollary 5.3 yield

$$\partial_z^2 m = 2 \frac{\langle l, m^2 \rangle^2 \langle l, b m^{-1} b \rangle}{(\beta \langle l, b \rangle)^3} b + \mathcal{O}(\rho^{-2}(\rho + |\sigma|)^{-2}) = \mathcal{O}(\rho^{-3}(\rho + |\sigma|)^{-2}). \tag{5.48}$$

Here, in the last step, we used  $\|b\| \lesssim 1$  and  $|\langle l, b m^{-1} b \rangle| \lesssim |\sigma| + \rho$  due to the expansion

$$\langle l, b m^{-1} b \rangle = \langle q^{-1} f_u (q^*)^{-1} q^* f_u q q^{-1} s (q^*)^{-1} q^* f_u q \rangle + \mathcal{O}(\rho) = \sigma + \mathcal{O}(\rho) \tag{5.49}$$

as well as  $|\beta\langle l, b \rangle| \sim \rho(\rho + |\sigma|)$  and  $\langle l, m^2 \rangle = \mathcal{O}(1)$ . The proof of (5.49) is a consequence of (5.14a), (5.15a), (5.15b), (5.26),  $m^{-1} = q^{-1}s(q^*)^{-1} + \mathcal{O}(\rho)$  by (5.2) as well as  $q \sim 1$ .

Similarly, owing to (5.44), (5.45) and (5.48), we obtain

$$\partial_z^3 m = 12 \frac{\langle l, m^2 \rangle^3 \langle l, bm^{-1}b \rangle^2}{(\beta\langle l, b \rangle)^5} b + \mathcal{O}(\rho^{-5}(\rho + |\sigma|)^{-3}) = \mathcal{O}(\rho^{-5}(\rho + |\sigma|)^{-3}).$$

We now estimate  $\partial_z^k m(z)$  for  $k > 3$ . To that end, we will fix a parameter  $\alpha > 1$  and prove that there are  $\rho_* \sim 1$ ,  $C_1 \sim_\alpha 1$  and  $C_2 \sim_\alpha 1$  such that, for  $k \in \mathbb{N}$ , we have

$$m^{(k)} := \partial_z^k m = \beta_k b + q_k, \tag{5.50}$$

where  $m = m(\tau)$  for  $\tau \in I_\theta$  satisfying  $\rho := \rho(\tau) \leq \rho_*$  and  $\beta_k \in \mathbb{C}$  and  $q_k \in \text{ran } Q$  satisfy

$$|\beta_k| \leq \frac{k! C_1 C_2^{k-1}}{k^\alpha} \rho^{-2k+1} (\rho + |\sigma|)^{-k}, \quad \|q_k\| \leq \frac{k! C_1 C_2^{k-1}}{k^\alpha} \rho^{-2k+2} (\rho + |\sigma|)^{-k}. \tag{5.51}$$

Here,  $\sim_\alpha$  indicates that the constants in the definition of the comparison relation  $\sim$  will depend on  $\alpha$ .

Before we prove (5.50) below, we note two auxiliary statements. First, as  $\partial_z m^{-1} = -m^{-1}(\partial_z m)m^{-1}$  it is easy to check the following version of the usual Leibniz-rule:

$$\begin{aligned} \partial_z^k m^{-1} &= \sum_{n=1}^k \sum_{\substack{a_1+\dots+a_n=k \\ 1 \leq a_i \leq k}} \frac{k!}{a_1! \dots a_n!} \\ &\quad \times (-1)^n m^{-1} m^{(a_1)} m^{-1} m^{(a_2)} \dots m^{-1} m^{(a_n)} m^{-1} \end{aligned} \tag{5.52}$$

for any  $k \in \mathbb{N}$ . Here, in the sum over  $a_1 + \dots + a_n = k$ , the order of  $a_1, \dots, a_n$  has to be taken into account since  $m^{-1}$  and  $m^{(a)}$  do not commute in general. Second, we also have the following auxiliary bound. For all  $k \in \mathbb{N}$ ,  $n \in \mathbb{N}$  with  $n \leq k$  and  $\alpha > 1$ , we have

$$\sum_{\substack{a_1+\dots+a_n=k \\ 1 \leq a_i \leq k}} \frac{1}{a_1^\alpha \dots a_n^\alpha} \leq \frac{(2^{\alpha+1} \zeta(\alpha))^{n-1}}{k^\alpha}, \tag{5.53}$$

where  $\zeta(\alpha) = \sum_{n=1}^\infty n^{-\alpha}$  is Riemann’s zeta function. The bound in (5.53) can be proven by induction.

We now show (5.50) and (5.51) by induction on  $k$ . The initial step of the induction with  $k = 1$  has been established in (5.45) with  $\beta_1 = \langle l, m^2 \rangle / (\beta\langle l, b \rangle)$ ,  $q_1 = B^{-1}Q[m^2]$  and some sufficiently large  $C_1 \sim 1$ . Next, we establish the induction step by proving (5.50) and (5.51) under the assumption that they

hold true for all derivatives of lower order. From the induction hypothesis, we conclude

$$\|m^{(a)}\| \leq \frac{k!C_1C_2^{a-1}}{k^\alpha} \frac{\|b\| + \rho}{\rho^{2a-1}(\rho + |\sigma|)^a} \tag{5.54}$$

for all  $a \in \mathbb{N}$  satisfying  $1 \leq a \leq k - 1$ . For  $k \geq 2$ , we differentiate (2.3)  $k$ -times and obtain

$$B[\partial_z^k m] = r_k := \partial_z^k m + m(\partial_z^k m^{-1})m. \tag{5.55}$$

By separating the contributions for  $n = 1$  and  $n \geq 2$  in (5.52), we conclude

$$r_k = \sum_{n=3}^k \sum_{\substack{a_1+\dots+a_n=k \\ 1 \leq a_i < k-1}} \frac{k!}{a_1! \dots a_n!} (-1)^n m^{(a_1)} m^{-1} \dots m^{-1} m^{(a_n)} + \sum_{a=1}^{k-1} \frac{k!}{a!(k-a)!} m^{(a)} m^{-1} m^{(k-a)}. \tag{5.56}$$

Since  $n$  is at least 3 in the first sum, we obtain from (5.54) and (5.53) that

$$\begin{aligned} & \sum_{n=3}^k \sum_{\substack{a_1+\dots+a_n=k \\ 1 \leq a_i < k-1}} \frac{k!}{a_1! \dots a_n!} \|m^{(a_1)} m^{-1} \dots m^{-1} m^{(a_n)}\| \\ & \leq \frac{k!}{k^\alpha} \frac{\|b\| + \rho}{\rho^{2k-3}(\rho + |\sigma|)^k} \sum_{n=3}^k C_1^n M_\alpha^{n-1} C_2^{k-n}, \end{aligned} \tag{5.57}$$

where  $M_\alpha := 2^{\alpha+2}\zeta(\alpha)\|m^{-1}\|(\|b\| + \rho)$ . A similar argument yields

$$\sum_{a=1}^{k-1} \frac{k!}{a!(k-a)!} \|m^{(a)} m^{-1} m^{(k-a)}\| \leq \frac{k!}{k^\alpha} \frac{\|b\| + \rho}{\rho^{2k-2}(\rho + |\sigma|)^k} C_1^2 M_\alpha C_2^{k-2}.$$

Thus, we choose  $C_2 \geq 2M_\alpha C_1$  and conclude

$$\|r_k\| \leq \frac{k!}{k^\alpha} \frac{\|b\| + \rho}{\rho^{2k-2}(\rho + |\sigma|)^k} \frac{M_\alpha C_1^2 C_2^k}{C_2^2(1 - M_\alpha C_1/C_2)}.$$

Therefore, we obtain the bound on  $\|q_k\|$  in (5.51) for  $C_2 \sim 1$  sufficiently large since  $q_k = Q[\partial_z^k m] = B^{-1}Q[r_k]$  and  $\|B^{-1}Q\| \lesssim 1$ . Moreover,  $\beta_k = \langle l, r_k \rangle / (\beta \langle l, b \rangle)$ . Hence, by using the decomposition of  $r_k$  in (5.56) and (5.57), we obtain

$$\begin{aligned} |\beta_k| & \leq \frac{k!C_1C_2^{k-1}}{k^\alpha} \frac{\|b\| + \rho}{\rho^{2k-1}(\rho + |\sigma|)^k} \frac{\|l\|\rho^2}{|\beta \langle l, b \rangle|} \frac{C_1^2 M_\alpha^2}{C_2^2(1 - M_\alpha C_1/C_2)} \\ & \quad + \sum_{a=1}^{k-1} \frac{k!}{a!(k-a)!} \frac{|\langle l, m^{(a)} m^{-1} m^{(k-a)} \rangle|}{|\beta \langle l, b \rangle|} \end{aligned}$$

We use (5.50) for  $m^{(a)}$  and  $m^{(k-a)}$  in the argument of the last sum, which yields

$$\begin{aligned} & \frac{1}{a!(k-a)!} \frac{|\langle l, m^{(a)}m^{-1}m^{(k-a)} \rangle|}{|\beta \langle l, b \rangle|} \\ & \leq \frac{|\beta_a|}{a!} \frac{|\beta_{k-a}|}{(k-a)!} \frac{|\langle l, bm^{-1}b \rangle|}{|\beta \langle l, b \rangle|} \\ & \quad + \frac{C_1^2 C_2^{k-2}}{a^\alpha (k-a)^\alpha \rho^{2k-1} (\rho + |\sigma|)^k} \frac{\rho^2 \|l\| \|m^{-1}\|}{|\beta \langle l, b \rangle|} (2\|b\| + \rho) \\ & \leq \frac{C_1^2 C_2^{k-2}}{a^\alpha (k-a)^\alpha \rho^{2k-1} (\rho + |\sigma|)^k} \frac{\rho(\rho + |\sigma|)}{|\beta \langle l, b \rangle|} \left( \frac{|\langle l, bm^{-1}b \rangle|}{\rho + |\sigma|} + \|l\| \|m^{-1}\| (2\|b\| + \rho) \right) \end{aligned}$$

Here, we applied (5.51) to estimate  $q_a$  and  $q_{k-a}$  as well as  $\beta_a$  and  $\beta_{k-a}$ . Since  $|\beta \langle l, b \rangle| \sim \rho(\rho + |\sigma|)$  as shown below (5.47) and  $|\langle l, bm^{-1}b \rangle| \lesssim |\sigma| + \rho$  due to (5.49), we obtain the bound on  $|\beta_k|$  in (5.51) by using (5.53) to perform the summation over  $a$ . This completes the induction argument, which yields (5.50) and (5.51) for all  $k \in \mathbb{N}$  by possibly increasing  $C_2 \sim 1$ . By choosing, say,  $\alpha = 2$ , we immediately conclude Lemma 5.7 for  $\tau \in I_\theta$  satisfying  $\rho(\tau) \leq \rho_*$ . If  $\rho(\tau) > \rho_*$  then  $\|B^{-1}\| \lesssim 1$ . Hence, a simple induction argument using (5.55) and (5.56), which hold true for  $\rho(\tau) > \rho_*$  as well, yields some  $C \sim 1$  such that

$$\|\partial_z^k m(\tau)\| \lesssim k! C^k$$

for all  $k \in \mathbb{N}$  satisfying  $k \geq 1$ . Since  $\rho(\tau) \lesssim 1$  for all  $\tau \in I_\theta$ , we obtain Lemma 5.7 in the missing regime.  $\square$

### 6 THE CUBIC EQUATION

The following Proposition 6.1 is the main result of this section. It asserts that  $m$  is determined by the solution to a cubic equation, (6.3) below, close to points  $\tau_0 \in \text{supp } \rho$  of small density  $\rho(\tau_0)$ . In Section 7, this cubic equation will allow for a classification of the small local minima of  $\tau \mapsto \rho(\tau)$ . To have a short notation for the elements of  $\text{supp } \rho$  of small density well separated from the edges of  $I$ , we introduce the set

$$\mathbb{D}_{\varepsilon, \theta} := \{\tau \in \text{supp } \rho \cap I : \rho(\tau) \in [0, \varepsilon], \text{ dist}(\tau, \partial I) \geq \theta\}$$

for  $\varepsilon > 0$  and  $\theta > 0$ .

The leading order terms of the cubic and quadratic coefficients in (6.3) are given by  $\psi(\tau_0)$  and  $\sigma(\tau_0)$ , respectively. For their definitions, we refer to Lemma 5.5 (i) and (5.12).

**PROPOSITION 6.1** (Cubic equation for shape analysis). *Let  $I \subset \mathbb{R}$  be an open interval and  $\theta \in (0, 1]$ . If Assumptions 4.5 hold true on  $I$  for some  $\eta_* \in (0, 1]$  then there are thresholds  $\rho_* \sim 1$  and  $\delta_* \sim 1$  such that, for all  $\tau_0 \in \mathbb{D}_{\rho_*, \theta}$ , the following hold true:*

(a) The eigenvectors  $l = l(\tau_0)$ ,  $b = b(\tau_0)$  and the spectral projection  $Q = Q(\tau_0)$  of  $B(\tau_0)$  exist as in Corollary 5.2. Consequently, for all  $\omega \in [-\delta_*, \delta_*]$ , we have

$$m(\tau_0 + \omega) - m(\tau_0) = \Theta(\omega)b + r(\omega), \tag{6.1}$$

where  $\Theta: [-\delta_*, \delta_*] \rightarrow \mathbb{C}$  and  $r: [-\delta_*, \delta_*] \rightarrow \mathcal{A}$  are defined by

$$\Theta(\omega) := \left\langle \frac{l}{\langle b, l \rangle}, m(\tau_0 + \omega) - m(\tau_0) \right\rangle, \quad r(\omega) := Q[m(\tau_0 + \omega) - m(\tau_0)]. \tag{6.2}$$

Moreover,  $b = b^* + \mathcal{O}(\rho)$  and  $l = l^* + \mathcal{O}(\rho)$  as well as  $b + b^* \sim 1$  and  $l + l^* \sim 1$  with  $\rho = \rho(\tau_0) = \langle \text{Im } m(\tau_0) \rangle / \pi$ .

(b) The function  $\Theta$  satisfies the cubic equation

$$\mu_3 \Theta^3(\omega) + \mu_2 \Theta^2(\omega) + \mu_1 \Theta(\omega) + \omega \Xi(\omega) = 0 \tag{6.3}$$

for all  $\omega \in [-\delta_*, \delta_*]$ . The complex coefficients  $\mu_3, \mu_2, \mu_1$  and  $\Xi$  in (6.3) fulfill

$$\mu_3 = \psi + \mathcal{O}(\rho), \tag{6.4a}$$

$$\mu_2 = \sigma + i\rho \left( 3\psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^2), \tag{6.4b}$$

$$\mu_1 = 2i\rho\sigma - 2\rho^2 \left( \psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^3), \tag{6.4c}$$

$$\Xi(\omega) = \pi(1 + \nu(\omega)) + \mathcal{O}(\rho), \tag{6.4d}$$

where  $\sigma = \sigma(\tau_0)$  as well as  $\psi = \psi(\tau_0)$  are the extensions of  $\sigma$  and  $\psi$  from (5.12) to the real line (see Lemma 5.5 (i)). For the error term  $\nu(\omega)$ , we have

$$|\nu(\omega)| \lesssim |\Theta(\omega)| + |\omega| \lesssim |\omega|^{1/3}. \tag{6.5}$$

for all  $\omega \in [-\delta_*, \delta_*]$ . Uniformly for  $\tau_0 \in \mathbb{D}_{\rho_*, \theta}$ , we have

$$\psi + \sigma^2 \sim 1. \tag{6.6}$$

(c) Moreover,  $\Theta(\omega)$  and  $r(\omega)$  are bounded by

$$|\Theta(\omega)| \lesssim \min \left\{ \frac{|\omega|}{\rho^2}, |\omega|^{1/3} \right\}, \tag{6.7a}$$

$$\|r(\omega)\| \lesssim |\Theta(\omega)|^2 + |\omega|, \tag{6.7b}$$

uniformly for all  $\omega \in [-\delta_*, \delta_*]$ .

(d1) If  $\rho > 0$  then  $\Theta$  and  $r$  are differentiable in  $\omega$  at  $\omega = 0$ .

(d2) If  $\rho = 0$  then we have

$$\begin{aligned} \operatorname{Im} \Theta(\omega) &\geq 0, & |\operatorname{Im} \nu(\omega)| &\lesssim \operatorname{Im} \Theta(\omega), \\ \|\operatorname{Im} r(\omega)\| &\lesssim (|\Theta(\omega)| + |\omega|)\operatorname{Im} \Theta(\omega), \end{aligned} \tag{6.8}$$

for all  $\omega \in [-\delta_*, \delta_*]$  and  $\operatorname{Re} \Theta$  is non-decreasing on the connected components of  $\{\omega \in [-\delta_*, \delta_*]: \operatorname{Im} \Theta(\omega) = 0\}$ .

(e) The function  $\sigma: \mathbb{D}_{\rho_*, \theta} \rightarrow \mathbb{R}$  is uniformly  $1/3$ -Hölder continuous.

The previous proposition is the analogue of Lemma 9.1 in [3]. It should also be compared to [8, Proposition 4.12], where the shape analysis was performed only in a neighbourhood of an edge and thus a lower order accuracy was sufficient. The cubic equation for  $\Theta$ , (6.3), will be obtained from an  $\mathcal{A}$ -valued quadratic equation for  $\Delta := m(\tau_0 + \omega) - m(\tau_0)$  and the results of Section 5. In fact, we have

$$(\operatorname{Id} - C_m S)[\Delta] = \omega m^2 + \frac{\omega}{2}(m\Delta + \Delta m) + \frac{1}{2}(mS[\Delta]\Delta + \Delta S[\Delta]m), \tag{6.9}$$

where  $\tau_0, \tau_0 + \omega \in I_\theta := \{\tau \in I: \operatorname{dist}(\tau, \partial I) \geq \theta\}$  and  $m := m(\tau_0)$  (see the proof of Proposition 6.1 in Section 6.3 below for a derivation of (6.9)). Projecting (6.9) onto the direction  $b$  and its complement, where  $b$  is the unstable direction of  $B$  defined in Corollary 5.2, yields the cubic equation, (6.3), for the contribution  $\Theta$  of  $\Delta$  parallel with  $b$ . In the next subsection, this derivation is presented in a more abstract and transparent setting of a general  $\mathcal{A}$ -valued quadratic equation. After that, the coefficients of the cubic equation are computed in Lemma 6.3 in the setup of (6.9) before we prove Proposition 6.1 in Section 6.3.

### 6.1 GENERAL CUBIC EQUATION

Let  $B, T: \mathcal{A} \rightarrow \mathcal{A}$  be linear maps,  $A: \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A}$  a bilinear map and  $K: \mathcal{A} \times \mathcal{A} \rightarrow \mathcal{A}$  a map. For  $\Delta, e \in \mathcal{A}$ , we consider the quadratic equation

$$B[\Delta] - A[\Delta, \Delta] - T[e] - K[e, \Delta] = 0. \tag{6.10}$$

We view this as an equation for  $\Delta$ , where  $e$  is a (small) error term. This quadratic equation is a generalization of the stability equation (6.9) for the Dyson equation, (2.3) (see (6.23) and (6.28) below for the concrete choices of  $B, T, A, e$  and  $K$  in the setting of (6.9)).

Suppose that  $B$  has a non-degenerate isolated eigenvalue  $\beta$  and a corresponding eigenvector  $b$ , i.e.,  $B[b] = \beta b$  and  $D_r(\beta) \cap \operatorname{Spec}(B) = \{\beta\}$  for some  $r > 0$ . We denote the spectral projection corresponding to  $\beta$  and its complementary projection by  $P$  and  $Q$ , respectively, i.e.,

$$P := -\frac{1}{2\pi i} \oint_{\partial D_r(\beta)} (B - \omega \operatorname{Id})^{-1} d\omega = \frac{\langle l, \cdot \rangle}{\langle l, b \rangle} b, \quad Q := \operatorname{Id} - P. \tag{6.11}$$

Here,  $l \in \mathcal{A}$  is an eigenvector of  $B^*$  corresponding to its eigenvalue  $\bar{\beta}$ , i.e.,  $B^*[l] = \bar{\beta}l$ . In the following, we will assume that

$$\begin{aligned} \|B^{-1}Q[x]\| &\lesssim \|x\|, & |\langle l, b \rangle|^{-1} + \|b\| + \|l\| &\lesssim 1, & \|A[x, y]\| &\lesssim \|x\|\|y\|, \\ \|T[e]\| &\lesssim \|e\|, & \|K[e, y]\| &\lesssim \|e\|\|y\| \end{aligned} \tag{6.12}$$

for all  $x, y \in \mathcal{A}$  and the  $e \in \mathcal{A}$  from (6.10). The guiding idea is that the main contribution in the decomposition

$$\Delta = \Theta b + Q[\Delta], \quad \Theta := \frac{\langle l, \Delta \rangle}{\langle l, b \rangle} \tag{6.13}$$

is given by  $\Theta$ , i.e., the coefficient of  $\Delta$  in the direction  $b$ , under the assumption that  $\Delta$  is small. If  $A = K = 0$  then this would be a simple linear stability analysis of the equation  $B[\Delta] = \text{small}$  around an isolated eigenvalue of  $B$ . The presence of the quadratic terms in (6.10) requires to follow second and third order terms carefully. In the following lemma, we show that the behaviour of  $\Theta$  is governed by a scalar-valued cubic equation (see (6.14) below) and that  $Q[\Delta]$  is indeed dominated by  $\Theta$ . The implicit constants in (6.12) are the model parameters in Section 6.1. We stress that the implicit constants in our estimates only depend on these model parameters and not on any other quantity.

LEMMA 6.2 (General cubic equation). *Let  $\beta$  be a non-degenerate isolated eigenvalue of  $B$ . Let  $\Delta \in \mathcal{A}$  and  $e \in \mathcal{A}$  satisfy (6.10),  $\Theta$  be defined as in (6.13) and the conditions in (6.12) hold true. Then there is  $\varepsilon \sim 1$  such that if  $\|\Delta\| \leq \varepsilon$  then  $\Theta$  satisfies the cubic equation*

$$\mu_3 \Theta^3 + \mu_2 \Theta^2 + \mu_1 \Theta + \mu_0 = \tilde{e}, \tag{6.14}$$

with some  $\tilde{e} = \mathcal{O}(|\Theta|^4 + |\Theta|\|e\| + \|e\|^2)$  and with coefficients

$$\begin{aligned} \mu_3 &= \langle l, A[b, B^{-1}QA[b, b]] + A[B^{-1}QA[b, b], b] \rangle, \\ \mu_2 &= \langle l, A[b, b] \rangle, \\ \mu_1 &= -\beta \langle l, b \rangle, \\ \mu_0 &= \langle l, T[e] \rangle. \end{aligned} \tag{6.15}$$

Moreover, we have

$$Q[\Delta] = B^{-1}QT[e] + \mathcal{O}(|\Theta|^2 + \|e\|^2). \tag{6.16}$$

If we additionally assume that  $\text{Im } \Delta, l, b \in \overline{\mathcal{A}}_+$  and  $K$  is linear in its second argument as well as

$$B[x]^* = B[x^*], \quad A[x, y]^* = A[x^*, y^*], \quad T[e]^* = T[e], \quad K[e, y]^* = K[e, y^*] \tag{6.17}$$

for all  $x, y \in \mathcal{A}$  then there are  $\varepsilon \sim 1$  and  $\delta \sim 1$  such that  $\|\Delta\| \leq \varepsilon$  and  $\|e\| \leq \delta$  also imply

$$\|\text{Im } Q[\Delta]\| \lesssim (|\Theta| + \|e\|)\text{Im } \Theta, \tag{6.18a}$$

$$|\text{Im } \tilde{e}| \lesssim (|\Theta|^3 + \|e\|)\text{Im } \Theta. \tag{6.18b}$$

*Proof.* Setting  $r := Q[\Delta]$ , the quadratic equation (6.10) reads as

$$\Theta\beta b + Br = T[e] + A[\Delta, \Delta] + K[e, \Delta]. \quad (6.19)$$

By applying  $Q$  and afterwards  $B^{-1}$  to the previous relation, we conclude that

$$\begin{aligned} r &= B^{-1}QT[e] + \Theta^2 B^{-1}QA[b, b] + e_1, \\ e_1 &:= \Theta B^{-1}Q(A[b, r] + A[r, b]) + B^{-1}QA[r, r] + B^{-1}QK[e, \Delta]. \end{aligned} \quad (6.20)$$

We have  $\|e_1\| \lesssim \|r\|\|\Theta\| + \|r\|^2 + \|e\|\|\Delta\|$  and  $\|r\| \lesssim \|e\| + |\Theta|^2 + \|e_1\|$ . Hence,

$$\|r\| \lesssim |\Theta|^2 + \|e\| + \|r\|^2 + \|e\|\|\Delta\|.$$

From the second bound in (6.12), we conclude  $\|P\| + \|Q\| \lesssim 1$  and, thus,  $\|r\| + |\Theta| \lesssim \|\Delta\|$ . By choosing  $\varepsilon \sim 1$  small enough, assuming  $\|\Delta\| \leq \varepsilon$  and using  $\|r\| \lesssim \|\Delta\|$  to incorporate  $\|r\|^2$  into the right-hand side, we obtain

$$\|r\| \lesssim |\Theta|^2 + \|e\|. \quad (6.21a)$$

By (6.12) and (6.13), we have  $\|\Delta\| \lesssim |\Theta| + \|r\|$ . This together with (6.21a) plugged into the estimate on  $\|e_1\|$  implies

$$\|e_1\| \lesssim |\Theta|^3 + \|e\|\|\Theta\| + \|e\|^2. \quad (6.21b)$$

This proves (6.16). Defining  $e_2 := e_1 + B^{-1}QT[e]$  yields  $\Delta = \Theta b + \Theta^2 B^{-1}QA[b, b] + e_2$ . By plugging this into (6.19) and computing the scalar product with  $\langle l, \cdot \rangle$ , we obtain

$$\begin{aligned} \Theta\beta\langle l, b \rangle &= \langle l, T[e] \rangle + \Theta^2\langle l, A[b, b] \rangle \\ &\quad + \Theta^3\langle l, A[b, B^{-1}QA[b, b]] + A[B^{-1}QA[b, b], b] \rangle - \tilde{e}, \end{aligned} \quad (6.22a)$$

where we introduced

$$\begin{aligned} \tilde{e} &:= -\langle l, K[e, \Delta] + \Theta^4 A[B^{-1}QA[b, b], B^{-1}QA[b, b]] \\ &\quad + A[\Delta, e_2] + A[e_2, \Delta] + A[e_2, e_2] \rangle. \end{aligned} \quad (6.22b)$$

Since  $\|e_2\| \lesssim |\Theta|^3 + \|e\|$  by (6.21b) and  $\|\Delta\| \lesssim |\Theta| + \|e\|$  by (6.21a) and (6.16), we conclude  $\tilde{e} = \mathcal{O}(|\Theta|^4 + |\Theta|\|e\| + \|e\|^2)$ . Therefore,  $\Theta$  satisfies (6.14) with the coefficients from (6.15).

For the rest of the proof, we additionally assume that the relations in (6.17) hold true. Taking the imaginary part of (6.20) and arguing similarly as after (6.20) yield

$$\begin{aligned} \|\operatorname{Im} e_1\| &\lesssim (\|r\| + \|e\|)\operatorname{Im} \Theta + (\|r\| + |\Theta| + \|e\|)\|\operatorname{Im} r\|, \\ \|\operatorname{Im} r\| &\lesssim |\Theta|\operatorname{Im} \Theta + \|\operatorname{Im} e_1\|. \end{aligned}$$

Hence, (6.18a) and  $\|\operatorname{Im} e_1\| \lesssim (|\Theta|^2 + \|e\|)\operatorname{Im} \Theta$  follow for  $\|\Delta\| \leq \varepsilon$  and  $\|e\| \leq \delta$  with some sufficiently small  $\varepsilon \sim 1$  and  $\delta \sim 1$ . From this and taking the imaginary part in (6.22b), we conclude (6.18b) as  $\|\operatorname{Im} \Delta\| \lesssim \operatorname{Im} \Theta$  by (6.18a) and  $\operatorname{Im} e_2 = \operatorname{Im} e_1$ . This completes the proof of Lemma 6.2.  $\square$



6.2 CUBIC EQUATION ASSOCIATED TO DYSON STABILITY EQUATION

Owing to (6.15), the coefficients  $\mu_3, \mu_2$  and  $\mu_1$  are completely determined by the bilinear map  $A$  and the operator  $B$ . For analyzing the Dyson equation, (2.3), owing to (6.9), the natural choices for  $A$  and  $B$  are

$$B := \text{Id} - C_m S, \quad A[x, y] := \frac{1}{2}(mS[x]y + yS[x]m) \tag{6.23}$$

with  $x, y \in \mathcal{A}$ . In particular,  $Q$  in (6.11) has to be understood with respect to  $B = \text{Id} - C_m S$ . In the next lemma, we compute  $\mu_3, \mu_2$  and  $\mu_1$  with these choices. This computation involves the inverse of  $\text{Id} - C_s F$ .

In order to directly ensure its invertibility, we will assume  $\text{Im } z > 0$ . This assumption will be removed in the proof of Proposition 6.1 in Section 6.3 below.

LEMMA 6.3 (Coefficients of the cubic for Dyson equation). *Let  $A$  and  $B$  be defined as in (6.23). If Assumptions 4.5 hold true on an interval  $I \subset \mathbb{R}$  for some  $\eta_* \in (0, 1]$  then there is a threshold  $\rho_* \sim 1$  such that, for  $z \in \mathbb{H}_{I, \eta_*}$  satisfying  $\rho(z) + \rho(z)^{-1} \text{Im } z \leq \rho_*$ , the coefficients of the cubic (6.14) have the expansions*

$$\mu_3 = \psi + \mathcal{O}(\rho + \rho^{-1} \text{Im } z), \tag{6.24a}$$

$$\mu_2 = \sigma + i\rho \left( 3\psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^2 + \rho^{-1} \text{Im } z), \tag{6.24b}$$

$$\mu_1 = -\pi\rho^{-1} \text{Im } z + 2i\rho\sigma - 2\rho^2 \left( \psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^3 + \text{Im } z + \rho^{-2} (\text{Im } z)^2), \tag{6.24c}$$

where  $\sigma = \sigma(z)$  and  $\psi = \psi(z)$  are defined as in (5.12). Moreover, we also have

$$\langle l, mS[b]b \rangle = \sigma + i\rho \left( 3\psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^2 + \rho^{-1} \text{Im } z). \tag{6.25}$$

*Proof.* In this proof, we use the convention that concatenation of maps on  $\mathcal{A}$  and evaluation of these maps in elements of  $\mathcal{A}$  are prioritized before the multiplication in  $\mathcal{A}$ , i.e.,

$$AB[b]c := (A[B[b]])c$$

if  $A$  and  $B$  are maps on  $\mathcal{A}$  and  $b, c \in \mathcal{A}$ . We will obtain all expansions in (6.24) from (6.15) by using the special choices for  $A$  and  $B$  from (6.23). Before starting with the proof of (6.24a), we establish a few identities. Recalling  $m = q^*uq$  from (3.2) and (3.4), we first notice the following alternative expression for  $A$

$$A[x, y] = \frac{1}{2} C_{q^*, q} [uFC_{q^*, q}^{-1}[x]C_{q^*, q}^{-1}[y] + C_{q^*, q}^{-1}[y]FC_{q^*, q}^{-1}[x]u] \tag{6.26}$$

with  $x, y \in \mathcal{A}$ . Owing to (4.21), the operators  $C_{q^*, q}$  and  $C_{q^*, q}^{-1}$  are bounded. We choose  $\rho_* \sim 1$  small enough so that Lemma 5.1 is applicable. By using

$u = s + i\text{Im } u + \mathcal{O}(\rho^2)$  due to (5.2) as well as (5.4), (5.5) and (5.14a) in (6.26), we obtain

$$A[b_0, b_0] = C_{q^*, q}[sf_u^2 + i\rho f_u^3] + \mathcal{O}(\rho^2 + \rho^{-1}\text{Im } z). \quad (6.27)$$

Combining (6.27) and (5.18) implies

$$B_0^{-1}Q_0A[b_0, b_0] = C_{q^*, q}(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] + \mathcal{O}(\rho + \rho^{-1}\text{Im } z).$$

We now prove the expansion (6.24a) for  $\mu_3$  by starting from (6.15) and using  $l = l_0 + \mathcal{O}(\rho)$ ,  $b = b_0 + \mathcal{O}(\rho)$  by (5.15),  $B^{-1}Q = B_0^{-1}Q_0 + \mathcal{O}(\rho)$  due to  $B = B_0 + \mathcal{O}(\rho)$  and Lemma 5.1 and the previous identities. This yields

$$\begin{aligned} \mu_3 &= \langle l_0, A[B_0^{-1}Q_0A[b_0, b_0], b_0] + A[b_0, B_0^{-1}Q_0A[b_0, b_0]] \rangle + \mathcal{O}(\rho) \\ &= \frac{1}{2} \langle f_u, uF(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2]f_u + uF[f_u](\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] \\ &\quad + f_uF(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] + (\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2]F[f_u] \rangle \\ &\quad + \mathcal{O}(\rho + \rho^{-1}\text{Im } z) \\ &= \langle sf_u^2, (\text{Id} + F)(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] \rangle + \mathcal{O}(\rho + \rho^{-1}\text{Im } z). \end{aligned}$$

Here, we also used  $F[f_u] = f_u + \mathcal{O}(\rho^{-1}\text{Im } z)$  by (5.5) and  $u = s + \mathcal{O}(\rho)$  by (5.2). This shows (6.24a).

In order to compute  $\mu_2$ , we define

$$\begin{aligned} b_1 &:= 2i\rho C_{q^*, q}(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2], \\ l_1 &:= -2i\rho C_{q, q^*}^{-1}(\text{Id} - FC_s)^{-1}Q_{s,F}^*F[sf_u^2]. \end{aligned}$$

Then we use (5.15a) as well as (5.15b) and obtain

$$\begin{aligned} \langle l, A[b, b] \rangle &= \langle l_0, A[b_0, b_0] \rangle + \langle l_1, A[b_0, b_0] \rangle + \langle l_0, A[b_1, b_0] \rangle + \langle l_0, A[b_0, b_1] \rangle \\ &\quad + \mathcal{O}(\rho^2 + \text{Im } z) \\ &= \langle sf_u^3 \rangle + i\rho \langle f_u^4 \rangle + 2i\rho \langle sf_u^2, (\text{Id} + 2F)(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] \rangle \\ &\quad + \mathcal{O}(\rho^2 + \rho^{-1}\text{Im } z) \\ &= \sigma + i\rho \left( 3\psi + \frac{\sigma^2}{\langle f_u^2 \rangle} \right) + \mathcal{O}(\rho^2 + \rho^{-1}\text{Im } z). \end{aligned}$$

Here, in the second step, we used (5.14a), (6.27) and the definition of  $l_1$  to compute the first and second term, (5.14a), the definition of  $b_1$  and (6.26) to compute the third and fourth term. In the last step, we then employed

$$\begin{aligned} \langle f_u^4 \rangle + \langle sf_u^2, 2(\text{Id} + 2F)(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] \rangle \\ &= \langle sf_u^2, (\text{Id} + 2(\text{Id} + 2F)(\text{Id} - C_sF)^{-1})Q_{s,F}[sf_u^2] \rangle \\ &\quad + \langle sf_u^2, P_{s,F}[sf_u^2] \rangle \\ &= 3 \langle sf_u^2, (\text{Id} + F)(\text{Id} - C_sF)^{-1}Q_{s,F}[sf_u^2] \rangle + \frac{\sigma^2}{\langle f_u^2 \rangle} \\ &\quad + \mathcal{O}(\rho^{-1}\text{Im } z). \end{aligned}$$

Here, we applied (5.17),  $C_s = C_s^*$  and  $C_s[sf_u^2] = sf_u^2$ . Since  $\mu_2 = \langle l, A[b, b] \rangle$  by (6.15), this completes the proof of (6.24b). A similar computation as the one for  $\mu_2$  yields (6.25).

Since  $\mu_1 = -\beta \langle l, b \rangle$  by (6.15), the expansion in (5.15c) immediately yields (6.24c). This completes the proof of the lemma.  $\square$

### 6.3 THE CUBIC EQUATION FOR THE SHAPE ANALYSIS

In this subsection, we will prove Proposition 6.1 by using Lemma 6.2 and Lemma 6.3. Therefore, in addition to the choices of  $A$  and  $B$  in (6.23), we choose  $\Delta = m(\tau_0 + \omega) - m(\tau_0)$ ,  $\tau_0, \tau_0 + \omega \in I$  and

$$e = \omega \mathbf{1}, \quad T[x] = xm^2, \quad K[x, y] = \frac{1}{2}(xmy + ymx) \tag{6.28}$$

for  $x, y \in \mathcal{A}$  with  $m = m(\tau_0)$  in (6.10).

*Proof of Proposition 6.1.* We choose  $\rho_* \sim 1$  such that Lemma 5.1 and Corollary 5.2 are applicable. We fix  $\tau_0 \in \mathbb{D}_{\rho_*, \theta}$  and set  $m = m(\tau_0)$ . The statements about  $l$  and  $b$  in (a) of Proposition 6.1 follow from Corollary 5.2. In particular,  $|\langle l, b \rangle| \sim 1$ . Thus, the conditions in (6.12) are a direct consequence of Assumptions 4.5, (4.21), Lemma 5.1 and Corollary 5.2. Furthermore, if  $\rho = 0$  then we have  $m = m^*$  and, thus, (6.17) follows. For  $\omega \in [-\delta_*, \delta_*]$ ,  $\delta_* := \theta/2$ , we set  $\Delta = m(\tau_0 + \omega) - m$ . Since  $\Theta(\omega)b = P[\Delta]$ ,  $r(\omega) = Q[\Delta]$  and  $P + Q = \text{Id}$ , we immediately obtain (6.1). This proves (a).

Next, we derive (6.9) for  $\Delta := m(z_0 + \omega) - m(z_0)$  and  $m := m(z_0)$  with  $z_0 := \tau_0 + i\eta$ ,  $\tau_0 \in \mathbb{D}_{\rho_*, \theta}$ ,  $\omega \in [-\delta_*, \delta_*]$  and  $\eta \in (0, \eta_*]$ . We subtract (2.3) evaluated at  $z = z_0$  from (2.3) evaluated at  $z = z_0 + \omega$  and obtain (6.9) with  $\Delta$  and  $m$  defined at  $z_0 = \tau_0 + i\eta$ . Directly taking the limit  $\eta \downarrow 0$  yields (6.9) with the original choices of  $\Delta$  and  $m$  at  $z_0 = \tau_0$  by the Hölder-continuity of  $m$  on  $\overline{\mathbb{H}}_{I', \eta_*}$ ,  $I' := \{\tau \in I : \text{dist}(\tau, \partial I) \geq \theta/2\}$ , due to Proposition 4.7.

Lemma 6.2 is applicable for  $|\omega| \leq \delta_*$  with some sufficiently small  $\delta_* \sim 1$  since this guarantees  $\|\Delta\| \leq \varepsilon$  owing to the Hölder-continuity of  $m$ . Hence, Lemma 6.2 yields a cubic equation for  $\Theta$  as defined in (6.2) with  $l = l(z_0)$ ,  $b = b(z_0)$  and  $z_0 = \tau_0 + i\eta$ . The coefficients of this cubic equation are given in Lemma 6.2. Owing to the uniform 1/3-Hölder continuity of  $z \mapsto m(z)$  on  $\overline{\mathbb{H}}_{I', \eta_*}$ , we conclude from the definition of  $\Theta$  and  $r := Q[\Delta]$  in (6.2), the boundedness of  $Q$  and  $B^{-1}Q$  as well as (6.16) that  $|\Theta(\omega)| \lesssim |\omega|^{1/3}$ , i.e., the second bound in (6.7a), and (6.7b) uniformly for  $\eta \in [0, \eta_*]$ .

We now compute the coefficients of the cubic in (6.3) for  $\tau_0 \in \mathbb{D}_{\rho_*, \theta}$ . Set  $z_0 := \tau_0 + i\eta$ . Note that for  $\eta = \text{Im } z_0 > 0$  these coefficients were already given in (6.24), so the only task is to check their limit behaviour as  $\eta \downarrow 0$ . Owing to (5.26), the expansions in (6.4a), (6.4b) and (6.4c) follow from (6.24a), (6.24b) and (6.24c), respectively, using the continuity of  $\sigma$ ,  $\psi$  and  $f_u$  on  $\overline{\mathbb{H}}_{\text{small}}$  by Lemma 5.5 and Lemma 5.4, respectively. We now show (6.4d). With the definitions of  $\tilde{e}$  and  $\mu_0$  from Lemma 6.2 (see (6.22b) and (6.15), respectively), we

set  $\Xi(\omega) := \omega^{-1}(\mu_0 - \tilde{e})$  for arbitrary  $|\omega| \leq \delta_*$ . Since  $l = C_{q,q^*}^{-1}[f_u] + \mathcal{O}(\rho + \rho^{-1}\eta)$  due to (5.14a) and (5.15b), as well as  $m^2 = (\operatorname{Re} m)^2 + \mathcal{O}(\rho) = C_{q^*,q} C_s[qq^*] + \mathcal{O}(\rho)$  due to  $\operatorname{Im} m \sim \rho \mathbf{1}$  and (5.2), we have

$$\omega^{-1}\mu_0 = \langle l^* m^2 \rangle = \langle f_u qq^* \rangle + \mathcal{O}(\rho + \rho^{-1}\eta) = \pi + \mathcal{O}(\rho + \rho^{-1}\eta). \quad (6.29)$$

Here, we also used  $C_s[f_u] = f_u$  in the second step and (5.19) in the last step. We set  $\nu(\omega) := -(\omega\pi)^{-1}\tilde{e}$ . We recall  $e = \omega \mathbf{1}$ . Since  $\tilde{e} = \mathcal{O}(|\Theta(\omega)|^4 + |\Theta(\omega)||\omega| + |\omega|^2)$  and  $|\Theta(\omega)| \lesssim |\omega|^{1/3}$ , we obtain (6.5). This yields (6.4d) by using (5.26) in (6.29). Since (5.35) implies (6.6), this completes the proof of (b) for  $\tau_0 \in \mathbb{D}_{\rho_*,\theta}$  and we assume  $\eta = 0$  in the following.

If  $\rho = \rho(\tau_0) > 0$  then (4.20) yields the missing first bound in (6.7a) completing the proof of part (c). Moreover, in this case, the definitions of  $\Theta$  and  $r$  imply their differentiability at  $\omega = 0$  due to Proposition 4.7. This shows (d1).

We now verify (d2). Since  $\rho = 0$ , we have  $\operatorname{Im} m(\tau_0) = 0$  and thus  $\operatorname{Im} \Theta(\omega) \geq 0$  by the positive semidefiniteness of  $\operatorname{Im} m(\tau_0 + \omega)$ . Since  $\mu_0$  is real as  $l$  and  $T[e]$  are self-adjoint, we obtain the second bound in (6.8) directly from (6.18b) and  $|\Theta(\omega)| \lesssim |\omega|^{1/3}$ . The third bound in (6.8) follows from (6.18a) and  $e = \omega \mathbf{1}$ . Since  $\rho = 0$  and hence  $b = C_{q^*,q}[f_u]$  by (5.15a) and  $l = C_{q,q^*}^{-1}[f_u]$  by (5.15b) are positive definite elements of  $\mathcal{A}$ ,  $\operatorname{Re} \Theta(\omega) + \langle l, m(\tau_0) \rangle / \langle l, b \rangle$  is the real part of the Stieltjes transform of a positive measure  $\mu$  evaluated on the real axis. The real part of a Stieltjes transform is non-decreasing on the connected components of the complement in  $\mathbb{R}$  of the support of its defining measure. Therefore, as the support of  $\mu$  is contained in  $\mathbb{R} \setminus \operatorname{int}(\{\omega \in [-\delta_*, \delta_*] : \operatorname{Im} \Theta(\omega) = 0\})$ , where  $\operatorname{int}$  denotes the interior, due to  $\operatorname{Im} m(\tau_0) = 0$ , we conclude that  $\operatorname{Re} \Theta(\omega)$  is non-decreasing on the connected components of  $\{\omega \in [-\delta_*, \delta_*] : \operatorname{Im} \Theta(\omega) = 0\}$ .

Lemma 5.5 (i) directly implies the Hölder-continuity in (e), which completes the proof of Proposition 6.1.  $\square$

## 7 CUBIC ANALYSIS

In this section, we prove Theorem 2.5. It will follow directly from its more precise and extended formulation, Theorem 7.1 below, which is the main result of this section. In contrast to Theorem 2.5, Theorem 7.1 treats all small local minima  $\tau_0$  of  $\rho$  – not only those, where  $\rho$  vanishes. Moreover, the error terms are effective and express explicitly the dependence on  $\rho(\tau_0)$ , thus, making the statement uniform in the value of the density at the local minimum. Theorem 7.1 describes the behaviour of  $\operatorname{Im} m$  close to local minima of  $\rho$  inside of  $\operatorname{supp} \rho$ . This behaviour is governed by the universal shape functions  $\Psi_{\text{edge}}: [0, \infty) \rightarrow \mathbb{R}$

and  $\Psi_{\min} : \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$\Psi_{\text{edge}}(\lambda) := \frac{\sqrt{(1+\lambda)\lambda}}{(1+2\lambda+2\sqrt{(1+\lambda)\lambda})^{2/3} + (1+2\lambda-2\sqrt{(1+\lambda)\lambda})^{2/3} + 1}, \tag{7.1a}$$

$$\Psi_{\min}(\lambda) := \frac{\sqrt{1+\lambda^2}}{(\sqrt{1+\lambda^2} + \lambda)^{2/3} + (\sqrt{1+\lambda^2} - \lambda)^{2/3} - 1} - 1. \tag{7.1b}$$

We remind the reader that the comparison relations  $\lesssim$ ,  $\gtrsim$  and  $\sim$  were defined Convention 3.4. In the following Theorem 7.1, the model parameters for these relations are given by  $c_1$ ,  $c_2$  and  $c_3$  in (3.10),  $k_3$  in (4.16) and  $\theta$  in the definition of  $I_\theta$  in (7.2) below.

**THEOREM 7.1** (Behaviour of  $\text{Im } m$  close to local minima of  $\rho$ ). *Let  $(a, S)$  be a data pair such that (3.10) is satisfied. Let  $m$  be the solution to the associated Dyson equation (2.3) and assume that (4.16) holds true on  $\mathbb{H}_{I, \eta_*}$  for some interval  $I \subset \mathbb{R}$  and some  $\eta_* \in (0, 1]$ . We write  $v := \pi^{-1} \text{Im } m$  and, for some  $\theta \in (0, 1]$ , we set*

$$I_\theta := \{\tau \in I : \text{dist}(\tau, \partial I) \geq \theta\}. \tag{7.2}$$

*Then there are thresholds  $\rho_* \sim 1$  and  $\delta_* \sim 1$  such that if  $\tau_0 \in \text{supp } \rho \cap I_\theta$  is a local minimum of  $\rho$  and  $\rho(\tau_0) \leq \rho_*$  then*

$$v(\tau_0 + \omega) = v(\tau_0) + h\Psi(\omega) + \mathcal{O}\left(\rho(\tau_0)|\omega|^{1/3}\mathbf{1}(|\omega| \lesssim \rho(\tau_0)^3) + \Psi(\omega)^2\right) \tag{7.3}$$

*for  $\omega \in [-\delta_*, \delta_*] \cap D$  with some  $h = h(\tau_0) \in \mathcal{A}$  satisfying  $h \sim 1$ . Moreover, the set  $D$  and the function  $\Psi$  depend only on the type of  $\tau_0$  in the following way:*

- (a) *Left edge: If  $\tau_0 \in (\partial \text{supp } \rho) \setminus \{\inf \text{supp } \rho\}$  is the infimum of a connected component of  $\text{supp } \rho$  and the lower edge of the corresponding gap is in  $I_\theta$ , i.e.,  $\tau_1 := \sup((-\infty, \tau_0) \cap \text{supp } \rho) \in I_\theta$ , then (7.3) holds true with  $v(\tau_0) = 0$ ,  $D = [0, \infty)$  and*

$$\Psi(\omega) = \Delta^{1/3} \Psi_{\text{edge}}\left(\frac{\omega}{\Delta}\right)$$

*where  $\Delta := \tau_0 - \tau_1$ . If  $\tau_0 = \inf \text{supp } \rho$ , or more generally  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0 - \varepsilon, \tau_0]$  with some  $\varepsilon \sim 1$ , then the same conclusion holds true with  $\Delta := 1$ .*

- (b) *Right edge: If  $\tau_0 \in \partial \text{supp } \rho$  is the supremum of a connected component then a similar statement as in the case of a left edge holds true.*
- (c) *Cusp: If  $\tau_0 \notin \partial \text{supp } \rho$  and  $\rho(\tau_0) = 0$  then (7.3) holds true with  $D = \mathbb{R}$  and  $\Psi(\omega) = |\omega|^{1/3}$ .*

(d) Internal minimum: If  $\tau_0 \notin \partial \text{supp } \rho$  and  $\rho(\tau_0) > 0$  then there is  $\tilde{\rho} \sim \rho(\tau_0)$  such that (7.3) holds true with  $D = \mathbb{R}$  and

$$\Psi(\omega) = \tilde{\rho} \Psi_{\min} \left( \frac{\omega}{\tilde{\rho}^3} \right).$$

Theorem 7.1 (a) and (b) are formulated in order to track the dependence of  $v$  near an edge on the gap size  $\Delta$ . The simpler case of a regular edge, i.e. when  $\Delta \sim 1$ , e.g. at the extreme edges, is detailed in Section 7.6, especially in Proposition 7.18 below.

Before we proceed we conclude Theorem 2.5 from Theorem 7.1.

*Proof of Theorem 2.5.* Since  $S \in \Sigma_{\text{flat}}$ , the condition (3.10) is satisfied for some positive constants  $c_1, c_2$  and  $c_3$ . Moreover, by possibly shrinking the interval  $I$  and using the continuity of  $m$ , we find constants  $k_3 > 0$  and  $\eta_* \in (0, 1]$  such that  $\|m(z)\| \leq k_3$  for all  $z \in \mathbb{H}_{I, \eta_*}$ . Thus, the assumptions of Theorem 7.1 are satisfied and Theorem 2.5 follows.  $\square$

We remark that working on the slightly smaller interval  $I_\theta$  instead of  $I$  ensures some stronger a-priori control on  $m$  due to the results in the previous sections, in particular, Proposition 4.7.

If the conditions of Theorem 7.1 hold true, i.e., the data pair  $(a, S)$  satisfies (3.10) and  $m$  satisfies (4.16) on  $\mathbb{H}_{I, \eta_*}$ , then Assumptions 4.5 are fulfilled on  $\mathbb{H}_{I, \eta_*}$  (compare Lemma 4.8 (ii)). In fact, Theorem 7.1 holds true under Assumptions 4.5 which will become apparent from the proof.

Theorem 7.1 contains the most important results of the shape analysis. When considering  $\rho = \langle v \rangle$  instead of  $v$  the coefficient in front of  $\Psi(\omega)$  in (7.3) can be precisely identified as demonstrated in part (i) of Theorem 7.2 below. Moreover, Theorem 7.2 contains additional information on the size of the connected components of  $\text{supp } \rho$  and the distance between local minima; these are collected in part (ii). Note that the same information were also proven in the commutative setup in Theorem 2.6 of [3] and Theorem 7.2 shows that they are also available in our general von Neumann algebra setup.

We remark that  $\Psi_{\min}(\omega) = \Psi_{\min}(-\omega)$  for  $\omega \in \mathbb{R}$  and, for  $\omega > 0, \Delta > 0$  and  $\tilde{\rho} > 0$ , we have

$$\Delta^{1/3} \Psi_{\text{edge}} \left( \frac{\omega}{\Delta} \right) \sim \min \left\{ \frac{\omega^{1/2}}{\Delta^{1/6}}, \omega^{1/3} \right\}, \tag{7.4a}$$

$$\tilde{\rho} \Psi_{\min} \left( \frac{\omega}{\tilde{\rho}^3} \right) \sim \min \left\{ \frac{\omega^2}{\tilde{\rho}^5}, \omega^{1/3} \right\}. \tag{7.4b}$$

In the following Theorem 7.2, the model parameters for the comparison relations  $\sim, \lesssim$  and  $\gtrsim$  are the constants  $k_1, \dots, k_8$  from Assumptions 4.5 and  $\theta$  in the definition of  $I_\theta$  in (7.2).

**THEOREM 7.2** (Behaviour of  $\rho$  near almost cusp points; Structure of the set of minima of  $\rho$ ). *Let  $I \subset \mathbb{R}$  be an open interval and  $\theta \in (0, 1]$ . If Assumptions 4.5*

hold true on  $I$  for some  $\eta_* \in (0, 1]$  (in particular, if the data pair  $(a, S)$  satisfies (3.10) and  $m$  satisfies (4.16) on  $\mathbb{H}_{I, \eta_*}$ ) then the following statements hold true

(i) There are thresholds  $\rho_* \sim 1$ ,  $\sigma_* \sim 1$  and  $\delta_* \sim 1$  such that if  $\tau_0 \in \text{supp } \rho \cap I_\theta$  is a local minimum of  $\rho$  satisfying  $\rho(\tau_0) \leq \rho_*$  then we set  $\Gamma := \sqrt{27}\pi/(2\psi)$  with  $\psi = \psi(\tau_0)$  defined as in Lemma 5.5 and have

(a) (Left edge with small gap) If  $\tau_0 \in \partial \text{supp } \rho \setminus \{\inf \text{supp } \rho\}$  is the infimum of a connected component of  $\text{supp } \rho$ ,  $|\sigma(\tau_0)| \leq \sigma_*$  and the lower edge of the gap lies in  $I_\theta$ , i.e.,  $\tau_1 := \sup((-\infty, \tau_0) \cap \text{supp } \rho) \in I_\theta$ , then

$$\begin{aligned} \rho(\tau_0 + \omega) &= (4\Gamma)^{1/3} \Psi(\omega) + \mathcal{O}(|\sigma(\tau_0)|\Psi(\omega) + \Psi(\omega)^2), \\ \Psi(\omega) &:= \Delta^{1/3} \Psi_{\text{edge}}\left(\frac{\omega}{\Delta}\right) \end{aligned} \tag{7.5a}$$

for all  $\omega \in [0, \delta_*]$ . Here,  $\Gamma \sim 1$  and  $\psi \sim 1$ .

(b) (Right edge with small gap) If  $\tau_0 \in \partial \text{supp } \rho \setminus \{\sup \text{supp } \rho\}$  is the supremum of a connected component then a similar statement as in the case of a left edge holds true.

(c) (Cusp) If  $\tau_0 \notin \partial \text{supp } \rho$  and  $\rho(\tau_0) = 0$  then

$$\rho(\tau_0 + \omega) = \frac{\Gamma^{1/3}}{4^{1/3}} |\omega|^{1/3} + \mathcal{O}(|\omega|^{2/3}) \tag{7.5b}$$

for all  $\omega \in [-\delta_*, \delta_*]$ . Here,  $\Gamma \sim 1$  and  $\psi \sim 1$ .

(d) (Nonzero local minimum) If  $\tau_0 \notin \partial \text{supp } \rho$  and  $\rho(\tau_0) > 0$  then

$$\begin{aligned} \rho(\tau_0 + \omega) &= \rho(\tau_0) \\ &+ \Gamma^{1/3} \Psi(\omega) \times \begin{cases} 1 + \mathcal{O}(\rho(\tau_0)^{1/2}), & \text{if } |\omega| \lesssim \rho(\tau_0)^{7/2}, \\ 1 + \mathcal{O}\left(\frac{\rho(\tau_0)^4}{|\omega|}\right), & \text{if } \rho(\tau_0)^{7/2} \lesssim |\omega| \lesssim \rho(\tau_0)^3, \\ 1 + \mathcal{O}(\Psi(\omega)), & \text{if } \rho(\tau_0)^3 \lesssim |\omega| \leq \delta_*, \end{cases} \\ \Psi(\omega) &:= \tilde{\rho} \Psi_{\min}\left(\frac{\omega}{\tilde{\rho}^3}\right), \quad \tilde{\rho} := \frac{\rho(\tau_0)}{\Gamma^{1/3}} \end{aligned} \tag{7.5c}$$

for all  $\omega \in \mathbb{R}$ . Here,  $\Gamma \sim 1$ ,  $\psi \sim 1$  and the terms  $1 + \mathcal{O}(\dots)$  indicate multiplicative error factors.

(ii) If  $\text{supp } \rho \cap I_\theta \neq \emptyset$  then  $\text{supp } \rho \cap I_\theta$  consists of  $K \sim 1$  intervals, i.e., there are  $\alpha_1, \dots, \alpha_K \in \partial \text{supp } \rho \cup \partial I_\theta$  and  $\beta_1, \dots, \beta_K \in \partial \text{supp } \rho \cup \partial I_\theta$ ,  $\alpha_i < \beta_i < \alpha_{i+1}$ , such that

$$\text{supp } \rho \cap \overline{I_\theta} = \bigcup_{i=1}^K [\alpha_i, \beta_i] \tag{7.6}$$

and  $\beta_i - \alpha_i \sim 1$  if  $\beta_i \neq \sup I_\theta$  and  $\alpha_i \neq \inf I_\theta$ .

For  $\rho_* > 0$ , we define the set  $\mathbb{M}_{\rho_*}$  of small local minima  $\tau$  of  $\rho$  which are not edges of  $\text{supp } \rho$ , i.e.,

$$\mathbb{M}_{\rho_*} := \{ \tau \in (\text{supp } \rho \setminus \partial \text{supp } \rho) \cap I_\theta : \rho(\tau) \leq \rho_*, \rho \text{ has a local minimum at } \tau \}. \tag{7.7}$$

There is a threshold  $\rho_* \sim 1$  such that, for all  $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$  satisfying  $\gamma_1 \neq \gamma_2$  and for all  $i = 1, \dots, K$ , we have

$$|\gamma_1 - \gamma_2| \sim 1, \quad |\alpha_i - \gamma_1| \sim 1, \quad |\beta_i - \gamma_1| \sim 1 \tag{7.8}$$

if  $\alpha_i \neq \inf I_\theta$  and  $\beta_i \neq \sup I_\theta$ .

The factors  $4^{1/3}$  and  $4^{-1/3}$  in the cases (a) and (c) of part (i) of Theorem 7.2 can be eliminated by redefining  $\Gamma$ ,  $\Psi_{\text{edge}}$  and  $\Psi_{\text{min}}$  to bring the leading term on the right-hand sides into the uniform  $\Gamma^{1/3}\Psi(\omega)$  form. We have not used these redefined versions of  $\Gamma$ ,  $\Psi_{\text{edge}}$  and  $\Psi_{\text{min}}$  here in order to be consistent with [3]. We remark that part (i) (a) and (b) of Theorem 7.2 cover only the case of  $\tau_0 \in \partial \text{supp } \rho$  with sufficiently small  $|\sigma(\tau_0)|$ . We will establish later that the smallness of  $|\sigma(\tau_0)|$  corresponds to the smallness of the adjacent gap  $\tau_0 - \tau_1$  (see Lemma 7.15 below). Together with the cusps and the small nonzero local minima, they form the *almost cusp points* of  $\rho$  (see the set  $P_{\text{cusp}}$  in (10.4) later). At the extreme edges  $\inf \text{supp } \rho$  and  $\sup \text{supp } \rho$ ,  $\sigma$  is not so small. Thus, the exclusion of these edges in the statement of Theorem 7.2 (a) and (b) is in fact superfluous. If  $|\sigma(\tau_0)|$  is not so small then  $\rho(\tau_0 + \omega)$  is well approximated by a rescaled version of  $(\omega_\pm)^{1/2}$  (positive and negative part of  $\omega$  for left and right edge, respectively). The precise statement and scaling are given in Proposition 7.18 below.

REMARK 7.3 (Scaling relations for  $\rho(z)$ ). Let  $I \subset \mathbb{R}$  be an open interval,  $\theta \in (0, 1]$  and  $\rho(z) := \langle \text{Im } m(z) \rangle / \pi$  for  $z \in \mathbb{H}$ . If Assumptions 4.5 hold true on  $I$  with  $\eta_* = 1$  then there are  $\varepsilon \sim 1$  and  $\rho_* \sim 1$  such that

- (i) (Inside support around an edge with small gap) Let  $\tau_0, \tau_1 \in \text{supp } \rho \cap I_\theta$  satisfy  $\tau_0 < \tau_1$  and  $(\tau_0, \tau_1) \cap \text{supp } \rho = \emptyset$ . We set  $\Delta := \tau_1 - \tau_0$ . For  $\omega \in [0, \varepsilon]$ , we have

$$\rho(\tau_0 - \omega + i\eta) \sim \rho(\tau_1 + \omega + i\eta) \sim \frac{(\omega + \eta)^{1/2}}{(\Delta + \omega + \eta)^{1/6}}$$

- (ii) (Inside a gap) Let  $\tau_0, \tau_1 \in \text{supp } \rho \cap I_\theta$  satisfy  $\tau_0 < \tau_1$  and  $(\tau_0, \tau_1) \cap \text{supp } \rho = \emptyset$ . We set  $\Delta := \tau_1 - \tau_0$ . Then, for  $\tau \in [\tau_0, \tau_1]$  and  $\eta \in [0, \varepsilon]$ , we have

$$\rho(\tau + i\eta) \sim \frac{\eta}{(\Delta + \eta)^{1/6}} \left( \frac{1}{(\tau_1 - \tau + \eta)^{1/2}} + \frac{1}{(\tau - \tau_0 + \eta)^{1/2}} \right).$$



- (iii) (Around a left edge with large gap) Let  $\tau_0 \in I_\theta \cap \partial \text{supp } \rho$  satisfy  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0 - \delta, \tau_0]$  and some  $\delta \sim 1$ . Then, for  $\omega \in [0, \varepsilon]$ , we have

$$\begin{aligned} \rho(\tau_0 + \omega + i\eta) &\sim (\omega + \eta)^{1/2}, \\ \rho(\tau_0 - \omega + i\eta) &\sim \frac{\eta}{(\omega + \eta)^{1/2}}. \end{aligned}$$

A similar statement holds true for a right edge  $\tau_0$ , i.e., if  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0, \tau_0 + \delta]$  and some  $\delta \sim 1$ .

- (iv) (Close to a local minimum) If  $\tau_0 \in \text{supp } \rho \setminus \partial \text{supp } \rho$  is a local minimum of  $\rho$  such that  $\rho(\tau_0) \leq \rho_*$  then, for all  $\omega \in [-\varepsilon, \varepsilon]$  and  $\eta \in [0, \varepsilon]$ , we have

$$\rho(\tau_0 + \omega + i\eta) \sim \rho(\tau_0) + (|\omega| + \eta)^{1/3}.$$

These scaling relations for  $\rho(z) = \langle \text{Im } m(z) \rangle / \pi$  are proven in the same way as the corresponding ones in Corollary A.1 of [3]. The proof in [3] simply relied on the fact that  $\langle \text{Im } m(z) \rangle$  is the harmonic extension of  $\pi\rho$  to the complex upper half-plane and the behavior of  $\rho$  close to its local minima and thus is applicable equally well in the current situation, due to Theorem 7.2.

### 7.1 SHAPE REGULAR POINTS

In the following definition, we introduce the notion of a *shape regular point* which collects the properties of  $m$  necessary for the proof of Theorem 7.1. Proposition 7.5 below explains how the statements of Theorem 7.1 are transferred to this more general setup. In fact, Lemma 4.8 (ii) and Proposition 6.1 show that, under the assumptions of Theorem 7.1, any point  $\tau_0 \in \text{supp } \rho \cap I$  of sufficiently small density  $\rho(\tau_0)$  is a shape regular point for  $m$  in the sense of Definition 7.4 below. By explicitly spelling out the properties of  $m$  really used in the proof of Theorem 7.1 we made our argument modular because a similar analysis around shape regular points will be applied in later works as well.

This modularity, however, requires to reinterpret the concept of comparison relations. In earlier sections we used the comparison relation  $\sim, \lesssim$  and the  $\mathcal{O}$ -notation introduced in Convention 3.4 to hide irrelevant constants in various estimates that depended only on the model parameters  $c_1, c_2, c_3$  from (3.10),  $k_3$  from (4.16) and  $\theta$  from (7.2), these are also the model parameters in Theorem 7.1. The model parameters in Theorem 7.2 are given by  $k_1, \dots, k_8$  in Assumptions 4.5 and  $\theta$  in the definition of  $I_\theta$ .

The formulation of Definition 7.4 also involves comparison relations instead of carrying constants; in the application these constants depend on the original model parameters. When Proposition 7.5 is proven, the corresponding constants directly depend on the constants in Definition 7.4, hence they also indirectly depend on the original model parameters when we apply it to the proof of Theorem 7.1. Since these dependences are somewhat involved and we do not want to overload the paper with different concepts of comparison relations, for simplicity, for the purpose of Theorem 7.1, the reader may think of

the implicit constants in every  $\sim$ -relation depending only on the original model parameters  $c_1, c_2, c_3, k_3$  and  $\theta$ .

DEFINITION 7.4 (Admissibility for shape analysis, shape regular points). Let  $m$  be the solution of the Dyson equation (2.3) associated to a data pair  $(a, S) \in \mathcal{A}_{\text{sa}} \times \Sigma$ .

(i) Let  $\tau_0 \in \mathbb{R}$ ,  $J \subset \mathbb{R}$  be an open interval with  $0 \in J$ ,  $\Theta: J \rightarrow \mathbb{C}$  and  $r: J \rightarrow \mathcal{A}$  be continuous functions and  $b \in \mathcal{A}$ . We say that  $m$  is  $(J, \Theta, b, r)$ -admissible for the shape analysis at  $\tau_0$  if the following conditions are satisfied:

- (a) The function  $m: \mathbb{H} \rightarrow \mathcal{A}$  has a continuous extension to  $\tau_0 + J$ , which we also denote by  $m$ . The relation (6.1) and the bounds (6.7a) as well as (6.7b) hold true for all  $\omega \in J$ .
- (b) The function  $\Theta$  satisfies the cubic equation (6.3) for all  $\omega \in J$  with the coefficients

$$\begin{aligned}\mu_3 &= \psi + \mathcal{O}(\rho), \\ \mu_2 &= \sigma + i3\psi\rho + \mathcal{O}(\rho^2 + \rho|\sigma|), \\ \mu_1 &= -2\rho^2\psi + i\kappa_1\rho\sigma + \mathcal{O}(\rho^3 + \rho^2|\sigma|), \\ \Xi(\omega) &= \kappa(1 + \nu(\omega)) + \mathcal{O}(\rho),\end{aligned}$$

where  $\rho := \langle \text{Im } m(\tau_0) \rangle / \pi$  and  $\psi, \kappa \geq 0$  as well as  $\sigma, \kappa_1 \in \mathbb{R}$  are some parameters satisfying (6.6) and  $\kappa, |\kappa_1| \sim 1$ . The function  $\nu: J \rightarrow \mathbb{C}$  satisfies (6.5).

- (c) The element  $b \in \mathcal{A}$  in (6.1) fulfils  $b = b^* + \mathcal{O}(\rho)$  and  $b + b^* \sim 1$ .
- (d1) If  $\rho > 0$  then  $\Theta$  and  $r$  are differentiable in  $\omega$  at  $\omega = 0$ .
- (d2) If  $\rho = 0$  then (6.8) holds true for all  $\omega \in J$  and  $\text{Re } \Theta$  is non-decreasing on the connected components of  $\{\omega \in J: \text{Im } \Theta(\omega) = 0\}$ .

(ii) Let  $\tau_0 \in \mathbb{R}$  and  $J \subset \mathbb{R}$  be an open interval with  $0 \in J$ . We say that  $\tau_0$  is a *shape regular point for  $m$  on  $J$*  if  $m$  is  $(J, \Theta, b, r)$ -admissible for the shape analysis at  $\tau_0$  for some continuous functions  $\Theta: J \rightarrow \mathbb{C}$  and  $r: J \rightarrow \mathcal{A}$  as well as  $b \in \mathcal{A}$ .

The key technical step in the proof of Theorem 7.1 is the following Proposition 7.5; it shows that Theorem 7.1 holds under more general weaker conditions, in fact shape admissibility is sufficient. For the proof of Theorem 7.1 we will first check shape regularity from Proposition 6.1 and then we will prove Proposition 7.5; both steps are done in Section 7.4 below.

PROPOSITION 7.5 (Theorem 7.1 under weaker assumptions). *For the solution  $m$  to the Dyson equation (2.3), we write  $v := \pi^{-1} \text{Im } m$ ,  $\rho = \langle v \rangle$ .*

There are thresholds  $\rho_* \sim 1$  and  $\delta_* \sim 1$  such that if  $\rho(\tau_0) \leq \rho_*$  and  $\tau_0 \in \text{supp } \rho$  is a local minimum of  $\rho$  as well as a shape regular point for  $m$  on  $J$  with an open interval  $J \subset \mathbb{R}$  satisfying  $0 \in J$  then (7.3) holds true for all  $\omega \in [-\delta_*, \delta_*] \cap J \cap D$ . Here, as in Theorem 7.1,  $h = h(\tau_0) \in \mathcal{A}$  with  $h \sim 1$  and  $D$  as well as  $\Psi$  depend only on the type of  $\tau_0$  in the following way:

Suppose that  $\tau_0 \in \partial \text{supp } \rho$  is the infimum of a connected component of  $\text{supp } \rho$ . If  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0 - \varepsilon, \tau_0]$  with some  $\varepsilon \sim 1$  (e.g.  $\tau_0 = \inf \text{supp } \rho$ ) and  $|\inf J| \gtrsim 1$ , then the conclusion of case (a) in Theorem 7.1 holds true with  $\Delta = 1$  and  $v(\tau_0) = 0$ .

If  $\tau_0 \neq \inf \text{supp } \rho$  and  $\tau_1 := \sup((-\infty, \tau_0) \cap \text{supp } \rho)$  is a shape regular point for  $m$ ,  $\Delta \lesssim 1$  with  $\Delta := \tau_0 - \tau_1$  and  $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim |\tau_0 - \tau_1|^{1/3}$  then the conclusion of case (a) in Theorem 7.1 holds true with this choice of  $\Delta$  as well as  $v(\tau_0) = 0$ .

Similarly to (a), the statement of case (b) in Theorem 7.1 can be translated to the current setup. The cases (c) and (d) of Theorem 7.1, cusp and internal minimum, respectively, hold true without any changes.

Furthermore, suppose that  $\tau_0 \in \text{supp } \rho$  is a shape regular point for  $m$  and  $\rho(\tau_0) = 0$ , then  $\tau_0$  is a cusp if  $\sigma(\tau_0) = 0$  and  $\tau_0$  is an edge, in particular  $\tau_0 \in \partial \text{supp } \rho$ , if  $\sigma(\tau_0) \neq 0$ .

Similarly, the following Proposition 7.6 is the analogue of Theorem 7.2 under the sole requirement of shape admissibility. Owing to the weaker assumptions, the error term in (7.9) as well as the result in (7.10) of Proposition 7.6 are weaker than the corresponding results in Theorem 7.2. We will first show Proposition 7.6 and then conclude Theorem 7.2 by using extra arguments for the stronger conclusions; both proofs will be presented in Section 7.5 below.

At a shape regular point  $\tau_0 \in \mathbb{R}$ , we set  $\Gamma := \sqrt{27}\kappa/(2\psi)$  (cf. Theorem 7.7 (i) below), where  $\kappa = \kappa(\tau_0)$  and  $\psi = \psi(\tau_0)$  are defined as in Definition 7.4 (i) (b).

PROPOSITION 7.6 (Behaviour of  $\rho$  near almost cusp points, set of minima of  $\rho$  under weaker assumptions). *Let  $m$  be the solution to the Dyson equation, (2.3), and  $\rho = \pi^{-1}\langle \text{Im } m \rangle$ . The following statements hold true*

(i) *There are thresholds  $\rho_* \sim 1$ ,  $\sigma_* \sim 1$  and  $\delta_* \sim 1$  such that if  $\tau_0 \in \text{supp } \rho$  is a shape regular point for  $m$  on an open interval  $J \subset \mathbb{R}$  with  $0 \in J$ ,  $\rho(\tau_0) \leq \rho_*$  and  $\tau_0$  is a local minimum of  $\rho$  then we have*

(a) *(Left edge with small gap) If  $\tau_0 \in \partial \text{supp } \rho \setminus \{\inf \text{supp } \rho\}$  is the infimum of a connected component of  $\text{supp } \rho$ ,  $|\sigma(\tau_0)| \leq \sigma_*$  and  $\tau_1 := \sup((-\infty, \tau_0) \cap \text{supp } \rho)$  is a shape regular point satisfying  $\Delta \lesssim 1$  for  $\Delta := \tau_0 - \tau_1$  and  $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim |\tau_0 - \tau_1|^{1/3}$  then (7.5a) for all  $\omega \in [0, \delta_*] \cap J$ .*

(b) *(Right edge with small gap) If  $\tau_0 \in \partial \text{supp } \rho \setminus \{\sup \text{supp } \rho\}$  is the supremum of a connected component then a similar statement as in the case of a left edge holds true.*

(c) (Cusp) If  $\tau_0 \notin \partial \text{supp } \rho$  and  $\rho(\tau_0) = 0$  then (7.5b) holds true for all  $\omega \in [-\delta_*, \delta_*] \cap J$ .

(d) (Internal minimum) If  $\tau_0 \notin \partial \text{supp } \rho$  and  $\rho(\tau_0) > 0$  then

$$\begin{aligned} \rho(\tau_0 + \omega) &= \rho(\tau_0) + \Gamma^{1/3} \Psi(\omega) + \mathcal{O} \left( \frac{|\omega|}{\rho(\tau_0)} \mathbf{1}(|\omega| \lesssim \rho(\tau_0)^3) + \Psi(\omega)^2 \right), \\ \Psi(\omega) &:= \tilde{\rho} \Psi_{\min} \left( \frac{\omega}{\tilde{\rho}^3} \right), \quad \tilde{\rho} := \frac{\rho(\tau_0)}{\Gamma^{1/3}} \end{aligned} \tag{7.9}$$

for all  $\omega \in [-\delta_*, \delta_*] \cap J$ .

(ii) Let  $I \subset \mathbb{R}$  be an open interval with  $\text{supp } \rho \cap I \neq \emptyset$  and  $|I| \lesssim 1$  and let  $m$  have a continuous extension to the closure  $\bar{I}$  of  $I$ . Let  $J \subset \mathbb{R}$  be an open interval with  $0 \in J$  and  $\text{dist}(0, \partial J) \gtrsim 1$  such that  $J + (\partial \text{supp } \rho) \cap I \subset I$ . We assume that all points in  $(\partial \text{supp } \rho) \cap I$  are shape regular points for  $m$  on  $J$  and all estimates in Definition 7.4 hold true uniformly on  $(\partial \text{supp } \rho) \cap I$ . If  $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim |\tau_0 - \tau_1|^{1/3}$  uniformly for all  $\tau_0, \tau_1 \in (\partial \text{supp } \rho) \cap I$  then  $\text{supp } \rho \cap I$  consists of  $K \sim 1$  intervals, i.e., there are  $\alpha_1, \dots, \alpha_K \in \partial \text{supp } \rho \cup \partial I$  and  $\beta_1, \dots, \beta_K \in \partial \text{supp } \rho \cup \partial I$ ,  $\alpha_i < \beta_i < \alpha_{i+1}$ , such that (7.6) holds true with  $I_\theta$  replaced by  $\bar{I}$  and  $\beta_i - \alpha_i \sim 1$  if  $\beta_i \neq \text{sup } I$  and  $\alpha_i \neq \text{inf } I$ .

If  $\mathbb{M}_{\rho_*}$  is defined as in (7.7) then there is a threshold  $\rho_* \sim 1$  such that if, in addition to the previous conditions in (ii), all points of  $(\mathbb{M}_{\rho_*} \cup \partial \text{supp } \rho) \cap I$  are shape regular points for  $m$  on  $J$  and all estimates in Definition 7.4 hold true uniformly on  $(\mathbb{M}_{\rho_*} \cup \partial \text{supp } \rho) \cap I$  then, for  $\gamma \in \mathbb{M}_{\rho_*}$ , we have  $|\alpha_i - \gamma| \sim 1$  and  $|\beta_i - \gamma| \sim 1$  if  $\alpha_i \neq \text{inf } I$  and  $\beta_i \neq \text{sup } I$ . Moreover, for any  $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$ , we have either

$$|\gamma_1 - \gamma_2| \sim 1, \quad \text{or} \quad |\gamma_1 - \gamma_2| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4. \tag{7.10}$$

If  $\rho(\gamma_1) = 0$  or  $\rho(\gamma_2) = 0$  then, for  $\gamma_1 \neq \gamma_2$ , only the first case occurs.

An important step towards Theorem 7.1 and Proposition 7.5 will be to prove similar behaviours for  $\Theta$  as  $\text{Im } \Theta$  is the leading term in  $v$ . These behaviours are collected in the following theorem, Theorem 7.7. It has weaker assumptions than those of Theorem 7.1 and those required in Proposition 7.5 – in particular, on the coefficient  $\mu_1$  in the cubic equation (6.3). However, these assumptions will be sufficient for the purpose of Theorem 7.7.

**THEOREM 7.7** (Abstract cubic equation). *Let  $\Theta(\omega)$  be a continuous solution to the cubic equation*

$$\mu_3 \Theta(\omega)^3 + \mu_2 \Theta(\omega)^2 + \mu_1 \Theta(\omega) + \omega \Xi(\omega) = 0 \tag{7.11}$$

for  $\omega \in J$ , where  $J \subset \mathbb{R}$  is an open interval with  $0 \in J$ . We assume that the

coefficients satisfy

$$\begin{aligned} \mu_3 &= \psi + \mathcal{O}(\rho), \\ \mu_2 &= \sigma + 3i\psi\rho + \mathcal{O}(\rho^2 + \rho|\sigma|), \\ \mu_1 &= -2\rho^2\psi + \mathcal{O}(\rho^3 + \rho|\sigma|), \\ \Xi(\omega) &= \kappa(1 + \nu(\omega)) + \mathcal{O}(\rho) \end{aligned}$$

with some fixed parameters  $\psi \geq 0, \rho \geq 0, \sigma \in \mathbb{R}$  and  $\kappa \sim 1$ . The cubic equation is assumed to be stable in the sense that

$$\psi + \sigma^2 \sim 1. \tag{7.12}$$

Moreover, for all  $\omega \in J$ , we require the following bounds on  $\nu$  and  $\Theta$ :

$$|\nu(\omega)| \lesssim |\omega|^{1/3}, \tag{7.13a}$$

$$|\Theta(\omega)| \lesssim |\omega|^{1/3}. \tag{7.13b}$$

Then the following statements hold true:

- (i) ( $\rho > 0$ ) For any  $\Pi_* \sim 1$ , there is a threshold  $\rho_* \sim 1$  such that if  $\rho \in (0, \rho_*]$  and  $|\sigma| \leq \Pi_*\rho^2$  then we have

$$\text{Im } \Theta(\omega) = \rho\Psi_{\min}\left(\Gamma\frac{\omega}{\rho^3}\right) + \mathcal{O}(\min\{\rho^{-1}|\omega|, |\omega|^{2/3}\}), \tag{7.14}$$

with  $\Gamma := \sqrt{27}\kappa/(2\psi)$ . Note that  $\Gamma \sim 1$  if  $\rho_* \sim 1$  is small enough.

- (ii) ( $\rho = 0$ ) If  $\rho = 0$  and we additionally assume  $\text{Im } \Theta(\omega) \geq 0$  for  $\omega \in J$ ,  $\text{Re } \Theta$  is non-decreasing on the connected components of  $\{\omega \in J : \text{Im } \Theta(\omega) = 0\}$  as well as

$$|\text{Im } \nu(\omega)| \lesssim \text{Im } \Theta(\omega) \tag{7.15}$$

for all  $\omega \in J$  then we have

- (a) If  $\sigma = 0$  then  $\text{Im } \Theta(\omega)$  has a cubic cusp at  $\omega = 0$ , i.e.,

$$\text{Im } \Theta(\omega) = \frac{\sqrt{3}}{2}\left(\frac{\kappa}{\psi}\right)^{1/3}|\omega|^{1/3} + \mathcal{O}(|\omega|^{2/3}). \tag{7.16}$$

- (b) If  $\sigma \neq 0$  then  $\text{Im } \Theta(\omega)$  has a square root edge at  $\omega = 0$ , i.e., there is  $c_* \sim 1$  such that

$$\text{Im } \Theta(\omega) = \begin{cases} c\hat{\Delta}^{1/3}\Psi_{\text{edge}}\left(\frac{|\omega|}{\hat{\Delta}}\right) + \mathcal{O}((|\nu(\omega)| + \varepsilon(\omega))\varepsilon(\omega)), & \text{if } \text{sign } \omega = \text{sign } \sigma, \\ 0, & \text{if } \omega \in -\text{sign } \sigma[0, c_*|\sigma|^3], \end{cases} \tag{7.17}$$

where  $\widehat{\Delta} \in (0, \infty)$ ,  $c \in (0, \infty)$  and  $\varepsilon: \mathbb{R} \rightarrow [0, \infty)$  are defined by

$$\begin{aligned} \widehat{\Delta} &:= \min \left\{ \frac{4}{27\kappa} \frac{|\sigma|^3}{\psi^2}, 1 \right\}, & c &:= 3\sqrt{\kappa} \frac{\widehat{\Delta}^{1/6}}{|\sigma|^{1/2}}, \\ \varepsilon(\omega) &:= \min \left\{ \frac{|\omega|^{1/2}}{\widehat{\Delta}^{1/6}}, |\omega|^{1/3} \right\}. \end{aligned} \quad (7.18)$$

We have  $\widehat{\Delta} \sim |\sigma|^3$  and  $c \sim 1$ . Moreover, for  $\text{sign } \omega = \text{sign } \sigma$ , we have

$$|\Theta(\omega)| \lesssim \varepsilon(\omega). \quad (7.19)$$

## 7.2 CUBIC EQUATIONS IN NORMAL FORM

The core of the proof of Theorem 7.7 is to bring (7.11) into a normal form by a change of variables. We will first explain the analysis of these normal forms, especially the mechanism of choosing the right branch of the solution based upon selection principles that will be derived from the constraints on  $\Theta$  given in Theorem 7.7. Then, in Section 7.3, we show how to bring (7.11) to these normal forms.

In the following proposition, we study a special solution  $\Omega(\lambda)$  to a one-parameter family of cubic equations in normal forms with constant term  $\Lambda(\lambda)$  (or  $2\Lambda(\lambda)$ ), where  $\Lambda(\lambda)$  is a perturbation of the identity map  $\lambda \mapsto \lambda$ . Here, a-priori, the real parameter  $\lambda$  is always contained in an (possibly unbounded) interval around 0. This range of definition will not be explicitly indicated in the statements but will be explicitly restricted for their conclusions. We compare the solution to this perturbed cubic equation with the solution to the cubic equation with constant term  $\lambda$ . Depending on the precise type of the cubic equation, the choice of the solution is based on some of the following *selection principles*

SP1  $\lambda \mapsto \Omega(\lambda)$  is continuous

SP2  $\Omega(0) = \Omega_0$  for some given  $\Omega_0 \in \mathbb{C}$

SP3  $\text{Im}(\Omega(\lambda) - \Omega(0)) \geq 0$ ,

SP4'  $|\text{Im } \Lambda(\lambda)| \leq \gamma |\lambda| |\text{Im } \Omega(\lambda)|$  for some  $\gamma > 0$  and  $\text{Re } \Omega(\lambda)$  is non-decreasing on the connected components of  $\{\lambda: \text{Im } \Omega(\lambda) = 0\}$ .

We use the notation SP4' to distinguish this selection principle from SP-4 which was introduced in Lemma 9.9 of [3].

We will make use of the following standard convention for complex powers.

**DEFINITION 7.8** (Complex powers). We define  $\mathbb{C} \setminus (-\infty, 0) \rightarrow \mathbb{C}$ ,  $\zeta \mapsto \zeta^\gamma$  for  $\gamma \in \mathbb{C}$  by  $\zeta^\gamma := \exp(\gamma \log \zeta)$ , where  $\log: \mathbb{C} \setminus (-\infty, 0) \rightarrow \mathbb{C}$  is a continuous branch of the complex logarithm with  $\log 1 = 0$ .

With this convention, we record Cardano's formula as follows:

PROPOSITION 7.9 (Cardano). *The three roots of  $\Omega^3 - 3\Omega + 2\zeta$ ,  $\zeta \in \mathbb{C}$ , are  $\widehat{\Omega}_+(\zeta)$ ,  $\widehat{\Omega}_-(\zeta)$  and  $\widehat{\Omega}_0(\zeta)$  which are defined by*

$$\begin{aligned} \widehat{\Omega}_\pm(\zeta) &:= \frac{1}{2}(\Phi_+(\zeta) + \Phi_-(\zeta)) \pm \frac{i\sqrt{3}}{2}(\Phi_+(\zeta) - \Phi_-(\zeta)), \\ \widehat{\Omega}_0(\zeta) &:= -(\Phi_+(\zeta) + \Phi_-(\zeta)), \end{aligned} \tag{7.20}$$

where

$$\Phi_\pm(\zeta) = \begin{cases} (\zeta \pm \sqrt{\zeta^2 - 1})^{1/3}, & \text{if } \operatorname{Re} \zeta \geq 1, \\ (\zeta \pm i\sqrt{1 - \zeta^2})^{1/3}, & \text{if } |\operatorname{Re} \zeta| < 1, \\ -(-\zeta \mp \sqrt{\zeta^2 - 1})^{1/3}, & \text{if } \operatorname{Re} \zeta \leq -1. \end{cases}$$

PROPOSITION 7.10 (Solution to the cubic in normal form). *Let  $\Omega(\lambda)$  satisfy SP1 and SP2.*

(i) (Non-zero local minimum) *Let  $\Omega_0 = \sqrt{3}(i + \chi_1)$  in SP2 and  $\Omega(\lambda)$  satisfy*

$$\Omega(\lambda)^3 + 3\Omega(\lambda) + 2\Lambda(\lambda) = 0, \quad \Lambda(\lambda) = (1 + \chi_2 + \mu(\lambda))\lambda + \chi_3, \tag{7.21}$$

with  $|\mu(\lambda)| \lesssim \alpha|\lambda|^{1/3}$ ,  $\alpha > 0$ . *Then there exist  $\delta \sim 1$  and  $\chi_* \sim 1$  such that if  $\alpha, |\chi_1|, |\chi_2|, |\chi_3| \leq \chi_*$  then*

$$\Omega(\lambda) - \Omega_0 = \widehat{\Omega}(\lambda) - i\sqrt{3} + \mathcal{O}((\alpha + |\chi_2| + |\chi_3|) \min\{|\lambda|, |\lambda|^{2/3}\}) \tag{7.22}$$

for all  $\lambda \in \mathbb{R}$  satisfying  $|\lambda| \leq \delta/\alpha^3$ , where  $\widehat{\Omega}(\lambda) := \Phi_{\text{odd}}(\lambda) + i\sqrt{3}\Phi_{\text{even}}(\lambda)$  and  $\Phi_{\text{odd}}$  and  $\Phi_{\text{even}}$  are the odd and even part of the function  $\Phi: \mathbb{C} \rightarrow \mathbb{C}$ ,  $\Phi(\zeta) := (\sqrt{1 + \zeta^2} + \zeta)^{1/3}$ , respectively.

Moreover, we have for  $|\lambda| \leq \delta/\alpha^3$  that

$$|\Omega(\lambda) - \Omega_0| \lesssim \min\{|\lambda|, |\lambda|^{1/3}\}. \tag{7.23}$$

In the following, we assume that  $\Omega(\lambda)$ , in addition to SP1 and SP2, also satisfies SP3 and SP4'.

(ii) (Simple edge) *Let  $\Omega_0 = 0$  in SP2 and  $\Omega(\lambda)$  be a solution to*

$$\Omega^2(\lambda) + \Lambda(\lambda) = 0, \quad \Lambda(\lambda) = (1 + \mu(\lambda))\lambda. \tag{7.24}$$

If  $|\mu(\lambda)| \leq \gamma^{2/3}|\lambda|^{1/3}$  for the  $\gamma > 0$  of SP4' then there is  $c_* \sim 1$  such that

$$\begin{aligned} \Omega(\lambda) &= \widehat{\Omega}(\lambda) + \mathcal{O}(|\mu(\lambda)||\lambda|^{1/2}), \\ \widehat{\Omega}(\lambda) &:= \begin{cases} i\lambda^{1/2}, & \text{if } \lambda \in [0, c_*\gamma^{-2}], \\ -(-\lambda)^{1/2}, & \text{if } \lambda \in [-c_*\gamma^{-2}, 0], \end{cases} \end{aligned} \tag{7.25}$$

and  $\operatorname{Im} \Omega(\lambda) = 0$  as well as  $\operatorname{Im} \mu(\lambda) = 0$  for all  $\lambda \in [-c_*\gamma^{-2}, 0]$ .

(iii) (Sharp cusp) Let  $\Omega_0 = 0$  in  $\text{SP}2$ ,  $\gamma \sim 1$  in  $\text{SP}4$  and  $\Omega(\lambda)$  be a solution to

$$\Omega^3(\lambda) + \Lambda(\lambda) = 0, \quad \Lambda(\lambda) = (1 + \mu(\lambda))\lambda. \tag{7.26}$$

If  $|\mu(\lambda)| \lesssim |\lambda|^{1/3}$  then there is  $\delta \sim 1$  such that

$$\begin{aligned} \Omega(\lambda) &= \widehat{\Omega}(\lambda) + \mathcal{O}(|\mu(\lambda)||\lambda|^{1/3}), \\ \widehat{\Omega}(\lambda) &:= \frac{1}{2} \begin{cases} (-1 + i\sqrt{3})\lambda^{1/3}, & \text{if } \lambda \in (0, \delta], \\ (1 + i\sqrt{3})|\lambda|^{1/3}, & \text{if } \lambda \in [-\delta, 0]. \end{cases} \end{aligned} \tag{7.27}$$

(iv) (Two nearby edges) Let  $\Omega_0 = s$  for some  $s \in \{\pm 1\}$  in  $\text{SP}2$ ,  $\gamma \sim 1$  in  $\text{SP}4$  and  $\Omega(\lambda)$  be a solution to

$$\Omega(\lambda)^3 - 3\Omega(\lambda) + 2\Lambda(\lambda) = 0, \quad \Lambda(\lambda) = (1 + \mu(\lambda))\lambda + s. \tag{7.28}$$

Then there are  $\delta \sim 1$ ,  $\varrho \sim 1$  and  $\gamma_* \sim 1$  such that if  $|\mu(\lambda)| \lesssim \widehat{\gamma}|\lambda|^{1/3}$  for some  $\widehat{\gamma} \in [0, \gamma_*]$  then

(a) We have

$$\Omega(\lambda) = \widehat{\Omega}_+(1 + |\lambda|) + \mathcal{O}(|\mu(\lambda)| \min\{|\lambda|^{1/2}, |\lambda|^{1/3}\}), \tag{7.29}$$

for all  $\lambda \in s(0, 2\delta/\widehat{\gamma}^3]$ . (Recall the definition of  $\widehat{\Omega}_+$  from (7.20).) Moreover, for all  $\lambda \in s(0, 2\delta/\widehat{\gamma}^3]$ , we have

$$|\Omega(\lambda) - \Omega_0| \lesssim \min\{|\lambda|^{1/2}, |\lambda|^{1/3}\}. \tag{7.30}$$

(b) For all  $\lambda \in -s(0, 2 - \varrho\widehat{\gamma}]$ , we have

$$\text{Im } \Omega(\lambda) \lesssim \widehat{\gamma}^{1/2}. \tag{7.31}$$

(c) We have

$$\text{Im } \Omega(-s(2 + \varrho\widehat{\gamma})) > 0. \tag{7.32}$$

The core of each part in Proposition 7.10 is choosing the correct cubic root. For the most complicated part (iv), we state this choice in the following auxiliary lemma. For its formulation, we introduce the intervals

$$I_1 := -s[-\lambda_1, 0), \quad I_2 := -s(0, \lambda_2], \quad I_3 := -s[\lambda_3, \lambda_1], \tag{7.33}$$

where we used the definitions

$$\lambda_1 := 2\frac{\delta}{\widehat{\gamma}^3}, \quad \lambda_2 := 2 - \varrho\widehat{\gamma}, \quad \lambda_3 := 2 + \varrho\widehat{\gamma}. \tag{7.34}$$

These definitions are modelled after (9.105) in [3]. We will choose  $\widehat{\gamma} = \widehat{\Delta}^{1/3}$  in the proof of Theorem 7.7 below. Then  $\lambda_1$  corresponds to an expansion range  $\delta$  in the  $\omega$  coordinate. Note that with the above choice of  $\widehat{\gamma}$ , we obtain the same



$\lambda_1$  as in (9.105) of [3]. However,  $\lambda_2$  and  $\lambda_3$  differ slightly from those in [3], where  $\lambda_{2,3}$  were set to be  $2 \mp \varrho|\sigma|$ . Nevertheless, we will see below that  $\widehat{\gamma} \sim |\sigma|$  but they are not equal in general.

For given  $\delta, \varrho \sim 1$ , we will always choose  $\gamma_* \sim 1$  so small that  $\widehat{\gamma} \leq \gamma_*$  implies

$$\lambda_1 \geq 4, \quad 1 \leq \lambda_2 < 2 < \lambda_3 \leq 3.$$

Therefore, the intervals in (7.33) are disjoint and nonempty.

LEMMA 7.11 (Choice of cubic roots in Proposition 7.10 (iv)). *Under the assumptions of Proposition 7.10 (iv), there are  $\delta, \varrho, \gamma_* \sim 1$  such that if  $\widehat{\gamma} \leq \gamma_*$  then we have*

$$\Omega|_{I_k} = \widehat{\Omega}_+ \circ \Lambda|_{I_k}$$

for  $k = 1, 2, 3$ . Here,  $\widehat{\Omega}_+$  is defined as in (7.20).

*Proof.* The proof is the same as the one of Lemma 9.14 in [3] but SP-4 in [3] is replaced by SP4' above. In that proof, SP-4 is used only in the part titled ‘‘Choice of  $a_2$ ’’. We redo this part here. Recall that  $a_2 = 0, \pm$  denoted the index such that  $\Omega|_{I_2} = \widehat{\Omega}_{a_2} \circ \Lambda|_{I_2}$  and our goal is to show  $a_2 = +$ . Similarly as in [3], we assume without loss of generality  $s = -1$ . Since  $\lim_{\lambda \downarrow -1} \widehat{\Omega}_-(\lambda) = 2$  and  $\Omega(0) = -1$  by SP2, we conclude  $a_2 \neq -$ .

We now prove  $a_2 \neq 0$ . To that end, we take the imaginary part of the cubic equation, (7.28), and obtain

$$3((\operatorname{Re} \Omega)^2 - 1)\operatorname{Im} \Omega = -2\lambda \operatorname{Im} \mu(\lambda) + (\operatorname{Im} \Omega)^3. \tag{7.35}$$

Suppose that  $a_2 = 0$ . From the definition of  $\widehat{\Omega}_0$ ,  $\Lambda(\lambda) = (1 + \mu(\lambda))\lambda - 1$  and  $|\mu(\lambda)| \lesssim \widehat{\gamma}|\lambda|^{1/3}$  we obtain

$$\operatorname{Re} \widehat{\Omega}_0(\Lambda(\lambda)) \leq -1 - c|\lambda|^{1/2} + C\widehat{\gamma}^{1/2}\lambda^{2/3}, \quad |\operatorname{Im} \widehat{\Omega}_0(\Lambda(\lambda))| \lesssim \widehat{\gamma}^{1/2}\lambda^{2/3}, \tag{7.36}$$

(compare (9.120) in [3]). Thus, from (7.35), we conclude

$$|\lambda|^{1/2}\operatorname{Im} \Omega \lesssim |\lambda|\operatorname{Im} \Omega$$

for small  $\lambda$  as  $|\operatorname{Im} \mu(\lambda)| \lesssim \operatorname{Im} \Omega$  by SP4' and  $|\operatorname{Im} \Lambda| = |\lambda|\operatorname{Im} \mu$ . Hence,  $\operatorname{Im} \Omega(\lambda) = 0$  for small enough  $|\lambda|$ . Thus,  $\operatorname{Re} \Omega$  is non-decreasing for such  $\lambda$  by SP4', but from  $\Omega(0) = -1$  and the first bound in (7.36) we conclude that  $\operatorname{Re} \Omega$  has to be decreasing if  $\Omega(\lambda) = \widehat{\Omega}_0(\Lambda(\lambda))$ . This contradiction shows  $a_2 \neq 0$ , hence,  $a_2 = +$ . The rest of the proof in [3] is unchanged.  $\square$

*Proof of Proposition 7.10.* For the proof of (i), we mainly follow the proof of Proposition 9.3 in [3] with  $\gamma_4 = \chi_1$ ,  $\gamma_5 = \chi_2$  and  $\gamma_6 = \chi_3$  in (9.35) and (9.37) of [3].

Following the careful selection of the correct solution of (7.21) (cf. (9.36) in [3]) by the selection principles till above (9.50) in [3] yields  $\Omega(\lambda) = \widehat{\Omega}(\Lambda(\lambda))$  and hence, in particular,  $\widehat{\Omega}(\chi_3) = \Omega_0 = \sqrt{3}(i + \chi_1)$ . ( $\widehat{\Omega} = \widehat{\Omega}_+$  in [3].) By defining

$$\Lambda_0(\lambda) := (1 + \chi_2 + \mu(\lambda))\lambda$$

and using  $|\mu(\lambda)| \lesssim \alpha|\lambda|^{1/3}$  instead of (9.54) in [3], we obtain

$$\begin{aligned} \widehat{\Omega}(\Lambda_0(\lambda)) - \widehat{\Omega}(0) &= \widehat{\Omega}(\lambda) - \widehat{\Omega}(0) + \mathcal{O}\left(\frac{(|\chi_2| + |\mu(\lambda)|)|\lambda|}{1 + |\lambda|^{2/3}}\right) \\ &= \widehat{\Omega}(\lambda) - \widehat{\Omega}(0) + \mathcal{O}((\alpha + |\chi_2|) \min\{|\lambda|, |\lambda|^{2/3}\}) \end{aligned}$$

instead of (9.56) in [3]. Thus, (9.57) in the proof of Proposition 9.3 in [3] yields

$$\widehat{\Omega}(\chi_3 + \Lambda_0(\lambda)) - \widehat{\Omega}(\chi_3) = \widehat{\Omega}(\lambda) - \widehat{\Omega}(0) + \mathcal{O}((\alpha + |\chi_2| + |\chi_3|) \min\{|\lambda|, |\lambda|^{2/3}\}).$$

Thus, we obtain (7.22) since  $\widehat{\Omega}(\chi_3) = \Omega_0$  and  $\widehat{\Omega}(0) = i\sqrt{3}$ . We remark that (7.23) is exactly (9.53) in [3].

The proof of (ii) resembles the proof of Lemma 9.11 in [3] but we replace assumption SP-4 of [3] by SP4'. Since  $\Omega(\lambda)$  solves (7.24), there is a function  $A: \mathbb{R} \rightarrow \{\pm\}$  such that  $\Omega(\lambda) = \widetilde{\Omega}_{A(\lambda)}(\Lambda(\lambda))$  for all  $\lambda \in \mathbb{R}$ . Here,  $\widetilde{\Omega}_{\pm}: \mathbb{C} \rightarrow \mathbb{C}$  denote the functions

$$\widetilde{\Omega}_{\pm}(\zeta) := \pm \begin{cases} i\zeta^{1/2}, & \text{if } \operatorname{Re} \zeta \geq 0, \\ -(-\zeta)^{1/2}, & \text{if } \operatorname{Re} \zeta < 0. \end{cases}$$

(Note that they were denoted by  $\widehat{\Omega}_{\pm}$  in (9.78) of [3]). By assumption, there is  $c_* \sim 1$  such that  $|\mu(\lambda)| < 1$  for all  $|\lambda| \leq c_*\gamma^{-2}$ . Hence, by SP1, we find  $a_+, a_- \in \{\pm\}$  such that  $A(\lambda) = a_{\pm}$  for  $\lambda \in \pm[0, c_*\gamma^{-2}]$ .

For  $\lambda \geq 0$ , we have

$$\operatorname{Im} \widetilde{\Omega}_-(\Lambda(\lambda)) = -\lambda^{1/2} + \mathcal{O}(\mu(\lambda)\lambda^{1/2}).$$

Thus, possibly shrinking  $c_* \sim 1$ , we obtain  $\operatorname{Im} \widetilde{\Omega}_-(\Lambda(\lambda)) < 0$  for  $\lambda \in (0, c_*\gamma^{-2}]$ . Therefore, the choice  $a_+ = -$  would contradict SP3 and we conclude  $a_+ = +$ . We now prove that  $a_- = +$ . Assume to the contrary that  $a_- = -$ . For small enough  $c_* \sim 1$ , we have

$$\begin{aligned} \operatorname{Re} \widetilde{\Omega}_-(\Lambda(\lambda)) &= |\lambda|^{1/2} \operatorname{Re} (1 + \mu(\lambda))^{1/2} \sim |\lambda|^{1/2}, \\ \operatorname{Im} \widetilde{\Omega}_-(\Lambda(\lambda)) &= |\lambda|^{1/2} \operatorname{Im} ((1 + \mu(\lambda))^{1/2}) \lesssim |\lambda|^{1/2} \end{aligned}$$

for  $\lambda \in [-c_*\gamma^{-2}, 0)$  by the definition of  $\widetilde{\Omega}_-$  and  $\Lambda$ . Hence, taking the imaginary part of (7.24) and using SP4' yield

$$|\lambda|^{1/2} \operatorname{Im} \Omega(\lambda) \lesssim \gamma|\lambda| \operatorname{Im} \Omega(\lambda)$$

for  $\lambda \in [-c_*\gamma^{-2}, 0)$ . By possibly shrinking  $c_* \sim 1$ , we obtain  $\operatorname{Im} \Omega(\lambda) = 0$  for  $\lambda \in [-c_*\gamma^{-2}, 0)$ . Hence, taking the imaginary part of (7.24) also yields  $\operatorname{Im} \mu(\lambda) = 0$  for  $\lambda \in [-c_*\gamma^{-2}, 0)$ . Thus, SP4' implies that  $\operatorname{Re} \Omega$  is non-decreasing on  $[-c_*\gamma^{-2}, 0)$  which contradicts  $\operatorname{Re} \widetilde{\Omega}_-(0) = 0$  and  $\operatorname{Re} \widetilde{\Omega}_-(\Lambda(\lambda)) \sim |\lambda|^{1/2} > 0$  for  $\lambda \in [-c_*\gamma^{-2}, 0)$  with small enough  $c_* \sim 1$ . Hence,  $a_- = +$  which

completes the selection of the main term  $\widehat{\Omega} = \widetilde{\Omega}_+$  in (7.25). The error term in (7.25) follows by estimating  $\widehat{\Omega}(\Lambda(\lambda))$  directly.

For the proof of (iii), we select the correct root of (7.26) as in the proof of Lemma 9.12 in [3] under SP4' instead of SP-4. Since  $\Omega(\lambda)$  solves (7.26) there is a function  $A: \mathbb{R} \rightarrow \{0, \pm\}$  such that

$$\Omega(\lambda) = \widetilde{\Omega}_{A(\lambda)}(\Lambda(\lambda))$$

for all  $\lambda \in \mathbb{R}$ . Here, we introduced the functions  $\widetilde{\Omega}_a: \mathbb{C} \rightarrow \mathbb{C}$ ,  $a = 0, \pm$ , defined by

$$\widetilde{\Omega}_0 := - \begin{cases} \zeta^{1/3}, & \text{if } \operatorname{Re} \zeta \geq 0, \\ -(-\zeta)^{1/3}, & \text{if } \operatorname{Re} \zeta < 0, \end{cases} \quad \widetilde{\Omega}_{\pm}(\zeta) := \frac{1 \mp i\sqrt{3}}{2} \widetilde{\Omega}_0(\zeta).$$

(Note that they were denoted by  $\widehat{\Omega}_a$ ,  $a \in \{0, \pm\}$ , in (9.87) of [3].) By SP1,  $A$  can only change its value at  $\lambda$  if  $\Lambda(\lambda) = 0$ . By choosing  $\delta \sim 1$  small enough and using  $|\mu(\lambda)| \lesssim |\lambda|^{1/3}$ , we have  $A(\lambda) = a_+$  and  $A(-\lambda) = a_-$  for some constants  $a_{\pm}$  and for all  $\lambda \in (0, \delta]$ .

We will now use SP3 and SP4' to determine the value of  $a_+$  and  $a_-$ . As in (9.91) of the proof of Lemma 9.12 in [3], we have

$$\pm(\operatorname{sign} \lambda) \operatorname{Im} \widetilde{\Omega}_{\pm}(\Lambda(\lambda)) = \frac{\sqrt{3}}{2} |\lambda|^{1/3} + \mathcal{O}(\mu(\lambda)\lambda^{1/3}) \geq |\lambda|^{1/3} - C|\lambda|^{2/3}.$$

By possibly shrinking  $\delta \sim 1$ , we conclude  $\operatorname{Im} \widetilde{\Omega}_-(\Lambda(\lambda)) < 0$  for  $\lambda \in (0, \delta]$  and  $\operatorname{Im} \widetilde{\Omega}_+(\Lambda(\lambda)) < 0$  for  $\lambda \in [-\delta, 0)$ . Hence, owing to SP3, we conclude  $a_+ \neq -$  and  $a_- \neq +$ .

Next, we will prove  $a_+ \neq 0$ . For  $\lambda \geq 0$ , we have

$$\operatorname{Re} \widetilde{\Omega}_0(\Lambda(\lambda)) \leq -\lambda^{1/3} + C\lambda^{2/3}, \quad \operatorname{Im} \widetilde{\Omega}_0(\Lambda(\lambda)) \lesssim \lambda^{2/3}.$$

Thus, assuming  $\Omega(\lambda) = \widetilde{\Omega}_0(\Lambda(\lambda))$  and estimating the imaginary part of (7.26) yield

$$\lambda^{2/3} \operatorname{Im} \Omega(\lambda) \lesssim (\operatorname{Im} \Omega(\lambda))^3 + |\operatorname{Im} \Lambda(\lambda)| \lesssim |\lambda| \operatorname{Im} \Omega(\lambda).$$

Hence, we possibly shrink  $\delta \sim 1$  and conclude  $\operatorname{Im} \Omega(\lambda) = 0$  for  $\lambda \in [0, \delta]$ . Therefore,  $\operatorname{Re} \Omega(\lambda)$  is non-decreasing on  $[0, \delta]$  by SP4'. Combined with  $\Omega_0 = 0$  and  $\operatorname{Re} \widetilde{\Omega}_0(\Lambda(\lambda)) \lesssim -\lambda^{1/3}$ , we obtain a contradiction. Hence, this implies  $a_+ \neq 0$ , i.e.,  $a_+ = +$ .

A similar argument excludes  $a_- = 0$  and we thus obtain  $a_- = -$ . Now, (7.27) is obtained from the definition of  $\widehat{\Omega} = \widetilde{\Omega}_+$ , which completes the proof of (iii).

For the proof of (iv), we remark that all estimates follow from Lemma 7.11 in the same way as they followed in [3] from Lemma 9.14 in [3]. Indeed, (7.29) is the same as (9.129) in [3]. The bound (7.30) is shown analogously to (9.129) and (9.130) in [3]. Moreover, (7.31) is (9.137) in [3] and (7.32) is obtained as (9.109) in [3]. This completes the proof of Proposition 7.10.  $\square$

## 7.3 PROOF OF THEOREM 7.7

Before we prove Theorem 7.7, we collect some properties of  $\Psi_{\text{edge}}$  and  $\Psi_{\text{min}}$  which will be useful in the following. We recall that  $\Psi_{\text{edge}}$  and  $\Psi_{\text{min}}$  were defined in (7.1).

LEMMA 7.12 (Properties of  $\Psi_{\text{min}}$  and  $\Psi_{\text{edge}}$ ). (i) Let  $\widehat{\Omega}$  be defined as in Proposition 7.10 (i). Then, for any  $\lambda \in \mathbb{R}$ , we have

$$\Psi_{\text{min}}(\lambda) = \frac{1}{\sqrt{3}} \text{Im} [\widehat{\Omega}(\lambda) - \widehat{\Omega}(0)]. \quad (7.37)$$

(ii) Let  $\widehat{\Omega}_+$  be defined as in (7.20). Then, for any  $\lambda \geq 0$ , we have

$$\Psi_{\text{edge}}(\lambda) = \frac{1}{2\sqrt{3}} \text{Im} \widehat{\Omega}_+(1 + 2\lambda). \quad (7.38)$$

(iii) There is a function  $\widetilde{\Psi}: [0, \infty) \rightarrow \mathbb{R}$  with uniformly bounded derivatives and  $\widetilde{\Psi}(0) = 0$  such that, for any  $\lambda \geq 0$ , we have

$$\Psi_{\text{edge}}(\lambda) = \frac{\lambda^{1/2}}{3} (1 + \widetilde{\Psi}(\lambda)), \quad |\widetilde{\Psi}(\lambda)| \lesssim \min\{\lambda, \lambda^{1/3}\}. \quad (7.39)$$

(iv) If  $|\varepsilon| \leq 1/3$  then, for any  $\lambda \geq 0$ , we have

$$\Psi_{\text{edge}}((1 + \varepsilon)\lambda) = (1 + \varepsilon)^{1/2} \Psi_{\text{edge}}(\lambda) + \mathcal{O}(\varepsilon \min\{\lambda^{3/2}, \lambda^{1/3}\}). \quad (7.40)$$

We remark that (7.37) appeared in the last displayed equation of the proof of Proposition 9.3 in [3]. The relation (7.38) was present in (9.127) of [3]. Note that (7.39) is (9.144) of [3]. The relation in (7.40) is identical to (9.145) in [3]. Moreover, we follow the proofs in [3].

*Proof.* For the proof of (i), we first cancel the summand  $-1$  in the definition of  $\Psi_{\text{min}}$  in (7.1b) and  $-\text{Im} \widehat{\Omega}(0)/\sqrt{3}$ . Multiplying the result by the denominator in (7.1b) yields a trivial identity.

Part (ii) follows from using  $1 + 2\lambda \geq 1$  in the definition of  $\text{Im} \widehat{\Omega}_+(1 + 2\lambda)$  in (7.20) and multiplying the result by the denominator in the definition of  $\Psi_{\text{edge}}$  in (7.1a).

Expanding the definition of  $\Psi_{\text{edge}}$  around  $\lambda^{1/2}/3$  directly implies (iii).

For the proof of (iv), we remark that  $1 + \varepsilon \sim 1$  for all  $\varepsilon \in [-1/3, 1/3]$ . If  $0 \leq \lambda \lesssim 1$  then (7.40) follows from (7.39). For  $\lambda \gtrsim 1$ , the expansion (7.40) is a consequence of (7.38) above as well as the stability of Cardano's solutions, (9.111) in Lemma 9.15 of [3].  $\square$

In the following proof of Theorem 7.7, we will choose appropriate normal coordinates  $\Omega$  and  $\Lambda$  in each case such that (7.11) turns into one of the cubic

equations in normal form from Proposition 7.10. This procedure has been similarly performed in the proofs of Proposition 9.3, Lemma 9.11, Lemma 9.12 and Section 9.2.2 in [3]. However, owing to the weaker error bounds here, we include the proof for the sake of completeness.

*Proof of Theorem 7.7.* We start with the proof of part (i) (cf. Proposition 9.3 in [3]). Owing to (7.13b) and  $|\Psi_{\min}(\lambda)| \lesssim |\lambda|^{1/3}$ , the statement of (7.14) is trivial for  $|\omega| \gtrsim 1$  since the error term dominates. Therefore, it suffices to prove (7.14) for  $|\omega| \leq \delta$  with some  $\delta \sim 1$ .

By possibly shrinking  $\rho_* \sim 1$ , we can assume that  $|\sigma| \leq \Pi_* \rho_*^2$  is small enough such that  $\psi \sim 1$  by (7.12). In the following,  $\gamma_\nu, \gamma_0, \gamma_1, \dots, \gamma_7 \in \mathbb{C}$  denote  $\omega$ -independent error terms such that  $|\gamma_k| \lesssim \rho$  for  $k = \nu, 0, 1, \dots, 7$ . Their precise forms can be determined along the proof but we omit them as they are irrelevant. We divide (7.11) by  $\mu_3$  and obtain

$$\Theta^3 + i3\rho(1 + \gamma_2)\Theta^2 - 2\rho^2(1 + \gamma_1)\Theta + (1 + \gamma_0 + (1 + \gamma_\nu)\nu(\omega))\frac{\kappa}{\psi}\omega = 0, \tag{7.41}$$

using  $|\mu_3| \sim 1$  and  $|\sigma| \leq \Pi_* \rho^2$ . We introduce the normal coordinates

$$\lambda := \Gamma \frac{\omega}{\rho^3}, \quad \Omega(\lambda) := \sqrt{3} \left[ (1 + \gamma_3) \frac{1}{\rho} \Theta \left( \frac{\rho^3}{\Gamma} \lambda \right) + i + \gamma_4 \right], \tag{7.42}$$

where  $\Gamma := \sqrt{27}\kappa/(2\psi)$ . Note that  $\Gamma \sim 1$  since  $\psi \sim 1$ . A straightforward computation starting from (7.41) shows that  $\Omega(\lambda)$  and  $\Lambda(\lambda)$  satisfy (7.21) with

$$\Lambda(\lambda) := (1 + \gamma_5 + \mu(\lambda))\lambda + \gamma_6, \quad \mu(\lambda) := (1 + \gamma_7)\nu \left( \frac{\rho^3}{\Gamma} \lambda \right),$$

i.e.,  $\chi_2 = \gamma_5$ ,  $\chi_3 = \gamma_6$  and  $\alpha = \rho$  by (7.13a). Hence, from (7.22) and (7.42), we obtain  $\delta \sim 1$  and  $\chi_* \sim 1$  such that

$$\begin{aligned} \text{Im } \Theta(\omega) &= \text{Im } \frac{\rho}{1 + \gamma_3} \frac{1}{\sqrt{3}} [\Omega(\lambda) - \Omega_0] \\ &= \rho \Psi_{\min} \left( \Gamma \frac{\omega}{\rho^3} \right) + \mathcal{O} \left( \rho^2 \min\{|\lambda|, |\lambda|^{1/3}\} + \rho^2 \min\{|\lambda|, |\lambda|^{2/3}\} \right) \end{aligned}$$

for  $|\lambda| \leq \delta/\rho^3$  if  $\rho \leq \min\{\chi_*, \rho_*\}$ . Here, we also used (7.23) to expand  $\rho/(1 + \gamma_3)$  and (7.37). By employing (7.42) again and replacing  $\rho_*$  by  $\min\{\chi_*, \rho_*\}$ , we conclude (7.14).

We now turn to the proof of part (ii) of Theorem 7.7. Since  $\rho = 0$ , the cubic equation (7.11) simplifies to the following equation

$$\psi\Theta(\omega)^3 + \sigma\Theta(\omega)^2 + \kappa(1 + \nu(\omega))\omega = 0. \tag{7.43}$$

We now prove Theorem 7.7 (ii) (a), i.e., the case  $\sigma = 0$  (cf. Lemma 9.12 in [3]). For any  $\delta \sim 1$ , the assertion is trivial for  $|\omega| \geq \delta$  since the error term dominates  $|\omega|^{1/3}$  and  $\text{Im } \Theta(\omega)$  in this case (compare (7.13b)). Therefore, it

suffices to prove the lemma for  $|\omega| \leq \delta$  with some  $\delta \sim 1$ . We choose the normal coordinates

$$\lambda := \omega, \quad \Omega(\lambda) := \left(\frac{\psi}{\kappa}\right)^{1/3} \Theta(\lambda),$$

and notice that the cubic equation (7.43) becomes (7.26) with  $\mu(\lambda) = \nu(\lambda)$ . The bound (7.13a) implies  $|\mu(\lambda)| \lesssim |\lambda|^{1/3}$ . Thus, (7.16) is a consequence of Proposition 7.10 (iii). This completes the proof of (ii) (a).

For the proof of Theorem 7.7 (ii) (b), we first show the following auxiliary lemma (cf. Lemma 9.11 in [3]).

LEMMA 7.13 (Simple edge). *Let the assumptions of Theorem 7.7 (ii) hold true. If  $\sigma \neq 0$  then there is  $c_* \sim 1$  such that*

$$\operatorname{Im} \Theta(\omega) = \begin{cases} \sqrt{\kappa} \left| \frac{\omega}{\sigma} \right|^{1/2} + \mathcal{O}(|\nu(\omega)| + |\sigma|^{-1} |\Theta(\omega)| \left| \frac{\omega}{\sigma} \right|^{1/2}), & \text{if } \operatorname{sign} \omega = \operatorname{sign} \sigma, \ |\omega| \leq c_* |\sigma|^3, \\ 0, & \text{if } \operatorname{sign} \omega = -\operatorname{sign} \sigma, \ |\omega| \leq c_* |\sigma|^3. \end{cases} \tag{7.44}$$

Moreover, we have  $|\Theta(\omega)| \lesssim |\omega/\sigma|^{1/2}$  for  $|\omega| \leq c_* |\sigma|^3$ .

*Proof.* Dividing (7.43) by  $\kappa\sigma$  yields

$$\left(1 + \frac{\psi}{\sigma} \Theta(\omega)\right) \frac{\Theta(\omega)^2}{\kappa} + (1 + \nu(\omega)) \frac{\omega}{\sigma} = 0. \tag{7.45}$$

We introduce  $\lambda, \Omega(\lambda)$  and  $\mu(\lambda)$  defined by

$$\lambda := \frac{\omega}{\sigma}, \quad \Omega(\lambda) := \frac{1}{\sqrt{\kappa}} \Theta(\sigma\lambda), \quad \mu(\lambda) := \frac{1 + \nu(\sigma\lambda)}{1 + \psi\sigma^{-1}\Theta(\sigma\lambda)} - 1.$$

In the normal coordinates  $\lambda$  and  $\Omega(\lambda)$ , (7.45) viewed as a quadratic equation, fulfills (7.24) with the above choice of  $\mu(\lambda)$ . Since  $|\psi\sigma^{-1}\Theta(\sigma\lambda)| \lesssim |\sigma|^{-2/3} |\lambda|^{1/3}$  by (7.13b), there is  $c_* \sim 1$  such that

$$|\mu(\lambda)| \lesssim |\nu(\sigma\lambda)| + |\sigma|^{-1} |\Theta(\sigma\lambda)| \lesssim |\sigma|^{-2/3} |\lambda|^{1/3}, \quad |\operatorname{Im} \mu(\lambda)| \lesssim |\sigma|^{-1} |\operatorname{Im} \Theta(\sigma\lambda)| \tag{7.46}$$

for  $|\lambda| \leq c_* |\sigma|^2$  by (7.13a), (7.13b) and (7.15). Hence, we apply Proposition 7.10 (ii) with  $\gamma \sim |\sigma|^{-1}$  in  $\operatorname{SP}4'$  and obtain (7.44) with an error term  $\mathcal{O}(|\mu(\lambda)| |\lambda|^{1/2})$  instead, as well as  $|\Theta(\omega)| \lesssim |\sigma|^{-1/2} |\omega|^{1/2}$ . Thus, the first bound in (7.46) completes the proof of (7.44).  $\square$

From the second case in (7.44), we conclude the second case in (7.17). The first case in (7.17) and (7.19) are trivial if  $|\omega| \gtrsim 1$  due to (7.13b) and (7.4a). Hence, it suffices to prove this case for  $|\omega| \leq \delta$  with some  $\delta \sim 1$ . If  $|\sigma| \gtrsim 1$  then the first case in (7.17) also follows from (7.44) with  $\delta := c_* |\sigma|^3$ . Indeed, from (7.39), we conclude

$$\sqrt{\kappa} \left| \frac{\omega}{\sigma} \right|^{1/2} = c \widehat{\Delta}^{1/3} \Psi_{\text{edge}} \left( \frac{|\omega|}{\Delta} \right) + \mathcal{O}(|\omega|^{3/2}),$$

where  $c$  and  $\widehat{\Delta}$  are defined as in (7.18). Since  $|\omega| \lesssim \varepsilon(\omega)$  for  $|\omega| \leq \delta$  and  $\varepsilon(\omega)$  defined as in (7.18) we obtain the first case in (7.17) if  $|\sigma| \gtrsim 1$ . Similarly,  $|\Theta(\omega)| \lesssim |\omega/\sigma|^{1/2}$  by Lemma 7.13 yields (7.19) if  $|\omega| \leq \delta$  and  $|\sigma| \gtrsim 1$ . Hence, it remains to show the first case in (7.17) and (7.19) if  $|\sigma| \leq \sigma_*$  for some  $\sigma_* \sim 1$ . In fact, we choose  $\sigma_* \sim 1$  so small that  $\psi \sim 1$  by (7.12) and  $\widehat{\Delta} < 1$  for  $|\sigma| \leq \sigma_*$ . In order to apply Proposition 7.10 (iv), we introduce

$$\lambda := \frac{2}{\widehat{\Delta}}\omega, \quad \Omega(\lambda) := 3\frac{\psi}{|\sigma|}\Theta\left(\frac{\widehat{\Delta}}{2}\lambda\right) + \text{sign } \sigma, \quad \mu(\lambda) := \nu\left(\frac{\widehat{\Delta}}{2}\lambda\right) \quad (7.47)$$

(cf. (9.96) and (9.99) in [3]). The cubic (7.43) takes the form (7.28) in the normal coordinates  $\lambda$  and  $\Omega(\lambda)$  with the above choice of  $\mu(\lambda)$  and  $s = \text{sign } \sigma$  in (7.28). By (7.13a), we have  $|\mu(\lambda)| \lesssim \widehat{\Delta}^{1/3}|\lambda|^{1/3}$ . We set  $\widehat{\gamma} := \widehat{\Delta}^{1/3}$ . Therefore, Proposition 7.10 (iv) and (7.38) yield  $\delta \sim 1$  and possibly smaller  $\sigma_* := \min\{\sigma_*, \gamma_*\} \sim 1$  such that the first case in (7.17) holds true for  $|\sigma| \leq \sigma_*$  and  $|\omega| \leq \delta$  as  $\mu(\lambda) = \nu(\omega)$  and  $\widehat{\Delta} \sim |\sigma|^3$ . Moreover, (7.30) implies (7.19) for  $|\omega| \leq \delta$ . This completes the proof of (ii) (b) and hence of Theorem 7.7.  $\square$

7.4 PROOF OF THEOREM 7.1 AND PROPOSITION 7.5

In this section, we prove Theorem 7.1 and Proposition 7.5. Some parts of the following proof resemble the proofs of Theorem 2.6, Proposition 9.3 and Proposition 9.8 in [3]. However, owing to the weaker assumptions, we present it here for the sake of completeness.

*Proof of Theorem 7.1 and Proposition 7.5.* We will only prove the statements in Proposition 7.5. Theorem 7.1 is a direct consequence of this proposition as well as Lemma 4.8 (ii) and Proposition 6.1.

Along the proof of Proposition 7.5, we will shrink  $\delta_* \sim 1$  such that (7.3) holds true for all  $\omega \in [-\delta_*, \delta_*] \cap J \cap D$ . We will transfer the expansions of  $\Theta$  in Theorem 7.7 to expansions of  $v$  by means of (6.1). To that end, we take the imaginary part of (6.1) and obtain

$$v(\tau_0 + \omega) = v(\tau_0) + \pi^{-1}\text{Re } b\text{Im } \Theta(\omega) + \pi^{-1}\text{Im } b\text{Re } \Theta(\omega) + \pi^{-1}\text{Im } r(\omega). \quad (7.48)$$

We first establish (7.3) at a shape regular point  $\tau_0 \in (\text{supp } \rho) \setminus \partial \text{supp } \rho$  which is a local minimum of  $\tau \mapsto \rho(\tau)$ . If  $\rho = \rho(\tau_0) = 0$ , i.e., the case of a cusp at  $\tau_0$ , case (c), then  $\sigma = 0$ . Indeed, if  $\sigma$  were not 0, then, by the second case in (7.17),  $\text{Im } \Theta(\omega)$  would vanish on one side of  $\tau_0$ . By the third bound in (6.8), this would imply the vanishing of  $\rho$  as well, contradicting to  $\tau_0 \in \text{supp } \rho \setminus \partial \text{supp } \rho$ . Hence, for any  $\delta_* \sim 1$ , (7.16) and (7.48) immediately yield (7.3) for  $\omega \in [-\delta_*, \delta_*] \cap J \cap D$  with  $h = (2\pi)^{-1}b\sqrt{3}(\kappa/\psi)^{1/3}$  using (6.7a), (6.7b) and  $b = b^*$  due to  $\rho = 0$ .

We now assume  $\rho > 0$  which corresponds to an internal nonzero minimum at  $\tau_0$ , case (d). Thus, the following lemma implies that the condition  $|\sigma| \leq \Pi_*\rho^2$ ,  $\sigma = \sigma(\tau_0)$ , needed to apply Theorem 7.7 (i) is fulfilled. We will prove Lemma 7.14 at the end of this section.

LEMMA 7.14 (Bound on  $|\sigma|$  at nonzero local minimum). *There are thresholds  $\rho_* \sim 1$  and  $\Pi_* \sim 1$  such that*

$$|\sigma(\tau_0)| \leq \Pi_* \rho(\tau_0)^2$$

for each shape regular point  $\tau_0 \in \text{supp } \rho$  which is a local minimum of  $\rho$  and satisfies  $0 < \rho(\tau_0) \leq \rho_*$ .

Hence, (7.14), (7.48) and (6.7b) yield (7.3) with  $\tilde{\rho} = \rho\Gamma^{-1/3}$  and  $h = \pi^{-1}\Gamma^{1/3}\text{Re } b$ . Here, we also used

$$\rho|\Theta(\omega)| + |\Theta(\omega)|^2 + |\omega| + \min\{\rho^{-1}|\omega|, |\omega|^{2/3}\} \lesssim \frac{|\omega|}{\rho} \mathbf{1}(|\omega| \lesssim \rho^3) + \Psi(\omega)^2, \tag{7.49}$$

which is a consequence of (6.7a), (7.4b) for  $|\omega| \lesssim 1$ , as well as  $\text{Re } b \sim 1$  and  $\text{Im } b = \mathcal{O}(\rho)$ . This completes the proof of (7.3) for shape regular points  $\tau_0 \in (\text{supp } \rho) \setminus \partial \text{supp } \rho$ , cases (c) and (d).

We now turn to the proof of (7.3) at an edge  $\tau_0$ , case (a), i.e., for a shape regular point  $\tau_0 \in \partial \text{supp } \rho$ . We first prove a version of (7.3) with  $\widehat{\Delta}$  in place of  $\Delta$ , (7.50) below. In a second step, we then replace  $\widehat{\Delta}$  by  $\Delta$  to obtain (7.3). Since  $\tau_0 \in \partial \text{supp } \rho$ , we have  $\rho = \rho(\tau_0) = 0$ . Therefore,  $v(\tau_0) = 0$  since  $\langle \cdot \rangle$  is a faithful trace and  $v(\tau_0)$  is positive semidefinite. As  $\tau_0 \in \partial \text{supp } \rho$ , we have  $\sigma(\tau_0) \neq 0$ . Indeed, assuming  $\sigma(\tau_0) = 0$ , using Theorem 7.7 (ii) (a), taking the imaginary part of (6.1) as well as applying the third bound in (6.8) and the second bound in (6.7a) yield the contradiction  $\tau_0 \in (\text{supp } \rho) \setminus \partial \text{supp } \rho$ . Recalling the definitions of  $\widehat{\Delta}$  and  $c$  from (7.18), (7.48) and (7.17) yield

$$v(\tau_0 + \omega) = \pi^{-1}c\widehat{\Psi}(\omega)b + \mathcal{O}(\widehat{\Psi}(\omega)^2), \quad \widehat{\Psi}(\omega) := \widehat{\Delta}^{1/3}\Psi_{\text{edge}}\left(\frac{|\omega|}{\widehat{\Delta}}\right) \tag{7.50}$$

for any  $\omega \in [-\delta_*, \delta_*] \cap J \cap D$  with  $\text{sign } \omega = \text{sign } \sigma$  and some  $\delta_* \sim 1$ . Here, we also used  $b = b^* \sim 1$ , the first bound in (6.5), (7.19) and  $\varepsilon(\omega) \sim \widehat{\Psi}(\omega)$  by (7.4b) to obtain

$$|\Theta(\omega)|^2 + |\omega| + (|\Theta(\omega)| + |\omega| + \varepsilon(\omega))\varepsilon(\omega) \lesssim \widehat{\Psi}(\omega)^2$$

for any  $\omega \in [-\delta_*, \delta_*] \cap J \cap D$  with  $\text{sign } \omega = \text{sign } \sigma$  and some  $\delta_* \sim 1$ . This means that we have shown (7.3) with  $\Psi$  replaced by  $\widehat{\Psi}$ .

We now replace  $\widehat{\Delta}$  by  $\Delta$  in (7.50) to obtain (7.3). To that end, we first assume that  $|\sigma| \gtrsim 1$  and  $\Delta \lesssim 1$ . The second part of (7.17) implies  $|\sigma|^3 \lesssim \Delta \lesssim 1$  and thus  $|\sigma|^3 \sim \Delta \sim 1$ . Since  $|\sigma|^3 \sim \widehat{\Delta}$  we conclude  $\widehat{\Delta} \sim \Delta$ . Therefore, we obtain

$$\widehat{\Delta}^{1/3}\Psi_{\text{edge}}\left(\frac{|\omega|}{\widehat{\Delta}}\right) = \left(\frac{\Delta}{\widehat{\Delta}}\right)^{1/6} \Delta^{1/3}\Psi_{\text{edge}}\left(\frac{|\omega|}{\Delta}\right) + \mathcal{O}(\min\{|\omega|^{3/2}, |\omega|^{1/3}\}).$$

Here, we used  $\Psi_{\text{edge}}(|\lambda|) \lesssim |\lambda|^{1/3}$  for  $|\lambda| \gtrsim 1$  and (7.39) otherwise. Applying this relation to (7.50) yields (7.3) for  $\omega \in [-\delta_*, \delta_*] \cap J \cap D$  with  $\text{sign } \omega = \text{sign } \sigma$ ,  $\delta_* \sim 1$  and  $h := \pi^{-1}c(\Delta/\widehat{\Delta})^{1/6}b \sim 1$  for  $|\sigma| \gtrsim 1$  and  $\Delta \lesssim 1$ .

The next lemma shows that  $|\sigma| \gtrsim 1$  at the edge of a gap of size  $\Delta \gtrsim 1$ . We postpone its proof until the end of this section.



LEMMA 7.15 ( *$\sigma$  at an edge of a large gap*). *Let  $\tau_0 \in \partial \text{supp } \rho$  be a shape regular point for  $m$  on  $J$ . If  $|\inf J| \gtrsim 1$  and there is  $\varepsilon \sim 1$  such that  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0 - \varepsilon, \tau_0]$  then  $|\sigma| \sim 1$ . We also have  $|\sigma| \sim 1$  if  $\sup J \gtrsim 1$  and  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0, \tau_0 + \varepsilon]$  and some  $\varepsilon \sim 1$ .*

Under the assumptions of the previous lemma, we set  $\Delta := 1$  and obtain trivially  $\widehat{\Delta} \sim 1 \sim \Delta$ . Thus, (7.50) implies (7.3) by the same argument as in the case  $\Delta \lesssim 1$ .

For  $|\sigma| \leq \sigma_*$  with some sufficiently small  $\sigma_* \sim 1$ , we will prove below with the help of the following Lemma 7.16 and (7.40) that replacing  $\widehat{\Delta}$  by  $\Delta$  in (7.50) yields an affordable error. We present the proof of Lemma 7.16 at the end of this section.

LEMMA 7.16 (*Size of small gap*). *Let  $\tau_0, \tau_1 \in \partial \text{supp } \rho$ ,  $\tau_1 < \tau_0$ , be two shape regular points for  $m$  on  $J_0$  and  $J_1$ , respectively, where  $J_0, J_1 \subset \mathbb{R}$  are two open intervals with  $0 \in J_0 \cap J_1$ . We assume  $|\inf J_0| \gtrsim 1$  and  $\sup J_1 \gtrsim 1$  as well as  $(\tau_1, \tau_0) \cap \text{supp } \rho = \emptyset$ . We set  $\Delta(\tau_0) := \tau_0 - \tau_1$ . Then there is  $\tilde{\sigma} \sim 1$  such that if  $|\sigma(\tau_0)| \leq \tilde{\sigma}$  and  $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim |\tau_0 - \tau_1|^{1/3}$  then*

$$\frac{\Delta(\tau_0)}{\widehat{\Delta}(\tau_0)} = 1 + \mathcal{O}(\sigma(\tau_0)).$$

The same statement holds true when  $\tau_0$  is replaced by  $\tau_1$  with  $\Delta(\tau_1) := \tau_0 - \tau_1$ .

From Lemma 7.16, we conclude that there is  $\gamma \in \mathbb{C}$  such that  $|\gamma| \lesssim 1$  and  $\Delta = (1 + \gamma|\sigma|)\widehat{\Delta}$ . By possibly shrinking  $\sigma_* \sim 1$ , we can assume that  $|\gamma\sigma| \leq 1/3$  for  $|\sigma| \leq \sigma_*$ . Thus, (7.40) in Lemma 7.12 (iv) yields

$$\widehat{\Delta}^{1/3} \Psi_{\text{edge}}\left(\frac{|\omega|}{\widehat{\Delta}}\right) = \left(\frac{\Delta}{\widehat{\Delta}}\right)^{1/6} \Delta^{1/3} \Psi_{\text{edge}}\left(\frac{|\omega|}{\Delta}\right) + \mathcal{O}\left(\min\left\{\frac{|\omega|^{3/2}}{\Delta^{5/6}}, |\omega|^{1/3}\right\}\right).$$

Hence, choosing  $h := \pi^{-1}c(\Delta/\widehat{\Delta})^{1/6}b$  as before and noticing  $h \sim 1$  yields (7.3) in the missing regime. This completes the proof of Proposition 7.5. As we have already explained, Theorem 7.1 follows immediately.  $\square$

The core of the proof of Lemma 7.14 is an effective monotonicity estimate on  $v$ , see (7.51) below, which is the analogue of (9.20) in Lemma 9.2 of [3]. Owing to the weaker assumptions on the coefficients of the cubic equation, we need to present an upgraded proof here.

*Proof of Lemma 7.14.* In the whole proof, we will use the notation of Definition 7.4. We will show below that there are  $\rho_* \sim 1$  and  $\Pi_* \sim 1$  such that

$$(\text{sign } \kappa_1 \sigma(\tau)) \partial_\tau v(\tau) \gtrsim \rho(\tau)^{-1} \tag{7.51}$$

for all  $\tau \in \mathbb{R}$  which satisfy  $\rho(\tau) \in (0, \rho_*]$  and  $|\sigma(\tau)| \geq \Pi_* \rho(\tau)^2$  and are admissible points for the shape analysis.

Now, we first conclude the statement of the lemma from (7.51) through a proof by contradiction. If  $\tau_0$  satisfies the conditions of Lemma 7.14 then  $\partial_\tau \rho(\tau_0) = 0$  as  $\tau_0$  is a local minimum of  $\rho$ . Assuming  $|\sigma(\tau_0)| \geq \Pi_* \rho(\tau_0)^2$  and applying  $\langle \cdot \rangle$  to (7.51) yield the contradiction  $\partial_\tau \rho(\tau_0) > 0$ .

For the proof of (7.51) we start by proving a relation for  $\partial_\tau v(\tau)$ . We divide (6.1) by  $\omega$ , use  $\Theta(0) = 0$  and  $r(0) = 0$  as well as take the limit  $\omega \rightarrow 0$  to obtain  $\partial_\tau m(\tau) = b \partial_\omega \Theta(0) + \partial_\omega r(0)$ . Taking the imaginary part of the previous relation yields

$$\pi \partial_\tau v(\tau) = \text{Im} [b \partial_\omega \Theta(0)] + \text{Im} \partial_\omega r(0). \tag{7.52}$$

We divide (6.7b) by  $\omega$ , employ the first bound in (6.7a) and obtain

$$\left\| \frac{r(\omega)}{\omega} \right\| \lesssim 1 + \left| \frac{\Theta(\omega)}{\omega} \right|^2 \lesssim 1 + \frac{|\omega|}{\rho^4}.$$

By sending  $\omega \rightarrow 0$  and using  $r(0) = 0$ , we conclude

$$\|\text{Im} \partial_\omega r(0)\| \lesssim 1. \tag{7.53}$$

We divide (6.3) by  $\mu_1 \omega$ , take the limit  $\omega \rightarrow 0$  and use  $\lim_{\omega \rightarrow 0} \Theta(\omega) = \Theta(0) = 0$  to obtain

$$\begin{aligned} \partial_\omega \Theta(0) &= -\frac{\Xi(0) \bar{\mu}_1}{|\mu_1|^2} = \frac{(\kappa + \mathcal{O}(\rho))(i\kappa_1 \rho \sigma + 2\rho^2 \psi + \mathcal{O}(\rho^3 + \rho^2 |\sigma|))}{4\rho^4 |\psi + \mathcal{O}(\rho + |\sigma|)|^2 + \rho^2 |\kappa_1 \sigma + \mathcal{O}(\rho^2 + \rho |\sigma|)|^2} \\ &= \frac{\kappa}{\rho} \frac{i\kappa_1 \sigma + 2\rho \psi + \mathcal{O}(\rho^2 + \rho |\sigma|)}{4\rho^2 |\psi + \mathcal{O}(\rho + |\sigma|)|^2 + |\kappa_1 \sigma + \mathcal{O}(\rho^2 + \rho |\sigma|)|^2}, \end{aligned} \tag{7.54}$$

where we employed  $|\mu_1|^2 = 4\rho^4 |\psi + \mathcal{O}(\rho + |\sigma|)|^2 + \rho^2 |\kappa_1 \sigma + \mathcal{O}(\rho^2 + \rho |\sigma|)|^2$  as  $\rho, \psi, \kappa_1, \sigma \in \mathbb{R}$ . Thus, we obtain

$$\rho |\text{Re} \partial_\omega \Theta(0)| \lesssim \frac{\rho + \rho |\sigma|}{\rho^2 |\psi + \mathcal{O}(\rho + |\sigma|)|^2 + |\kappa_1 \sigma + \mathcal{O}(\rho^2 + \rho |\sigma|)|^2}. \tag{7.55}$$

Therefore, using  $b = b^* + \mathcal{O}(\rho)$ ,  $b + b^* \sim 1$ ,  $\kappa \sim 1$  and  $|\kappa_1| \sim 1$  yields

$$\begin{aligned} (\text{sign } \kappa_1 \sigma) \text{Im} [b \partial_\omega \Theta(0)] &\gtrsim \frac{\rho^{-1} |\sigma| + \mathcal{O}(\rho + |\sigma|) + \mathcal{O}(\rho + \rho |\sigma|)}{|\sigma + \mathcal{O}(\rho^2 + \rho |\sigma|)|^2 + \rho^2 |\psi + \mathcal{O}(\rho + |\sigma|)|^2} \\ &\gtrsim \frac{|\sigma|}{|\sigma|^2 + \rho^2} \frac{1}{\rho}. \end{aligned}$$

Here, in the first step, the error term  $\mathcal{O}(\rho + \rho |\sigma|)$  in the numerator originates from the second term in

$$\begin{aligned} (\text{sign } \kappa_1 \sigma) \text{Im} [b \partial_\omega \Theta(0)] &= (\text{sign } \kappa_1 \sigma) (\text{Re } b \text{Im} \partial_\omega \Theta(0) + \text{Im } b \text{Re} \partial_\omega \Theta(0)) \\ &\gtrsim (\text{sign } \kappa_1 \sigma) \text{Im} \partial_\omega \Theta(0) - \rho |\text{Re} \partial_\omega \Theta(0)| \end{aligned} \tag{7.56}$$

and applying (7.55) to it. We applied (7.54) to the first term on the right-hand side of (7.56). In the last estimate, we used  $\psi, |\sigma|, \rho \lesssim 1$  and  $|\sigma| \geq \Pi_* \rho^2$

for some large  $\Pi_* \sim 1$  as well as  $\rho \leq \rho_*$  for some small  $\rho_* \sim 1$ . Employing  $|\sigma| \geq \Pi_* \rho^2$  once more, the factor  $|\sigma|/(|\sigma|^2 + \rho^2)$  on the right-hand side scales like  $(1 + |\sigma|)^{-1} \gtrsim 1$ . Hence, we conclude from (7.52) and (7.53) that

$$(\text{sign } \kappa_1 \sigma) \partial_\tau v(\tau) \gtrsim \frac{1}{\rho} + \mathcal{O}(1).$$

By choosing  $\rho_* \sim 1$  sufficiently small, we obtain (7.51). This completes the proof of Lemma 7.14.  $\square$

*Proof of Lemma 7.15.* We prove both cases,  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0 - \varepsilon, \tau_0]$  or for all  $\tau \in [\tau_0, \tau_0 + \varepsilon]$ , in parallel. We can assume that  $|\sigma| \leq \tilde{\sigma}$  for any  $\tilde{\sigma} \sim 1$  as the statement trivially holds true otherwise. We choose  $(\delta, \varrho, \gamma_*)$  as in Proposition 7.10 (iv),  $\widehat{\Delta}$  as in (7.18), normal coordinates  $(\lambda, \Omega(\lambda))$  as in (7.47) as well as  $\widehat{\gamma} = \widehat{\Delta}^{1/3}$  and  $s = \text{sign } \sigma$ . We set  $\lambda_3 := 2 + \varrho \widehat{\Delta}^{1/3}$  (cf. (7.34)) and  $\omega_3 := \widehat{\Delta} \lambda_3 / 2$ . There is  $\tilde{\sigma} \sim 1$  such that  $\widehat{\Delta} \leq \gamma_*^3$  for  $|\sigma| \leq \tilde{\sigma}$  due to  $\widehat{\Delta} \sim |\sigma|^3$  by (6.6) and the definition of  $\widehat{\Delta}$  in (7.18). Hence,  $\omega_3 \leq C|\sigma|^3$  and, by possibly shrinking  $\tilde{\sigma} \sim 1$ , we obtain  $-\omega_3 \text{sign } \sigma \in J$  for  $|\sigma| \leq \tilde{\sigma}$  due to the assumption on  $J$  ( $|\inf J| \gtrsim 1$  or  $\sup J \gtrsim 1$ ). From (7.32), we obtain  $\text{Im } \Omega(-\lambda_3 \text{sign } \sigma) > 0$ . Hence,  $\text{Im } \Theta(-\omega_3 \text{sign } \sigma) > 0$ . From the third bound in (6.8), the second bound in (6.7a) and  $\omega_3 \lesssim |\sigma|^3$ , we conclude  $v(-\omega_3 \text{sign } \sigma) > 0$  for  $|\sigma| \leq \tilde{\sigma}$  and sufficiently small  $\tilde{\sigma} \sim 1$ . Thus,  $\rho(-\omega_3 \text{sign } \sigma) > 0$  which implies  $\omega_3 > \varepsilon$ . Therefore,  $|\sigma|^3 \gtrsim \omega_3 > \varepsilon \sim 1$  which completes the proof of Lemma 7.15.  $\square$

We finish this section by proving Lemma 7.16. It is similarly proven as Lemma 9.17 in [3]. We present the proof due to the weaker assumptions of Lemma 7.16. The main difference is the proof of (7.58) below (cf. (9.138) in [3]). In [3],  $\Theta$  could be explicitly represented in terms of  $m$ , i.e.,

$$\Theta(\omega) = \langle f, m(\tau_0 + \omega) - m(\tau_0) \rangle$$

(cf. (9.8) and (8.10c) in [3] with  $\alpha = 0$ ). In our setup,  $b$  and  $r$  do not necessarily define an orthogonal decomposition (cf. (6.1)).

*Proof of Lemma 7.16.* Let  $(\delta, \varrho, \gamma_*)$  be chosen as in Proposition 7.10 (iv). We choose  $\widehat{\Delta}$  as in (7.18) and normal coordinates as in (7.47) as well as  $\widehat{\gamma} = \widehat{\Delta}^{1/3}$  and  $s = \text{sign } \sigma$ . We assume  $\widehat{\Delta} \leq \gamma_*^3$  in the following and define  $\lambda_3$  as in (7.34). By using  $|\inf J_0| \gtrsim 1$  as in the proof of Lemma 7.15, we find  $\tilde{\sigma} \sim 1$  such that  $-\omega_3 \in J_0$  for  $\omega_3 := \lambda_3 \widehat{\Delta} / 2$  and  $|\sigma| \leq \tilde{\sigma}$ . Thus,  $-\Delta = \tau_1 - \tau_0 \in J_0$ . We set

$$\lambda_0 := \inf\{\lambda > 0: \text{Im } \Omega(\lambda) > 0\}$$

and remark that  $\lambda_0 = 2\Delta/\widehat{\Delta}$  due to the definition of  $\Delta$  and the third bound in (6.8). From (7.32), we conclude  $\lambda_0 \leq \lambda_3$ . Thus,  $\Delta \leq \widehat{\Delta}(1 + \mathcal{O}(\widehat{\gamma})) =$

$\widehat{\Delta}(1 + \mathcal{O}(|\sigma|))$  as  $\varrho \sim 1$  and  $\widehat{\gamma} \sim |\sigma|$ . Therefore, it suffices to show the opposite bound,

$$\Delta \geq \widehat{\Delta}(1 + \mathcal{O}(|\sigma|)). \tag{7.57}$$

If  $\lambda_0 \geq \lambda_2 := 2 - \varrho\widehat{\Delta}^{1/3}$  (cf. (7.34)) then we have (7.57) as  $\widehat{\Delta}^{1/3} \sim |\sigma|$  and  $\varrho \sim 1$ . If  $\lambda_0 < \lambda_2$  then we will prove below that

$$\text{Im } \Omega(\lambda_0 + \xi) \gtrsim \xi^{1/2} \tag{7.58}$$

for  $\xi \in [0, 1]$ . From (7.31), we then conclude

$$c_0(\lambda_2 - \lambda_0)^{1/2} \leq \text{Im } \Omega(\lambda_2) \leq C_1|\sigma|^{1/2}$$

as  $\widehat{\gamma} \sim |\sigma|$ . Hence,

$$\lambda_0 \geq \lambda_2 - (C_1/c_0)^2|\sigma| \geq 2 - C|\sigma|,$$

where we used  $\lambda_2 = 2 - \varrho\widehat{\gamma}$  and  $\varrho \sim 1$  in the last step. This shows (7.57) also in the case  $\lambda_0 < \lambda_2$ . Therefore, the proof of the lemma will be completed once (7.58) is proven.

In order to prove (7.58), we translate it into the coordinates  $\omega$  relative to  $\tau_0$  and  $v$ . From  $\lambda_0 < \lambda_2$ , we obtain

$$\Delta < (1 - \varrho\widehat{\Delta}^{1/3})\widehat{\Delta} \lesssim |\sigma|^3. \tag{7.59}$$

Since

$$\pi v(\tau_0 - \Delta - \widetilde{\omega}) = b\text{Im } \Theta(-\Delta - \widetilde{\omega}) + \text{Im } r(-\Delta - \widetilde{\omega}),$$

the bound (7.58) would follow from

$$v(\tau_0 - \Delta - \widetilde{\omega}) \gtrsim \widehat{\Delta}(\tau_0)^{-1/6}|\widetilde{\omega}|^{1/2} \tag{7.60}$$

for sufficiently small  $\Delta \lesssim |\sigma|^3 \leq \widetilde{\sigma}^3$  and  $\widetilde{\omega} \leq \widetilde{\delta}$  due to the third bound in (6.8). Since  $v(\tau_1) = 0$  and  $\tau_1 = \tau_0 - \Delta$  is a shape regular point, we conclude from (7.50) that

$$v(\tau_1 - \widetilde{\omega}) \gtrsim \widehat{\Delta}(\tau_1)^{-1/6}|\widetilde{\omega}|^{1/2}$$

for  $|\widetilde{\omega}| \leq \delta$ . Therefore, it suffices to show that

$$\widehat{\Delta}(\tau_1) \lesssim \widehat{\Delta}(\tau_0) \tag{7.61}$$

in order to verify (7.60). Owing to  $|\sigma(\tau_0) - \sigma(\tau_1)| \lesssim \Delta^{1/3}$  and (7.59), we have

$$|\sigma(\tau_1)| \lesssim |\sigma(\tau_0)| + \Delta^{1/3} \lesssim |\sigma(\tau_0)|.$$

We allow for a smaller choice of  $\widetilde{\sigma} \sim 1$  and assume  $\psi(\tau_1) \sim \psi(\tau_0) \sim 1$  by (6.6). Assuming without loss of generality  $\widehat{\Delta}(\tau_0) < 1$  and  $\widehat{\Delta}(\tau_1) < 1$ , we obtain (7.61) by the definition of  $\widehat{\Delta}$  in (7.18). We thus get (7.61) and hence (7.60). This proves (7.58) and completes the proof of Lemma 7.16.  $\square$

7.5 PROOFS OF THEOREM 7.2 AND PROPOSITION 7.6

*Proof of Proposition 7.6.* We start with the proof of part (i). We apply  $\langle \cdot \rangle$  to (7.3), use  $\rho = \langle v \rangle$  and obtain  $\langle h \rangle$  from the definitions of  $h$  in the four cases given in the proof of Proposition 7.5. Indeed, by using the relations

$$\langle b \rangle = \pi + \mathcal{O}(\rho), \quad c^3 = 4\Gamma, \tag{7.62}$$

which are proven below, as well as Lemma 7.16 in the cases (a) and (b) and the stronger error estimate (7.49) in case (d), we conclude part (i) of Proposition 7.6 up to the proof of (7.62).

The first relation in (7.62) follows from applying  $\langle \cdot \rangle$  to (5.15a) and using (5.14a), Corollary D.2 with  $\tau_0 \in \text{supp } \rho$ , the cyclicity of  $\langle \cdot \rangle$  and (5.19). The second relation in (7.62) is a consequence of the definition of  $c$  in (7.18) and the definition of  $\Gamma$  in Theorem 7.7 (i). This completes the proof of part (i).

We now turn to the proof of part (ii) of Proposition 7.6 and assume that all points of  $(\partial \text{supp } \rho) \cap I$  are shape regular for  $m$  and all estimates in Definition 7.4 hold true uniformly on this set. As in the proof of Proposition 7.5, we conclude  $\sigma(\tau_0) \neq 0$  for all  $\tau_0 \in (\partial \text{supp } \rho) \cap I$ . Owing to  $\text{dist}(0, \partial J) \gtrsim 1$  and the Hölder-continuity of  $\sigma$  on  $(\partial \text{supp } \rho) \cap I$ , Proposition 7.5 is applicable to every  $\tau_0 \in (\partial \text{supp } \rho) \cap I$ . Hence, (7.4a) and  $\text{dist}(0, \partial J) \gtrsim 1$  imply the existence of  $\delta_1, c_1 \sim 1$  such that

$$\rho(\tau_0 + \omega) \geq c_1 |\omega|^{1/2} \tag{7.63}$$

for all  $\omega \in -\text{sign } \sigma(\tau_0)[0, \delta_1]$  and  $\tau_0 \in (\partial \text{supp } \rho) \cap I$ . In particular,  $\tau_0 - \text{sign } \sigma(\tau_0)[0, \delta_1] \subset \text{supp } \rho$  for all  $\tau_0 \in (\partial \text{supp } \rho) \cap I$ . Since  $|I| \lesssim 1$ , this implies that  $\text{supp } \rho \cap I$  consists of finitely many intervals  $[\alpha_i, \beta_i]$  with lengths  $\gtrsim 1$ , and, thus, their number  $K$  satisfies  $K \sim 1$  as  $\delta_1 \sim 1$  and  $\beta_i - \alpha_i \geq \delta_1$  if  $\beta_i \neq \sup I$  and  $\alpha_i \neq \inf I$ .

Additionally, we now assume that the elements of  $\mathbb{M}_{\rho_*}$  are shape regular points for  $m$  on  $J$  and all estimates in Definition 7.4 hold true uniformly on  $\mathbb{M}_{\rho_*}$ . By possibly shrinking  $\rho_* \sim 1$ , we conclude from (7.63) that  $|\alpha_i - \gamma| \sim 1$  and  $|\beta_i - \gamma| \sim 1$  for any  $i = 1, \dots, K$  and  $\gamma \in \mathbb{M}_{\rho_*}$ .

Suppose now that  $\tau_0 \in \mathbb{M}_{\rho_*}$  with  $\rho(\tau_0) = 0$ . Then part (i) and  $\text{dist}(0, \partial J) \gtrsim 1$  yield the existence of  $\delta_2, c_2 \sim 1$  such that

$$\rho(\tau_0 + \omega) \geq c_2 |\omega|^{1/3}$$

for all  $|\omega| \leq \delta_2$ . By possibly further shrinking  $\rho_* \sim 1$ , we thus obtain  $|\tau_0 - \gamma| \sim 1$  for all  $\gamma \in \mathbb{M}_{\rho_*} \setminus \{\tau_0\}$ . We thus conclude (7.10) in this case.

Finally, let  $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$  with  $\rho(\gamma_1), \rho(\gamma_2) > 0$ . Then applying (i) with  $\tau_0 = \gamma_1$  and  $\tau_0 = \gamma_2$  yields

$$\begin{aligned} \Psi_1(\omega) + \Psi_2(\omega) &\lesssim |\omega|^{1/3} (\rho(\gamma_1) \mathbf{1}(|\omega| \lesssim \rho(\gamma_1)^3) + \rho(\gamma_2) \mathbf{1}(|\omega| \lesssim \rho(\gamma_2)^3)) \\ &\quad + \Psi_1(\omega)^2 + \Psi_2(\omega)^2, \end{aligned}$$

where we defined  $\omega = \gamma_2 - \gamma_1$  and

$$\Psi_1(\omega) := \tilde{\rho}_1 \Psi_{\min}\left(\frac{|\omega|}{\tilde{\rho}_1^3}\right), \quad \Psi_2(\omega) := \tilde{\rho}_2 \Psi_{\min}\left(\frac{|\omega|}{\tilde{\rho}_2^3}\right)$$

with  $\tilde{\rho}_1 \sim \rho(\gamma_1)$  and  $\tilde{\rho}_2 \sim \rho(\gamma_2)$  (cf. Corollary 9.4 in [3]). Thus, we obtain either  $|\omega| \sim 1$  or  $|\omega| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$ . This completes the proof of (7.10) and hence the one of Proposition 7.6.  $\square$

Finally, we use Proposition 7.6 and a Taylor expansion of  $\rho$  around a nonzero local minimum  $\tau_0$  to obtain the stronger conclusions of Theorem 7.2.

*Proof of Theorem 7.2.* We start with the proof of part (i). Let  $\tau_0 \in \text{supp } \rho \cap I_\theta$  satisfy the conditions of Theorem 7.2 (i). Then, by Proposition 6.1, the conditions of Proposition 7.6 (i) are fulfilled and all conclusions in Theorem 7.2 (i) apart from the case  $|\omega| \lesssim \rho(\tau_0)^{7/2}$  in (7.5c) follow from Proposition 7.6 (i) and (7.4b).

For the proof of the missing case, we fix a local minimum  $\tau_0 \in \text{supp } \rho \cap I_\theta$  of  $\rho$  such that  $\rho(\tau_0) \leq \rho_*$ . We set  $\rho := \rho(\tau_0)$ . Owing to the 1/3-Hölder continuity of  $\rho$  by Proposition 4.7, there is  $\varepsilon \sim 1$  such that  $\rho(\tau_0 + \omega) \sim \rho$  if  $|\omega| \leq \varepsilon \rho^3$ . In particular,  $\rho(\tau_0 + \omega) > 0$  and using Lemma 5.7 with  $k = 2, 3$  to compute the second order Taylor expansion of  $\rho$  around  $\tau_0$  yields

$$f_{\tau_0}(\omega) := \rho(\tau_0 + \omega) - \rho(\tau_0) = \frac{c}{\rho^5} \omega^2 + \mathcal{O}\left(\frac{|\omega|^3}{\rho^8}\right) \tag{7.64}$$

for all  $\omega \in \mathbb{R}$  satisfying  $|\omega| \leq \varepsilon \rho^3$ , where  $c = c(\tau_0)$  satisfies  $0 \leq c \lesssim 1$ . On the other hand,  $\tau_0$  is a shape regular point by Proposition 6.1 and a nonzero local minimum of  $\rho$ . Hence, Proposition 7.6 (i) (d) implies

$$f_{\tau_0}(\omega) = \rho \Psi_{\min}\left(\Gamma \frac{\omega}{\rho^3}\right) + \mathcal{O}\left(\frac{|\omega|}{\rho}\right) = \frac{\Gamma^2}{18\rho^5} \omega^2 + \mathcal{O}\left(\frac{|\omega|^3}{\rho^8} + \frac{|\omega|}{\rho}\right) \tag{7.65}$$

for  $|\omega| \leq \varepsilon \rho^3$ , where  $\Gamma = \Gamma(\tau_0)$ . Here, we also used the second order Taylor expansion of  $\Psi_{\min}$  defined in (7.1b) in the second step. Note that  $\Gamma \sim 1$  since  $\psi + \sigma^2 \sim 1$  by (5.35) and  $|\sigma| \lesssim \rho^2$  by Lemma 7.14.

We compare (7.64) and (7.65) and conclude

$$\frac{c}{\rho^5} \omega^2 = \frac{\Gamma^2}{18\rho^5} \omega^2 + \mathcal{O}\left(\frac{|\omega|^3}{\rho^8} + \frac{|\omega|}{\rho}\right)$$

for  $|\omega| \leq \varepsilon \rho^3$ . Choosing  $\omega = \rho^{7/2}$  and solving for  $c$  yield

$$c = \frac{\Gamma^2}{18} + \mathcal{O}(\rho^{1/2}). \tag{7.66}$$

By starting from the expansion of  $f_{\tau_0}$  in (7.64), using the Taylor expansion of  $\Psi_{\min}$  and (7.4b), we obtain (7.5c) in the last missing regime  $|\omega| \lesssim \rho^{7/2}$ .

We now turn to the proof of (ii) of Theorem 7.2. By Proposition 6.1, the conditions of Proposition 7.6 (ii) are satisfied on  $I' := I \cap [-3\kappa, 3\kappa]$ , where  $\kappa := \|a\| + 2\|S\|^{1/2}$ . Since  $\|a\| \lesssim 1$  and  $\|S\| \leq \|S\|_{2 \rightarrow \|\cdot\|} \lesssim 1$  by Assumptions 4.5, we have  $|I'| \lesssim 1$ . Moreover,  $\text{supp } \rho \subset I'$  by (2.5a). Hence, by Proposition 7.6, it suffices to estimate the distance  $|\gamma_1 - \gamma_2|$ , where  $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$  satisfy  $\gamma_1 \neq \gamma_2$ . Let  $\gamma_1, \gamma_2 \in \mathbb{M}_{\rho_*}$ . By (7.10) in Proposition 7.6 (ii), we know a dichotomy: either  $|\gamma_1 - \gamma_2| \gtrsim 1$  or  $|\gamma_1 - \gamma_2| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$ . For  $\gamma_1 \neq \gamma_2$ , we now exclude the second case by using the expansions obtained in the proof of (i). If  $\rho_* \sim 1$  is chosen sufficiently small then  $c(\gamma_1) \sim 1$  and  $c(\gamma_2) \sim 1$  by (7.66). Hence, by assuming  $|\gamma_1 - \gamma_2| \lesssim \min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$ , we obtain  $\rho(\gamma_2) > \rho(\gamma_1)$  from the expansion of  $f_{\tau_0}(\omega)$  in (7.64) with  $\tau_0 = \gamma_1$  and  $\omega = \gamma_2 - \gamma_1$ . Similarly, as  $c(\gamma_2) \sim 1$ , the expansion of  $f_{\tau_0}(\omega)$  in (7.64) with  $\tau_0 = \gamma_2$  and  $\omega = \gamma_1 - \gamma_2$  implies  $\rho(\gamma_1) > \rho(\gamma_2)$ . This is a contradiction. Therefore, the distance of two small local minima of  $\rho$  is much bigger than  $\min\{\rho(\gamma_1), \rho(\gamma_2)\}^4$  and the dichotomy above completes the proof of (ii).  $\square$

7.6 CHARACTERISATIONS OF A REGULAR EDGE

In this subsection, we introduce the concept of *regular edges* of the self-consistent support and give several equivalent characterisations relying on the cubic analysis of the previous sections. We assume that  $S$  is flat and  $a$  is bounded, i.e., that (3.10) is satisfied. In particular, owing to Proposition 2.3, there is a Hölder continuous probability density  $\rho: \mathbb{R} \rightarrow [0, \infty)$  such that

$$\langle m(z) \rangle = \int_{\mathbb{R}} \frac{\rho(\tau)}{\tau - z} d\tau,$$

where  $m$  is the solution to the Dyson equation, (2.3). We now define regular edges of  $\rho$  as in [8].

DEFINITION 7.17 (Regular edge). We call  $\tau_0 \in \partial \text{supp } \rho$  a *regular edge* if the limit

$$\lim_{\text{supp } \rho \ni \tau \rightarrow \tau_0} \frac{\rho(\tau)}{\sqrt{|\tau - \tau_0|}} = \frac{\gamma_{\text{edge}}^{3/2}}{\pi}$$

exists for some  $\gamma_{\text{edge}}$  that satisfies  $0 < c_* \leq \gamma_{\text{edge}} \leq c^* < \infty$  for some constants  $c_*$  and  $c^*$ .

The following proposition provides several equivalent characterisations of a regular edge.

PROPOSITION 7.18 (Characterisations of a regular edge). *Let  $a$  and  $S$  satisfy (3.10) and  $m$  be the solution of the corresponding Dyson equation, (2.3). Suppose for some  $\tau_0 \in \partial \text{supp } \rho$ , there are  $m_* > 0$  and  $\delta > 0$  such that*

$$\|m(\tau + i\eta)\| \leq m_* \tag{7.67}$$

for all  $\tau \in [\tau_0 - \delta, \tau_0 + \delta]$  and  $\eta \in (0, \delta]$ . We set  $\sigma := \sigma(\tau_0)$ . Then the following statements are equivalent:

- (i) The point  $\tau_0$  is a regular edge of  $\rho$ .
- (ii) There are  $0 < c_* < c^* < \infty$  such that

$$c_* \leq \liminf_{\text{supp } \rho \ni \tau \rightarrow \tau_0} \frac{\rho(\tau)}{\sqrt{|\tau - \tau_0|}} \leq \limsup_{\text{supp } \rho \ni \tau \rightarrow \tau_0} \frac{\rho(\tau)}{\sqrt{|\tau - \tau_0|}} \leq c^*$$

- (iii) There are positive constants  $\sigma_*$  and  $\sigma^*$  such that

$$\sigma_* \leq |\sigma| \leq \sigma^*.$$

- (iv) There is  $\delta_* > 0$  such that

$$\rho(\tau_0 + \omega) = \begin{cases} \frac{\pi^{1/2}}{|\sigma|^{1/2}}|\omega|^{1/2} + \mathcal{O}(|\omega|), & \text{if } \text{sign } \omega = \text{sign } \sigma, \\ 0, & \text{if } \text{sign } \omega = -\text{sign } \sigma, \end{cases}$$

for all  $\omega \in [-\delta_*, \delta_*]$ . In particular, we have  $\gamma_{\text{edge}} = \pi/|\sigma|^{1/3}$ .

- (v) There is  $\delta_{\text{gap}} > 0$  such that

$$\rho(\tau) = 0$$

for all  $\tau \in [\tau_0, \tau_0 + \delta_{\text{gap}}]$  or for all  $\tau \in [\tau_0 - \delta_{\text{gap}}, \tau_0]$ .

All constants in (i) – (v) depend effectively on each other as well as possibly  $c_1, c_2, c_3$  from (3.10) as well as  $\delta$  and  $m_*$  from (7.67).

In our recent work [8] on the universality of the local eigenvalue statistics at regular edges parts of Proposition 7.18 have already been proven. In fact, in Theorem 4.1 of [8], we showed that (i) implies (iii) and (iv). The new implications in Proposition 7.18, however, require the cubic shape analysis of the previous subsections which was not available in [8]. Using our preceding analysis, the proof of Proposition 7.18 is quite short. In the proof, the comparison relation  $\sim$  is understood with respect to  $c_1, c_2, c_3$  from (3.10) as well as  $\delta$  and  $m_*$  from (7.67).

*Proof.* For the entire proof, we remark that, by Lemma 4.8 (ii), the conditions of Proposition 6.1 are satisfied. Moreover,  $\rho(\tau_0) = 0$  due to the continuity of  $\rho$  and  $\tau_0 \in \partial \text{supp } \rho$ . Before establishing the equivalence of (i) – (v), we show that  $\sigma \neq 0$  and there is  $c \sim 1$ , depending only on the constants in (3.10) as well as  $\delta$  and  $m_*$ , such that

$$\rho(\tau_0 + \omega) = \begin{cases} \frac{\pi^{1/2}}{|\sigma|^{1/2}}|\omega|^{1/2} + \mathcal{O}\left(\frac{|\omega|}{|\sigma|^2}\right), & \text{if } \text{sign } \omega = \text{sign } \sigma, \\ 0, & \text{if } \text{sign } \omega = -\text{sign } \sigma, \end{cases} \tag{7.68}$$

for all  $\omega \in [-c|\sigma|^3, c|\sigma|^3]$ .



By Proposition 6.1, we find  $\delta_0 \sim 1$ , depending only on the constants in (3.10) as well as  $\delta$  and  $m_*$ , such that taking the imaginary part of (6.1) and applying  $\langle \cdot \rangle$  to the result yield

$$\begin{aligned} \rho(\tau_0 + \omega) &= \text{Im}(\Theta(\omega)\pi^{-1}\langle b \rangle) + \pi^{-1}\langle \text{Im } r(\omega) \rangle \\ &= \text{Im } \Theta(\omega) + \mathcal{O}((|\Theta(\omega)| + |\omega|)\text{Im } \Theta(\omega)) \end{aligned} \tag{7.69}$$

for  $|\omega| \leq \delta_0$ . Here, we used  $\langle b \rangle = \pi$  by (7.62) in the proof of Proposition 7.6 as well as the third bound in (6.8) in the second step.

By Proposition 6.1 the assumptions of Theorem 7.7 (ii) are satisfied with  $\kappa = \pi$ . Hence, from Theorem 7.7 (ii) (a), (7.69) and  $|\Theta(\omega)| \lesssim |\omega|^{1/3}$  by (6.7a), we conclude that  $\sigma \neq 0$  as  $\tau_0 \in \partial \text{supp } \rho$ . From (7.69) and Lemma 7.13, we, thus, conclude (7.68) as  $|\sigma| \lesssim 1$ ,  $|\Theta(\omega)| \lesssim |\omega/\sigma|^{1/2}$  by Lemma 7.13 and, hence,  $|\nu(\omega)| \lesssim |\Theta(\omega)| + |\omega| \lesssim |\omega/\sigma|^{1/2}$  by the first bound in (6.5). This completes the proof of (7.68).

We now show that the statements (i) – (v) are equivalent. Trivially, (i) implies (ii). Moreover, if (ii) holds true then (7.68) yields (iii). Clearly, (iv) is implied by (iii) due to (7.68). Furthermore, (v) is trivially satisfied if (iv) holds true. We now prove that (v) implies (iii). By Proposition 6.1,  $\tau_0$  is a shape regular point. Thus, (iii) is a consequence of (v) by Lemma 7.15. Finally, (iii) implies (i) due to (7.68). This completes the proof of Proposition 7.18.  $\square$

8 BAND MASS FORMULA – PROOF OF PROPOSITION 2.6

Before proving Proposition 2.6, we state an auxiliary lemma which will be proven at the end of this section.

LEMMA 8.1. *Let  $(a, S)$  be a data pair,  $m$  the solution of the associated Dyson equation (2.3) and  $\rho$  the corresponding self-consistent density of states from Definition 2.2. We assume  $\|a\| \leq k_0$  and  $S[x] \leq k_1\langle x \rangle \mathbf{1}$  for all  $x \in \overline{\mathcal{A}}_+$  and for some  $k_0, k_1 > 0$ . Then we have*

(i) *If  $\tau \in \mathbb{R} \setminus \text{supp } \rho$  then there is  $m(\tau) = m(\tau)^* \in \mathcal{A}$  such that*

$$\lim_{\eta \downarrow 0} \|m(\tau + i\eta) - m(\tau)\| = 0.$$

*Moreover,  $m(\tau)$  is invertible and satisfies the Dyson equation, (2.3), at  $z = \tau$ . There is  $C > 0$ , depending only on  $k_0, k_1$  and  $\text{dist}(\tau, \text{supp } \rho)$ , such that  $\|m(\tau)\| \leq C$  and  $\|(\text{Id} - (1 - t)C_{m(\tau)}S)^{-1}\| \leq C$  all  $t \in [0, 1]$ .*

(ii) *Fix  $\tau \in \mathbb{R} \setminus \text{supp } \rho$ . Let  $m_t$  be the solution of (2.3) associated to the data pair*

$$(a_t, S_t) := (a - tS[m(\tau)], (1 - t)S)$$

*for  $t \in [0, 1]$  and  $\rho_t$  the corresponding self-consistent density of states. Then, for any  $t \in [0, 1]$ , we have*

$$\lim_{\eta \downarrow 0} \|m_t(\tau + i\eta) - m(\tau)\| = 0. \tag{8.1}$$

Moreover, there is  $c > 0$ , depending only on  $k_0, k_1$  and  $\text{dist}(\tau, \text{supp } \rho)$ , such that  $\text{dist}(\tau, \text{supp } \rho_t) \geq c$  for all  $t \in [0, 1]$ .

*Proof of Proposition 2.6.* We start with the proof of (i) and notice that the existence of  $m(\tau)$  has been proven in Lemma 8.1 (i). In order to verify (2.10), we consider the continuous flow of data pairs  $(a_t, S_t)$  from Lemma 8.1 (ii) and the corresponding solutions  $m_t$  of the Dyson equation, (2.3), and prove

$$\rho_t((-\infty, \tau)) = \langle \mathbf{1}_{(-\infty, 0)}(m_t(\tau)) \rangle \tag{8.2}$$

for all  $t \in [0, 1]$ . Note that  $\text{dist}(\tau, \text{supp } \rho_t) \geq c$  for all  $t \in [0, 1]$  by Lemma 8.1 (ii).

In particular, by Lemma 8.1 (ii),  $m_t(\tau) = m(\tau)$  is constant along the flow, and with it the right-hand side of (8.2). The identity (8.2) obviously holds for  $t = 1$ , because  $m_1(z) = (a - Sm(\tau) - z)^{-1}$  is the resolvent of a self-adjoint element and  $m(\tau)$  satisfies (2.3) at  $z = \tau$  by Lemma 8.1 (i). Thus it remains to verify that the left-hand side of (8.2) stays constant along the flow as well. This will show (8.2) for  $t = 0$  which is (2.10).

First we conclude from the Stieltjes transform representation (2.4) of  $m_t$  that

$$\rho_t((-\infty, \tau)) = -\frac{1}{2\pi i} \oint \langle m_t(z) \rangle dz, \tag{8.3}$$

where the contour encircles  $[\min \text{supp } \rho_t, \tau)$  counterclockwise, passing through the real line only at  $\tau$  and to the left of  $\min \text{supp } \rho_t$ , and we extended  $m_t(z)$  analytically to a neighbourhood of the contour (set  $m_t(\bar{z}) := m_t(z)^*$  for  $z \in \mathbb{H}$  and use Lemma D.1 (iv) close to the real axis to conclude analyticity in a neighbourhood of the contour).

We now show that the left-hand side of (8.3) does not change along the flow. Indeed, differentiating the right-hand side of (8.3) with respect to  $t$  and writing  $m_t = m_t(z)$  yield

$$\begin{aligned} \frac{d}{dt} \oint \langle m_t(z) \rangle dz &= \oint \langle \partial_t m_t(z) \rangle dz \\ &= \oint \langle (C_{m_t^*}^{-1} - S_t)^{-1} [\mathbf{1}], S[m(\tau)] - S[m_t] \rangle dz \\ &= \oint \langle (\partial_z m_t)(S[m(\tau)] - S[m_t]) \rangle dz \\ &= \oint \partial_z \left( \langle m_t S[m(\tau)] \rangle - \frac{1}{2} \langle m_t S[m_t] \rangle \right) dz = 0. \end{aligned}$$

Here, in the second step, we used  $\partial_t m_t(z) = (C_{m_t^*}^{-1} - S_t)^{-1} [-S[m_t] - S[m(\tau)]]$  obtained by differentiating the Dyson equation, (2.3), for the data pair  $(a_t, S_t)$  defined in Lemma 8.1 (ii) and the definition of the scalar product, (2.1). In the third step, we employed  $(C_{m_t^*}^{-1} - S_t)^{-1} [\mathbf{1}] = (\partial_z m_t(z))^*$  which follows from differentiating the Dyson equation, (2.3), for the data pair  $(a_t, S_t)$  with respect

to  $z$ . Finally, we used that  $m_t$  is holomorphic in a neighbourhood of the contour. This completes the proof of (i) of Proposition 2.6.

For the proof of (ii), we fix a connected component  $J$  of  $\text{supp } \rho$ . Let  $\tau_1, \tau_2 \in \mathbb{R} \setminus \text{supp } \rho$  satisfy  $\tau_1 < \tau_2$  and  $[\tau_1, \tau_2] \cap \text{supp } \rho = J$ . By (2.10), we have

$$n\rho(J) = n(\rho((-\infty, \tau_2)) - \rho((-\infty, \tau_1))) = \text{Tr}(P_2) - \text{Tr}(P_1) = \text{rank}P_2 - \text{rank}P_1,$$

where  $P_i := \pi(\mathbf{1}_{(-\infty, 0)}(m(\tau_i)))$  are orthogonal projections in  $\mathbb{C}^{n \times n}$  for  $i = 1, 2$ . Hence,  $n\rho(J) \in \mathbb{Z}$ . Since  $0 < n\rho(J) \leq n$  by definition of  $\text{supp } \rho$ , we conclude  $n\rho(J) \in \{1, \dots, n\}$ , which immediately implies that  $\text{supp } \rho$  has at most  $n$  connected components. This completes the proof of Proposition 2.6.  $\square$

*Proof of Lemma 8.1.* In part (i), the existence of the limit  $m(\tau) \in \mathcal{A}$  follows immediately from the implication (v)  $\Rightarrow$  (iii) of Lemma D.1. The invertibility of  $m(\tau)$  can be seen by multiplying (2.3) at  $z = \tau + i\eta$  by  $m(\tau + i\eta)$  and taking the limit  $\eta \downarrow 0$ . This also implies that  $m(\tau)$  satisfies (2.3) at  $z = \tau$ . In order to bound  $\|(\text{Id} - (1 - t)C_{m(\tau)}S)^{-1}\|$ , we recall the definitions of  $q, u$  and  $F$  from (3.1) and (3.4), respectively, and compute

$$\text{Id} - (1 - t)C_mS = C_{q^*, q}(\text{Id} - (1 - t)C_uF)C_{q^*, q}^{-1}$$

for  $m = m(z)$  with  $z \in \mathbb{H}$ . Hence, by (D.1), Lemma 4.8 (i) and Lemma B.2, we obtain  $\|(\text{Id} - (1 - t)C_mS)^{-1}\| \lesssim (1 - (1 - t)\|F\|_2)^{-1} \leq (1 - \|F\|_2)^{-1} \leq C$  for all  $z \in \tau + iN$ , where the set  $N \subset (0, 1]$  with an accumulation point at 0 is given in Lemma D.1 (ii). Taking the limit  $\eta \downarrow 0$  under the constraint  $\eta \in N$  and possibly increasing  $C$  yield the desired uniform bound. This completes the proof of (i).

We start the proof of (ii) with an auxiliary result. Similarly as in the proof of (i), we see that  $\text{Id} - (1 - t)C_{m^*, m}S$  is invertible for  $m = m(z)$ ,  $z \in \tau + iN$  with  $N$  as before. Since  $\|F(z)\|_2 \leq 1 - C^{-1}$  for  $z \in \tau + iN$  as in the proof of (i), Lemma B.3 implies that  $(\text{Id} - (1 - t)C_{u^*, u}F)^{-1}$ ,  $F = F(z)$ , and, thus,  $(\text{Id} - (1 - t)C_{m^*, m}S)^{-1} = C_{q^*, q}(\text{Id} - (1 - t)C_{u^*, u}F)^{-1}C_{q^*, q}^{-1}$  are positivity-preserving for  $z \in \tau + iN$ . Taking the limit  $\eta = \text{Im } z \downarrow 0$  in  $N$  shows that  $(\text{Id} - (1 - t)C_{m(\tau)}S)^{-1}$  is positivity-preserving for any  $t \in [0, 1]$ . Moreover, (B.12) with  $x = \mathbf{1}$  yields

$$(\text{Id} - (1 - t)C_{m^*, m}S)^{-1}[\mathbf{1}] = C_{q^*, q}(\text{Id} - (1 - t)C_{u^*, u}F)^{-1}C_{q^*, q}^{-1}[\mathbf{1}] \geq \mathbf{1}. \tag{8.4}$$

Since (8.4) holds true uniformly for  $z \in \tau + iN$  and  $t \in [0, 1]$ , taking the limit  $\eta = \text{Im } z \downarrow 0$  in  $N$ , we obtain

$$(\text{Id} - (1 - t)C_{m(\tau)}S)^{-1}[\mathbf{1}] \geq \mathbf{1} \tag{8.5}$$

for all  $t \in [0, 1]$ .

We fix  $t \in [0, 1]$ . We write  $m = m(\tau)$  and define  $\Phi_t : \mathcal{A} \times \mathbb{R} \rightarrow \mathcal{A}$  through

$$\begin{aligned} \Phi_t(\Delta, \eta) := & (\text{Id} - (1 - t)C_mS)[\Delta] - \frac{i\eta}{2}(m\Delta + \Delta m) - i\eta m^2 \\ & - \frac{1}{2}(1 - t)(\Delta S[\Delta]m + mS[\Delta]\Delta) \end{aligned}$$

In order to show (8.1), we apply the implicit function theorem (see e.g. Lemma D.4 below) to  $\Phi_t(\Delta, \eta) = 0$ . It is applicable as  $\Phi_t(0, 0) = 0$  and  $\partial_1 \Phi_t(0, 0) = \text{Id} - (1 - t)C_m S$  which is invertible by (i). Hence, we obtain an  $\varepsilon > 0$  and a continuously differentiable function  $\Delta_t: (-\varepsilon, \varepsilon) \rightarrow \mathcal{A}$  such that  $\Phi_t(\Delta_t(\eta), \eta) = 0$  for all  $\eta \in (-\varepsilon, \varepsilon)$  and  $\Delta_t(0) = 0$ . We now show that  $\Delta_t(\eta) + m(\tau) = m_t(\tau + i\eta)$  for all sufficiently small  $\eta > 0$  by appealing to the uniqueness of the solution to the Dyson equation, (2.3), with the choice  $z = \tau + i\eta$ ,  $a = a_t$  and  $S = S_t = (1 - t)S$ . In fact,  $m = m(\tau)$  and  $m_t = m_t(\tau + i\eta)$  with  $\eta > 0$  satisfy the Dyson equations

$$-m^{-1} = \tau - a + S[m], \quad -m_t^{-1} = \tau + i\eta - a + tS[m] + (1 - t)S[m_t] \quad (8.6)$$

and  $m_t$  is the unique solution of the second equation under the constraint  $\text{Im } m_t > 0$  (compare the remarks around (2.3)). A straightforward computation using the first relation in (8.6) and  $\Phi_t(\Delta_t(\eta), \eta) = 0$  reveals that  $\Delta_t(\eta) + m(\tau)$  solves the second equation in (8.6) for  $m_t$ . Moreover, differentiating  $\Phi_t(\Delta_t(\eta), \eta) = 0$  with respect to  $\eta$  at  $\eta = 0$  yields

$$\begin{aligned} \partial_\eta \text{Im } \Delta_t(\eta = 0) &= (\text{Id} - (1 - t)C_m S)^{-1}[m^2] \\ &\geq \|m^{-1}\|^{-2}(\text{Id} - (1 - t)C_m S)^{-1}[\mathbf{1}] \geq \|m^{-1}\|^{-2}\mathbf{1}. \end{aligned}$$

Here, we used that  $(\text{Id} - (1 - t)C_m S)^{-1}$  is compatible with the involution  $*$  and  $m = m^*$  in the first step. Then we employed the invertibility of  $m$ ,  $m^2 \geq \|m^{-1}\|^{-2}\mathbf{1}$  and the positivity-preserving property of  $(\text{Id} - (1 - t)C_m S)^{-1}$  in the second step and, finally, (8.5) in the last step. Hence,  $\text{Im}(\Delta_t(\eta) + m(\tau)) = \text{Im } \Delta_t(\eta) > 0$  for all sufficiently small  $\eta > 0$ . The uniqueness of the solution to the Dyson equation for  $m_t$ , the second relation in (8.6), implies  $\Delta_t(\eta) + m(\tau) = m_t(\tau + i\eta)$  for all sufficiently small  $\eta > 0$  and all  $t \in [0, 1]$ . Therefore, the continuity of  $\Delta_t$  as a function of  $\eta$ ,  $\Delta_t(\eta) \rightarrow \Delta_t(0) = 0$ , yields (8.1).

We now conclude from the implication (iii)  $\Rightarrow$  (v) of Lemma D.1 that  $\text{dist}(\tau, \text{supp } \rho_t) \geq \varepsilon$  for some  $\varepsilon > 0$ . Lemma D.1 is applicable since  $\|a_t\| \leq k_0 + k_1 C$  (cf. Lemma B.2 (i) and Lemma 8.1 (i)) and  $S_t[x] \leq S[x] \leq k_1 \langle x \rangle \mathbf{1}$  for all  $x \in \overline{\mathcal{A}}_+$ . For any  $t \in [0, 1]$ , statement (iii) in Lemma D.1 holds true with the same  $m = m(\tau)$  by (8.1) and  $S$  replaced by  $S_t = (1 - t)S$ . By (i),  $\|m\| \leq C$  and  $\|(\text{Id} - (1 - t)C_m S)^{-1}\| \leq C$  for all  $t \in [0, 1]$ . Hence, owing to Lemma D.1 (v), there is  $\varepsilon > 0$ , depending only on  $k_0, k_1$  and  $\text{dist}(\tau, \text{supp } \rho)$ , such that  $\text{dist}(\tau, \text{supp } \rho_t) \geq \varepsilon$  for all  $t \in [0, 1]$ . Here,  $\varepsilon$  depends only on  $k_0, k_1$  and  $\text{dist}(\tau, \text{supp } \rho)$  due to the exclusive dependence of  $C$  from (i) on the quantities and the effective dependence of the constants in Lemma D.1 on each other (see final remark in Lemma D.1). The uniformity of  $\varepsilon$  in  $t$  is a consequence of the uniformity of  $C$  from (i) in  $t$ . This completes the proof of Lemma 8.1.  $\square$

### 9 DYSON EQUATION FOR KRONECKER RANDOM MATRICES

In this section we present an application of the theory developed in this work to Kronecker random matrices, i.e., block correlated random matrices with

variance profiles within the blocks, and their limits. Kronecker matrices were introduced in [7], where several applications of them were outlined. In Lemma 9.1 and Lemma 9.3 below, we will provide some sufficient checkable conditions that ensure the flatness of  $S$  and the boundedness of  $\|m(z)\|$ , the main assumptions of Proposition 2.4, Theorem 2.5 and Theorem 7.1, for the self-consistent density of states of Kronecker random matrices.

9.1 THE KRONECKER SETUP

We fix  $K \in \mathbb{N}$  and a probability space  $(\mathfrak{X}, \pi)$  that we view as a possibly infinite set of indices. We consider the von Neumann algebra

$$\mathcal{A} = \mathbb{C}^{K \times K} \otimes L^\infty(\mathfrak{X}), \tag{9.1}$$

with the tracial state

$$\langle \kappa \otimes f \rangle = \frac{\text{Tr } \kappa}{K} \int_{\mathfrak{X}} f d\pi.$$

For  $K = 1$  the algebra  $\mathcal{A}$  is commutative and this setup was previously considered in [3, 1]. Now let  $(\alpha_\mu)_{\mu=1}^{\ell_1}, (\beta_\nu)_{\nu=1}^{\ell_2}$  be families of matrices in  $\mathbb{C}^{K \times K}$  with  $\alpha_\mu = \alpha_\mu^*$  self-adjoint and let  $(s^\mu)_{\mu=1}^{\ell_1}, (t^\nu)_{\nu=1}^{\ell_2}$  be families of non-negative bounded functions in  $L^\infty(\mathfrak{X}^2)$  and suppose that all  $s^\mu$  are symmetric,  $s^\mu(x, y) = s^\mu(y, x)$ . Then we define the self-energy operator  $S : \mathcal{A} \rightarrow \mathcal{A}$  as

$$S(\kappa \otimes f) := \sum_{\mu=1}^{\ell_1} \alpha_\mu \kappa \alpha_\mu \otimes S_\mu f + \sum_{\nu=1}^{\ell_2} (\beta_\nu \kappa \beta_\nu^* \otimes T_\nu f + \beta_\nu^* \kappa \beta_\nu \otimes T_\nu^* f), \tag{9.2}$$

where the bounded operators  $S_\mu, T_\nu, T_\nu^* : L^\infty(\mathfrak{X}) \rightarrow L^\infty(\mathfrak{X})$  act as

$$\begin{aligned} (S_\mu f)(x) &= \int_{\mathfrak{X}} s^\mu(x, y) f(y) \pi(dy), & (T_\nu f)(x) &= \int_{\mathfrak{X}} t^\nu(x, y) f(y) \pi(dy), \\ (T_\nu^* f)(x) &= \int_{\mathfrak{X}} t^\nu(y, x) f(y) \pi(dy). \end{aligned}$$

Furthermore we fix a self-adjoint  $a = a^* \in \mathcal{A}$ . With these data we will consider the Dyson equation, (2.3).

The following lemma provides sufficient conditions that ensure flatness of  $S$  and boundedness of  $\|m(z)\|$  uniformly in  $z$  up to the real line. We begin with some preparations. We use the notation  $x \mapsto v_x$  for  $x \in \mathfrak{X}$  and an element  $v \in \mathbb{C}^{K \times K} \otimes L^\infty(\mathfrak{X})$ , interpreting it as a function on  $\mathfrak{X}$  with values in  $\mathbb{C}^{K \times K}$ . We also introduce the functions  $\gamma \in L^\infty(\mathfrak{X}^2)$  via

$$\begin{aligned} \gamma(x, y) := & \left( \int_{\mathfrak{X}} (|s^\mu(x, u) - s^\mu(y, u)|^2 + |t^\nu(x, u) - t^\nu(y, u)|^2 \right. \\ & \left. + |t^\nu(u, x) - t^\nu(u, y)|^2) \pi(du) \right)^{1/2} \end{aligned} \tag{9.3}$$

and  $\Gamma : (0, \infty)^2 \rightarrow L^\infty(\mathfrak{X}), (\Lambda, \tau) \mapsto \Gamma_{\Lambda, \cdot}(\tau)$  through

$$\Gamma_{\Lambda, x}(\tau) := \left( \int_{\mathfrak{X}} \left( \frac{1}{\tau} + \|a_x - a_y\| + \gamma(x, y)\Lambda \right)^{-2} \pi(dy) \right)^{1/2}. \tag{9.4}$$

Here, we denoted by  $\|\cdot\|$  the operator norm on  $\mathbb{C}^{K \times K}$  induced by the Euclidean norm on  $\mathbb{C}^K$ . The two functions  $\gamma$  and  $\Gamma$  will be important to quantify the modulus of continuity of the data  $(a, S)$ .

LEMMA 9.1. *Let  $m$  be the solution of the Dyson equation, (2.3), on the von Neumann algebra  $\mathcal{A}$  from (9.1) associated to the data  $(a, S)$  with  $S$  defined as in (9.2).*

- (i) *Define  $\Gamma(\tau) := C_{\text{Kr}} \operatorname{ess\,inf}_x \Gamma_{1, x}(\tau)$  with  $C_{\text{Kr}} := (4 + 4K(\ell_1 + \ell_2) \max_{\mu, \nu} (\|\alpha_\mu\|^2 + \|\beta_\nu\|^2))^{1/2}$ , where  $\Gamma_{\Lambda, x}(\tau)$  was introduced in (9.4) and assume that for some  $z \in \mathbb{H}$  the  $L^2$ -upper bound  $\|m(z)\|_2 \leq \Lambda$  for some  $\Lambda \geq 1$  is satisfied. Then we have the uniform upper bound*

$$\|m(z)\| \leq \frac{\Gamma^{-1}(\Lambda^2)}{\Lambda}, \tag{9.5}$$

where we interpret the right-hand side as  $\infty$  if  $\Lambda$  is not in the range of the strictly monotonously increasing function  $\Gamma$ .

- (ii) *Suppose that the kernels of the operators  $S^\mu$  and  $T^\nu$ , used to define  $S$  in (9.2), are bounded from below, i.e.,  $\operatorname{ess\,inf}_{x, y} s^\mu(x, y) > 0$  and  $\operatorname{ess\,inf}_{x, y} t^\nu(x, y) > 0$ . Suppose further that*

$$\inf_{\kappa} \frac{1}{\operatorname{Tr} \kappa} \left( \sum_{\mu=1}^{\ell_1} \alpha_\mu \kappa \alpha_\mu + \sum_{\nu=1}^{\ell_2} (\beta_\nu \kappa \beta_\nu^* + \beta_\nu^* \kappa \beta_\nu) \right) > 0, \tag{9.6}$$

where the infimum is taken over all positive definite  $\kappa \in \mathbb{C}^{K \times K}$ . Then  $S$  is flat, i.e.,  $S \in \Sigma_{\text{flat}}$  (cf. (2.2b)).

- (iii) *Let  $S$  be flat, hence,  $\Lambda := 1 + \sup_{z \in \mathbb{H}} \|m(z)\|_2 < \infty$ . Then (9.5) holds true with this  $\Lambda$ .*

- (iv) *If  $a = 0$  then, for each  $\varepsilon > 0$ , (9.5) holds true on  $|z| \geq \varepsilon$  with  $\Lambda := 1 + 2\varepsilon^{-1}$ .*

*Proof of Lemma 9.1.* We adapt the proof of Proposition 6.6 in [3] to our non-commutative setting in order to prove (i). Recall the definition of  $\gamma(x, y)$  in (9.3). Estimating the norm  $\|m\|_2$  from below, we find

$$\begin{aligned} \|m\|_2^2 &= \frac{1}{K} \operatorname{Tr} \int \frac{\pi(dy)}{m_y^{-1}(m_x^*)^{-1}} \\ &\geq \operatorname{Tr} \int_{\mathfrak{X}} \frac{C_{\text{Kr}}^2 \pi(dy)}{m_x^{-1}(m_x^*)^{-1} + \|a_x - a_y\|^2 + \gamma(x, y)^2 \|m\|_2^2} \\ &\geq C_{\text{Kr}}^2 (\Gamma_{\|m\|_2, x}(\|m_x\|))^2, \end{aligned} \tag{9.7}$$

for  $\pi$ -almost all  $x \in \mathfrak{X}$ , where we used

$$\begin{aligned} \frac{1}{4}m_y^{-1}(m_y^*)^{-1} &\leq m_x^{-1}(m_x^*)^{-1} + (a_y - a_x)(a_y - a_x)^* \\ &\quad + ((Sm)_x - (Sm)_y)((Sm)_x - (Sm)_y)^* \\ &\leq m_x^{-1}(m_x^*)^{-1} + \|a_x - a_y\|^2 \\ &\quad + K(\ell_1 + \ell_2) \max_{\mu, \nu} (\|\alpha_\mu\|^2 + \|\beta_\nu\|^2) \gamma(x, y)^2 \|m\|_2^2. \end{aligned} \tag{9.8}$$

We conclude  $\Lambda \geq \Lambda^{-1}\Gamma(\Lambda\|m_x\|)$  for any upper bound  $\Lambda \geq 1$  on  $\|m\|_2$ . In particular, (9.5) follows.

We turn to the proof of (ii). We view a positive element  $r \in \mathcal{A}_+$  as a function  $r : [0, 1] \rightarrow \mathbb{C}^{K \times K}$  with values in positive semidefinite matrices. Then we find

$$(Sr)_x \geq c \int_{\mathfrak{X}} \left( \sum_{\mu=1}^{\ell_1} \alpha_\mu r_y \alpha_\mu + \sum_{\nu=1}^{\ell_2} (\beta_\nu r_y \beta_\nu^* + \beta_\nu^* r_y \beta_\nu) \right) \pi(dy),$$

as quadratic forms on  $\mathbb{C}^{K \times K}$  for almost every  $x \in \mathfrak{X}$ . The claim follows now immediately from (9.6). Part (iii) is a direct consequence of (i) and (ii) as well as (3.11). For the proof of part (iv), we use part (i) and (2.6) if  $a = 0$ .  $\square$

### 9.2 $N \times N$ -KRONECKER RANDOM MATRICES

As an application of the general Kronecker setup introduced above, we consider the *matrix Dyson equation* associated to Kronecker random matrices. Let  $X_\mu, Y_\nu \in \mathbb{C}^{N \times N}$  be independent centered random matrices such that  $Y_\nu = (y_{ij}^\nu)$  has independent entries and  $X_\mu = (x_{ij}^\mu)$  has independent entries up to the Hermitian symmetry constraint  $X_\mu = X_\mu^*$ . Suppose that the entries of  $\sqrt{N}X_\mu, \sqrt{N}Y_\nu$  have uniformly bounded moments,  $\mathbb{E}(|x_{ij}^\mu|^p + |y_{ij}^\nu|^p) \leq N^{-p/2}C_p$  and define their variance profiles through

$$s^\mu(i, j) := N\mathbb{E}|x_{ij}^\mu|^2, \quad t^\nu(i, j) := N\mathbb{E}|y_{ij}^\nu|^2.$$

Then we are interested in the asymptotic spectral properties of the Hermitian *Kronecker random matrix*

$$H := A + \sum_{\mu=1}^{\ell_1} \alpha_\mu \otimes X_\mu + \sum_{\nu=1}^{\ell_2} (\beta_\nu \otimes Y_\nu + \beta_\nu^* \otimes Y_\nu^*) \in \mathbb{C}^{K \times K} \otimes \mathbb{C}^{N \times N}, \tag{9.9}$$

as  $N \rightarrow \infty$ . Here the expectation matrix  $A$  is assumed to be bounded,  $\|A\| \leq C$ , and block diagonal, i.e.

$$A = \sum_{i=1}^N a_i \otimes E_{ii}, \tag{9.10}$$

with  $E_{ii} = (\delta_{il}\delta_{ik})_{l,k=1}^N \in \mathbb{C}^{N \times N}$  and  $a_i \in \mathbb{C}^{K \times K}$ . In [7] it was shown that the resolvent  $G(z) = (H - z)^{-1}$  of the Kronecker matrix  $H$  is well approximated

by the solution  $M(z)$  of a Dyson equation of Kronecker type, i.e., on the von Neumann algebra  $\mathcal{A}$  in (9.1) with self-energy  $S$  from (9.2) and  $a = A \in \mathcal{A}$ , when we choose  $\mathfrak{X} = \{1, \dots, N\}$  and  $\pi$  the uniform probability distribution. In other words,  $L^\infty(\mathfrak{X}) = \mathbb{C}^N$  with entrywise multiplication.

9.3 LIMITS OF KRONECKER RANDOM MATRICES

Now we consider limits of Kronecker random matrices  $H \in \mathbb{C}^{N \times N}$  with piecewise Hölder-continuous variance profiles as  $N \rightarrow \infty$ . In this situation we can make sense of the continuum limit for the solution  $M(z)$  of the associated matrix Dyson equation. The natural setup here is  $(\mathfrak{X}, \pi) = ([0, 1], dx)$ . We fix a partition  $(I_l)_{l=1}^L$  of  $[0, 1]$  into intervals of positive length, i.e.,  $[0, 1] = \dot{\cup}_l I_l$  and consider non-negative profile functions  $s^\mu, t^\nu : [0, 1]^2 \rightarrow \mathbb{R}$  that are Hölder-continuous with Hölder exponent  $1/2$  on each rectangle  $I_l \times I_k$ . We also fix a function  $a : [0, 1] \rightarrow \mathbb{C}^{K \times K}$  that is  $1/2$ -Hölder continuous on each  $I_l$ . In this piecewise Hölder-continuous setup the Dyson equation on  $\mathcal{A}$  with data pair  $(a, S)$  describes the asymptotic spectral properties of Kronecker random matrices with fixed variance profiles  $s^\mu$  and  $t^\nu$ , i.e., the random matrices  $H$  introduced in Subsection 9.2 if their variances are given by

$$\mathbb{E}|x_{ij}^\mu|^2 = \frac{1}{N} s^\mu\left(\frac{i}{N}, \frac{j}{N}\right), \quad \mathbb{E}|y_{ij}^\nu|^2 = \frac{1}{N} t^\nu\left(\frac{i}{N}, \frac{j}{N}\right),$$

and the matrices  $a_i$  in (9.10) by  $a_i = a(\frac{i}{N})$ .

LEMMA 9.2. *Suppose that  $a, s^\mu$  and  $t^\nu$  are piecewise Hölder-continuous with Hölder exponent  $1/2$  as described above. The empirical spectral distribution of the Kronecker random matrix  $H$ , defined in (9.9), with eigenvalues  $(\lambda_i)_{i=1}^{KN}$  converges weakly in probability to the self-consistent density of states  $\rho$  associated to the Dyson equation with data pair  $(a, S)$  as defined in (9.2), i.e., for any  $\varepsilon > 0$  and  $\varphi \in C(\mathbb{R})$  we have*

$$\mathbb{P}\left(\left|\frac{1}{KN} \sum_{i=1}^{KN} \varphi(\lambda_i) - \int_{\mathbb{R}} \varphi d\rho\right| > \varepsilon\right) \rightarrow 0, \quad N \rightarrow \infty.$$

*Proof of Lemma 9.2.* It suffices to prove convergence of the Stieltjes transforms, i.e., in probability  $\frac{1}{KN} \text{Tr}_{KN} G(z) \rightarrow \langle m(z) \rangle$  for every fixed  $z \in \mathbb{H}$ , where  $G(z) = (H - z)^{-1}$  is the resolvent of the Kronecker matrix  $H$  and  $m(z)$  is the solution to the Dyson equation with data  $(a, S)$ .

First we use the Theorem 2.7 from [7] to show that  $\frac{1}{KN} \text{Tr}_{KN} G(z) - \frac{1}{N} \sum_{i=1}^N \text{Tr}_K m_i(z) \rightarrow 0$  in probability, where  $M_N = (m_1, \dots, m_N) \in (\mathbb{C}^{K \times K})^N$  denotes the solution to a Dyson equation formulated on the von Neumann algebra  $\mathbb{C}^{K \times K} \otimes \mathbb{C}^N$  with entrywise multiplication on vectors in  $\mathbb{C}^N$  as explained in Subsection 9.2. We recall that in this setup the discrete kernels for  $S_\mu$  and  $T_\nu$  from the definition of  $S$  in (9.2) are given by  $N\mathbb{E}|x_{ij}^\mu|^2$  and  $N\mathbb{E}|y_{ij}^\nu|^2$ , respectively, and  $a = \sum_{i=1}^N a(\frac{i}{N}) \otimes e_i$ . To distinguish this discrete



data pair from the continuum limit over  $\mathbb{C}^{K \times K} \otimes L^\infty[0, 1]$ , we denote it by  $(a_N, S_N)$ . Note that in Theorem 2.7 of [7] the test functions were compactly supported in contrast to the function  $\tau \mapsto 1/(\tau - z)$  that we used here. However, by Theorem 2.4 of [7] and since the self-consistent density of states is compactly supported (cf. (2.5a) and  $\|S\| \lesssim 1$ ) no eigenvalues can be found beyond a certain bounded interval, ensuring that non compactly supported test function are allowed as well.

Now it remains to show that  $\langle M_N \rangle \rightarrow \langle m \rangle$  as  $N \rightarrow \infty$  for all  $z \in \mathbb{H}$ . For this purpose we embed  $\mathbb{C}^N$  into  $L^\infty[0, 1]$  via  $Pv := \sum_{i=1}^N v_i \mathbf{1}_{[(i-1)/N, i/N]}$ . With this identification  $M_N$  and  $m$  satisfy Dyson equations on the same space  $\mathbb{C}^{K \times K} \otimes L^\infty[0, 1]$ . Evaluating these two equations at  $z + i\eta$ , for a fixed  $z \in \mathbb{H}$  and any  $\eta \geq 0$ , and subtracting them from each other yield

$$B[\Delta] = m(S_N - S)[m]\Delta + C_m(S_N - S)[\Delta] + mS_N[\Delta]\Delta + C_m(S_N - S)[m] - m(a_N - a)\Delta - C_m[a_N - a],$$

where  $m = m(z + i\eta)$ ,  $M_N = M_N(z + i\eta)$ ,  $B = \text{Id} - C_m S$  and  $\Delta = M_N - m$ . Using the imaginary part of  $z$  we have  $\text{dist}(z + i\eta, \text{supp } \rho) \geq \text{Im } z > 0$ . By (3.22), (3.23), (3.11a) and (3.11c) of [7] we infer  $\|m\| + \|B^{-1}\|_2 \leq C$  for all  $\eta \geq 0$  with a constant  $C$  depending on  $\text{Im } z$ . Note that although the proofs in [7] were performed on  $\mathbb{C}^{N \times N}$  all estimates were uniform in  $N$  and all algebraic relations in these proof translate to the current setting on a finite von Neumann algebra. Using  $\|S_N - S\|_2 \leq \|S_N - S\|$  as well as  $\|S_N\| \leq C$  and possibly increasing  $C$ , we thus obtain

$$\|\Delta\|_2 \leq C(\Psi_N + \|\Delta\|_2^2), \quad \Psi_N := \|a_N - a\| + \|S_N - S\|,$$

where  $\Delta = \Delta(z + i\eta)$ , for all  $\eta \geq 0$ . We choose  $N_0$  sufficiently large such that  $2\Psi_N C^2 \leq 1/4$  for all  $N \geq N_0$  and define  $\eta_* := \sup\{\eta \geq 0: \|\Delta(z + i\eta)\|_2 \geq 2C\Psi_N\}$ . Since  $\|M_N\| + \|m\| \rightarrow 0$  for  $\eta \rightarrow \infty$ , we conclude  $\eta_* < \infty$ .

We now prove  $\eta_* = 0$ . For a proof by contradiction, we suppose  $\eta_* > 0$ . Then, by continuity,  $\|\Delta(\tau + i\eta_*)\|_2 = 2C\Psi_N$ . Since  $2\Psi_N C^2 \leq 1/4$ , we have  $\|\Delta(z + i\eta_*)\|_2 \leq 4C\Psi_N/3 < 2C\Psi_N = \|\Delta(z + i\eta_*)\|_2$ . From this contradiction, we conclude  $\eta_* = 0$ . Therefore, for  $N \geq N_0$ , we have

$$|M_N(z) - m(z)| \leq \|\Delta(z)\|_2 \leq 2C\Psi_N = 2C(\|S_N - S\| + \|a_N - a\|).$$

Since the right-hand side converges to zero as  $N \rightarrow \infty$ , due to the piecewise Hölder-continuity of the profile functions, and since  $z$  was arbitrary, we obtain  $\langle M_N \rangle \rightarrow \langle m \rangle$  as  $N \rightarrow \infty$  for all  $z \in \mathbb{H}$ . This completes the proof of Lemma 9.2.  $\square$

The boundedness of the solution to the Dyson equation in  $L^2$ -norm already implies uniform boundedness in the piecewise Hölder-continuous setup.

LEMMA 9.3. *Suppose that  $a$ ,  $s^u$  and  $t^v$  are piecewise 1/2-Hölder continuous and that  $\sup_{z \in \mathbb{D}} \|m(z)\|_2 < \infty$  for some domain  $\mathbb{D} \subseteq \mathbb{H}$ . Then we have the uniform bound  $\sup_{z \in \mathbb{D}} \|m(z)\| < \infty$ .*

In particular, if the random matrix  $H$  is centered, i.e.,  $a = 0$ , then  $m(z)$  is uniformly bounded as long as  $z$  is bounded away from zero; and if  $H$  is flat in the limit, i.e.,  $S$  is flat, then  $\sup_{z \in \mathbb{H}} \|m(z)\| < \infty$ .

*Proof.* By (i) of Lemma 9.1 the proof reduces to checking that  $\lim_{\tau \rightarrow \infty} \Gamma(\tau) = \infty$  for piecewise 1/2-Hölder continuous data in the special case  $(\mathfrak{X}, \pi) = ([0, 1], dx)$ . But this is clear since in that case  $\|a_x - a_y\|^2 + \gamma(x, y)^2 \leq C|x - y|$  implies that the integral in (9.4) is at least logarithmically divergent as  $\tau \rightarrow \infty$ .  $\square$

**COROLLARY 9.4** (Band mass quantization). *Let  $\rho$  be the self-consistent density of states for the Dyson equation with data pair  $(a, S)$  and  $\tau \in \mathbb{R} \setminus \text{supp } \rho$ . Then*

$$\rho((-\infty, \tau)) \in \left\{ \frac{1}{K} \sum_{l=1}^L k_l |I_l| : k_l = 1, \dots, K \right\}.$$

*In particular, in the  $L = 1$  case when  $s^\mu, t^\mu$  and  $a$  are 1/2-Hölder continuous on all of  $[0, 1]^2$  and  $[0, 1]$ , respectively, then  $\rho(J)$  is an integer multiple of  $1/K$  for every connected component  $J$  of  $\text{supp } \rho$  and there are at most  $K$  such components.*

*Proof.* Fix  $\tau \in \mathbb{R} \setminus \text{supp } \rho$ . We denote by  $x \mapsto m_x(\tau)$  the self-adjoint solution  $m(\tau)$  viewed as a function of  $x \in [0, 1]$  with values in  $\mathbb{C}^{K \times K}$ . As is clear from the Dyson equation this function inherits the regularity of the data, i.e., it is continuous on each interval  $I_l$ . By the band mass formula (2.10) we have

$$\rho((-\infty, \tau)) = \frac{1}{K} \sum_{l=1}^L \int_{I_l} \text{Tr } \mathbf{1}_{(-\infty, 0)}(m_x(\tau)) dx = \frac{1}{K} \sum_{l=1}^L k_l |I_l|,$$

where  $k_l = \text{Tr } \mathbf{1}_{(-\infty, 0)}(m_x(\tau)) \in \{0, \dots, K\}$  is continuous in  $x \in I_l$  with discrete values and therefore does not depend on  $x$ .  $\square$

**REMARK 9.5.** We extend the conjecture from Remark 2.9 of [1] to the Kronecker setting. We expect that in the piecewise 1/2-Hölder continuous setting of the current section, the number of connected components of the self-consistent spectrum  $\text{supp } \rho$  is at most  $K(2L - 1)$ .

10 PERTURBATIONS OF THE DATA PAIR

In this section, as an application of our results in Sections 4 to 7, we show that the Dyson equation, (2.3), is stable against small general perturbations of the data pair  $(a, S)$  consisting of the bare matrix  $a$  and the self-energy operator  $S$ . To that end, let  $T \subset \mathbb{R}$  contain 0,  $S_t: \mathcal{A} \rightarrow \mathcal{A}$ ,  $t \in T$ , be a family of positivity-preserving operators and  $a_t = a_t^* \in \mathcal{A}$ ,  $t \in T$ , be a family of self-adjoint elements. We set  $S := S_{t=0}$  and  $a := a_{t=0}$  and will always assume that

there are  $c_1, \dots, c_5 > 0$  such that

$$c_1 \langle x \rangle \mathbf{1} \leq S[x] \leq c_2 \langle x \rangle \mathbf{1}, \quad \|a\| \leq c_3, \quad \|S - S_t\| \leq c_4 t, \quad \|a - a_t\| \leq c_5 t \tag{10.1}$$

for all  $x \in \overline{\mathbb{A}}_+$  and for all  $t \in T$ . For any  $t \in T$ , let  $m_t$  be the solution to the Dyson equation associated to the data pair  $(a_t, S_t)$ , i.e.,

$$-m_t(z)^{-1} = z\mathbf{1} - a_t + S_t[m_t(z)] \tag{10.2}$$

for  $z \in \mathbb{H}$  (cf. (2.3)). We also set  $m := m_{t=0}$ .

The main result of this section, Proposition 10.1 below, states that  $\|m_t(z) - m(z)\|$  is small for sufficiently small  $t$  and all  $z$  away from points, where  $m(z)$  blows up. Depending on the location of  $z$ , there are three cases for the estimate: we obtain the best estimate of order  $|t|$  on  $\|m_t(z) - m(z)\|$  in the bulk, the estimate is weaker, of order  $|t|^{1/2}$ , if  $z$  is close to a regular edge and the weakest, of order  $|t|^{1/3}$ , if  $z$  is close to an (almost) cusp point.

We now introduce these concepts precisely. For a given  $m_* > 0$ , we define the set  $P_m := P_m^{m_*} \subset \mathbb{H}$ , where  $\|m(z)\|$  is larger than  $m_*$ , i.e.,

$$P_m^{m_*} := \{\tau \in \mathbb{R} : \sup_{\eta > 0} \|m(\tau + i\eta)\| > m_*\}.$$

For any fixed  $m_* > 0$  and  $\delta > 0$ , we introduce the set  $\mathbb{D}_{\text{bdd}}$  of points of distance at least  $\delta$  from  $P_m$ , i.e.,

$$\mathbb{D}_{\text{bdd}} := \mathbb{D}_{\text{bdd}}^{m_*, \delta} := \{z \in \mathbb{H} : \text{dist}(z, P_m) \geq \delta\}. \tag{10.3}$$

Note that  $\|m(z)\| \leq \max\{m_*, \delta^{-1}\}$  for all  $z \in \mathbb{D}_{\text{bdd}}$  as  $\|m(z)\| \leq (\text{dist}(z, \text{supp } \rho))^{-1}$  by (3.7).

We now introduce the concept of the *bulk*. Since  $S \in \Sigma_{\text{flat}}$ , the self-consistent density of states of  $m$  (cf. Definition 2.2) has a continuous density  $\rho: \mathbb{R} \rightarrow [0, \infty)$  with respect to the Lebesgue measure (cf. Proposition 2.3). We also write  $\rho$  for the harmonic extension of  $\rho$  to  $\mathbb{H}$  which satisfies  $\rho(z) = \langle \text{Im } m(z) \rangle / \pi$  for  $z \in \mathbb{H}$ . For  $\rho_* > 0$  and  $\delta_s > 0$ , we denote those points, where  $\rho$  is bigger than  $\rho_*$  or which are at least  $\delta_s$  away from  $\text{supp } \rho$ , by

$$\begin{aligned} \mathbb{D}_{\text{bulk}} &:= \mathbb{D}_{\text{bulk}}^{\rho_*} := \{z \in \mathbb{H} : \rho(z) \geq \rho_*\}, \\ \mathbb{D}_{\text{out}} &:= \mathbb{D}_{\text{out}}^{\delta_s} := \{z \in \mathbb{H} : \text{dist}(z, \text{supp } \rho) \geq \delta_s\}, \end{aligned}$$

respectively. We remark that, for fixed  $\rho_*$  and  $\delta_s$ , we have the inclusion  $\mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}} \subset \mathbb{D}_{\text{bdd}}$  for all sufficiently large  $m_*$  and sufficiently small  $\delta$  by (3.12).

For  $\tau \in \mathbb{R} \setminus \text{supp } \rho$ , let  $\Delta(\tau)$  denote the size of the largest interval that contains  $\tau$  and is contained in  $\mathbb{R} \setminus \text{supp } \rho$ . For  $\rho_* > 0$  and  $\Delta_* > 0$ , we define the set  $P_{\text{cusp}} = P_{\text{cusp}}^{\rho_*, \Delta_*} \subset \mathbb{R}$  of *almost cusp points* through

$$\begin{aligned} P_{\text{cusp}}^{\rho_*, \Delta_*} &:= \{\tau \in \text{supp } \rho \setminus \partial \text{supp } \rho : \tau \text{ is a local minimum of } \rho, \rho(\tau) \leq \rho_*\} \\ &\cup \{\tau \in \mathbb{R} \setminus \text{supp } \rho : \Delta(\tau) \leq \Delta_*\}. \end{aligned} \tag{10.4}$$

For some  $\delta_c > 0$ , we denote those points which are at least  $\delta_c$  away from almost cusp points by

$$\mathbb{D}_{\text{nocusp}} := \{z \in \mathbb{H} : \text{dist}(z, P_{\text{cusp}}) \geq \delta_c\}.$$

We remark that  $\mathbb{D} = \mathbb{D}_{\text{bdd}} \cap \mathbb{D}_{\text{cusp}}$ , where  $\mathbb{D}$  denotes the set of points which are away from  $P_m$  and  $P_{\text{cusp}}$ . More precisely, for some  $\delta > 0$ , we define

$$\mathbb{D} := \{z \in \mathbb{H} : \text{dist}(z, P_m) \geq \delta, \text{dist}(z, P_{\text{cusp}}) \geq \delta\}.$$

In this section, the model parameters are given by  $c_1, \dots, c_5$  from (10.1) as well as the fixed parameters  $m_*, \delta, \rho_*, \delta_s, \Delta_*$  and  $\delta_c$  from the definitions of  $P_m, \mathbb{D}_{\text{bdd}}, \mathbb{D}_{\text{bulk}}, \mathbb{D}_{\text{out}}, P_{\text{cusp}}$ , and  $\mathbb{D}_{\text{nocusp}}$ , respectively. Thus, the comparison relation  $\sim$  (compare Convention 3.4) is understood with respect to these parameters throughout this section.

PROPOSITION 10.1. *If the self-adjoint element  $a = a_{t=0}, a_t$  in  $\mathcal{A}$  and the positivity-preserving operators  $S = S_{t=0}, S_t$  on  $\mathcal{A}$  satisfy (10.1) for each  $t \in T$  then there is  $t_* \sim 1$  such that*

- (a) *Uniformly for all  $z \in \mathbb{D}_{\text{bdd}}$  and for all  $t \in [-t_*, t_*] \cap T$ , we have*

$$\|m_t(z) - m(z)\| \lesssim |t|^{1/3}.$$

*In particular,  $\|m_t(z)\| \lesssim 1$  uniformly for all  $z \in \mathbb{D}_{\text{bdd}}$  and for all  $t \in [-t_*, t_*] \cap T$ .*

- (b) *(Bulk and away from support of  $\rho$ ) Uniformly for all  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$  and for all  $t \in [-t_*, t_*] \cap T$ , we have*

$$\|m_t(z) - m(z)\| \lesssim |t|.$$

- (c) *(Away from almost cusps) Uniformly for all  $z \in \mathbb{D}_{\text{nocusp}} \cap \mathbb{D}_{\text{bdd}}$  and for all  $t \in [-t_*, t_*] \cap T$ , we have*

$$\|m_t(z) - m(z)\| \lesssim |t|^{1/2}.$$

In order to simplify the notation, we set  $\Delta m_t = \Delta m_t(z) = m_t(z) - m(z)$ . The behaviour of  $\Delta m_t$  will be governed by a scalar-valued cubic equation (see (10.6) below). This is the origin of the cubic root  $|t|^{1/3}$  in the general estimate on  $\|m_t(z) - m(z)\|$  in Proposition 10.1. In the special cases,  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$  and  $z \in \mathbb{D}_{\text{nocusp}}$ , the cubic equation simplifies to a linear or quadratic equation, respectively, which yield the improved estimates  $\tilde{|t|}$  and  $\tilde{|t|}^{1/2}$ , respectively.

We now define two positive auxiliary functions  $\tilde{\xi}_1(z)$  and  $\tilde{\xi}_2(z)$  for  $z \in \mathbb{D}_{\text{bdd}}$  which will control the coefficients in the cubic equation mentioned above. For their definitions, we distinguish several subdomains of  $\mathbb{D}_{\text{bdd}}$ . The slight ambiguity of the definitions due to overlaps between these domains does, however, not affect the validity of the following statements as the different versions of  $\tilde{\xi}_1$  as well as  $\tilde{\xi}_2$  are comparable with each other with respect to the comparison relation  $\sim$  and  $\tilde{\xi}_1$  as well as  $\tilde{\xi}_2$  are only used in bounds with respect to this comparison relation. For  $\rho_* \sim 1$  and  $\delta_* \sim 1$ , we define

- BULK: If  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$  then we set

$$\tilde{\xi}_1(z) := \tilde{\xi}_2(z) := 1. \tag{10.5a}$$

- AROUND A REGULAR EDGE: If  $z = \tau_0 + \omega + i\eta \in \mathbb{D}_{\text{nocusp}} \cap \mathbb{D}_{\text{bdd}}$  with some  $\tau_0 \in \partial \text{supp } \rho$ ,  $\omega \in [-\delta_*, \delta_*]$  and  $\eta \in (0, \delta_*]$  then we set

$$\tilde{\xi}_1(z) := (|\omega| + \eta)^{1/2}, \quad \tilde{\xi}_2(z) := 1. \tag{10.5b}$$

- CLOSE TO AN INTERNAL EDGE WITH A SMALL GAP: Let  $\alpha, \beta \in (\partial \text{supp } \rho) \setminus P_m$  satisfy  $\beta < \alpha$  and  $(\beta, \alpha) \cap \text{supp } \rho = \emptyset$ . We set  $\Delta := \alpha - \beta$ . If  $z \in \mathbb{D}_{\text{bdd}}$  satisfies  $z = \alpha - \omega + i\eta$  or  $z = \beta + \omega + i\eta$  for some  $\omega \in [-\delta_*, \Delta/2]$  and  $\eta \in (0, \delta_*]$  then we define

$$\tilde{\xi}_1(z) := (|\omega| + \eta)^{1/2} (|\omega| + \eta + \Delta)^{1/6}, \quad \tilde{\xi}_2(z) := (|\omega| + \eta + \Delta)^{1/3} \tag{10.5c}$$

- AROUND A SMALL INTERNAL MINIMUM: If  $z = \tau_0 + \omega + i\eta \in \mathbb{D}_{\text{bdd}}$ , where  $\tau_0 \in \text{supp } \rho \setminus \partial \text{supp } \rho$  is a local minimum of  $\rho$  with  $\rho(\tau_0) \leq \rho_*$ ,  $\omega \in [-\delta_*, \delta_*]$  and  $\eta \in (0, \delta_*]$  then we define

$$\tilde{\xi}_1(z) := (\rho(\tau_0) + (|\omega| + \eta)^{1/3})^2, \quad \tilde{\xi}_2(z) := \rho(\tau_0) + (|\omega| + \eta)^{1/3}. \tag{10.5d}$$

We remark that  $\tau_0 \in \partial \text{supp } \rho$  is a *regular edge* if  $\rho(\tau) = 0$  for all  $\tau \in [\tau_0 - \varepsilon, \tau_0]$  or  $\tau \in [\tau_0, \tau_0 + \varepsilon]$  for some  $\varepsilon \sim 1$ . In fact,  $\mathbb{D}_{\text{nocusp}} \cap \mathbb{D}_{\text{bdd}} \cap \partial \text{supp } \rho$  consists only of regular edges.

In the proof of Proposition 10.1, we will use the following two lemmas, whose proofs we postpone until the end of this section.

LEMMA 10.2. *Let  $\mathbb{D}_{\text{bdd}}$  be defined as in (10.3). Let  $a, S$  and  $(a_t)_{t \in T}$  and  $(S_t)_{t \in T}$  satisfy (10.1). Then there is  $\varepsilon_1 \sim 1$  such that if  $\|\Delta m_t(z)\| \leq \varepsilon_1$  for some  $z \in \mathbb{D}_{\text{bdd}}$ ,  $t \in T$ , then there are  $l, b \in \mathcal{A}$  depending on  $z$  such that  $\Theta_t := \langle l, \Delta m_t \rangle / \langle l, b \rangle$  satisfies a cubic inequality*

$$|\Theta_t^3 + \xi_2 \Theta_t^2 + \xi_1 \Theta_t| \lesssim |t| \tag{10.6}$$

with complex coefficients  $\xi_1$  and  $\xi_2$  depending on  $z$  and  $t$ . The function  $\Theta_t$  depends continuously on  $\text{Im } z$  and we also have  $|\Theta_t| \lesssim \|\Delta m_t\|$  as well as  $\|\Delta m_t\| \lesssim |\Theta_t| + |t|$  for all  $t \in T$ .

The coefficients,  $\xi_1$  and  $\xi_2$ , behave as follows: There are  $\delta_* \sim 1$ ,  $\rho_* \sim 1$  and  $c_* \sim 1$  such that, with the appropriate definitions of  $\tilde{\xi}_1$  and  $\tilde{\xi}_2$  from (10.5), we have

- If  $z \in \mathbb{D}_{\text{bdd}}$  satisfies the conditions for (10.5a) or (10.5c) with  $\omega \in [c_* \Delta, \Delta/2]$  then we have

$$|\xi_1(z)| \sim \tilde{\xi}_1(z), \quad |\xi_2(z)| \lesssim \tilde{\xi}_2(z). \tag{10.7a}$$

- If  $z \in \mathbb{D}_{\text{bdd}}$  satisfies the conditions for (10.5c) with  $\omega \in [-\delta_*, c_*\Delta]$  or (10.5b) or (10.5d) then we have

$$|\xi_1(z)| \sim \tilde{\xi}_1(z), \quad |\xi_2(z)| \sim \tilde{\xi}_2(z). \tag{10.7b}$$

All implicit constants in this lemma are uniform for any  $t \in T$ .

LEMMA 10.3. For  $0 < \eta_* < \eta^* < \infty$ , let  $\xi_1, \xi_2: [\eta_*, \eta^*] \rightarrow \mathbb{C}$  be complex-valued functions and  $\tilde{\xi}_1, \tilde{\xi}_2, d: [\eta_*, \eta^*] \rightarrow \mathbb{R}^+$  be continuous. Suppose that some continuous function  $\Theta: [\eta_*, \eta^*] \rightarrow \mathbb{C}$  satisfies the cubic inequality

$$|\Theta^3 + \xi_2\Theta^2 + \xi_1\Theta| \lesssim d \tag{10.8}$$

on  $[\eta_*, \eta^*]$  as well as

$$|\Theta| \lesssim \min\left\{d^{1/3}, \frac{d^{1/2}}{\tilde{\xi}_2^{1/2}}, \frac{d}{\tilde{\xi}_1}\right\} \tag{10.9}$$

at  $\eta_*$ . If one of the following two sets of relations holds true:

- 1) (i)  $\tilde{\xi}_2^3/d, \tilde{\xi}_1^3/d^2, \tilde{\xi}_1^2/(d\tilde{\xi}_2)$  are monotonically increasing functions,  
 (ii)  $|\xi_1| \sim \tilde{\xi}_1, |\xi_2| \sim \tilde{\xi}_2$ ,  
 (iii)  $d^2/\tilde{\xi}_1^3 + d\tilde{\xi}_2/\tilde{\xi}_1^2$  at  $\eta^*$  is sufficiently small depending on the implicit constants in 1) (ii) as well as (10.8) and (10.9).
- 2) (i)  $\tilde{\xi}_1^3/d^2$  is a monotonically increasing function,  
 (ii)  $|\xi_1| \sim \tilde{\xi}_1, |\xi_2| \lesssim \tilde{\xi}_1^{1/2}$ .

then, on  $[\eta_*, \eta^*]$ , we have the bound

$$|\Theta| \lesssim \min\left\{d^{1/3}, \frac{d^{1/2}}{\tilde{\xi}_2^{1/2}}, \frac{d}{\tilde{\xi}_1}\right\}. \tag{10.10}$$

*Proof of Proposition 10.1.* We start the proof by introducing the control parameter  $M(t)$ . Let  $\tilde{\xi}_1$  and  $\tilde{\xi}_2$  be defined as in (10.5). For  $t \in \mathbb{R}$ , we set

$$M(t) := \min\{|t|^{1/3}, \tilde{\xi}_2^{-1/2}|t|^{1/2}, \tilde{\xi}_1^{-1}|t|\}. \tag{10.11}$$

We remark that  $M$  also depends on  $z$  as  $\tilde{\xi}_1$  and  $\tilde{\xi}_2$  depend on  $z$ .

We will prove below that there are  $t_* \sim 1$  and  $C \sim 1$  such that, for any fixed  $t \in [-t_*, t_*] \cap T \setminus \{0\}$  (if this set is nonempty) and  $z \in \mathbb{D}_{\text{bdd}}$ , we have the implication

$$\|\Delta m_t(\text{Re } z + i\eta)\| \leq \varepsilon_1 \quad \text{for all } \eta \geq \text{Im } z \quad \Rightarrow \quad \|\Delta m_t(z)\| \leq CM(t), \tag{10.12}$$

where  $\varepsilon_1 \sim 1$  is from Lemma 10.2.

Armed with (10.12), by possibly shrinking  $t_* \sim 1$ , we can assume that  $2Ct_*^{1/3} \leq \varepsilon_1$ . We fix  $\tau \in \mathbb{R}$  and  $t \in [-t_*, t_*] \cap T \setminus \{0\}$  and set

$$\eta_* := \sup\{\eta > 0 : \|\Delta m_t(\tau + i\eta)\| \geq 2CM(t)\}.$$

Here, we use the convention  $\eta_* = -\infty$  if the set is empty. Note that  $\|\Delta m_t(\tau + i\eta)\| \leq 2\eta^{-1}$  since  $m$  and  $m_t$  are Stieltjes transforms. Hence,  $\eta_* < \infty$  as  $t \neq 0$ . We prove now that  $\eta_* \leq \inf\{\text{Im } z : z \in \mathbb{D}_{\text{bdd}}, \text{Re } z = \tau\}$ . For a proof by contradiction, we suppose that there is  $z_* \in \mathbb{D}_{\text{bdd}}$  such that  $\text{Re } z_* = \tau$  and  $\text{Im } z_* = \eta_*$  (note that if  $\tau + i\eta \in \mathbb{D}_{\text{bdd}}$  then  $\tau + i\eta' \in \mathbb{D}_{\text{bdd}}$  for any  $\eta' \geq \eta$ ). Since  $\Delta m_t$  is continuous in  $z$ , we have  $\|\Delta m_t(z_*)\| = 2CM(t)$ . Thus,  $\|\Delta m_t(\tau + i\eta)\| \leq 2Ct_*^{1/3} \leq \varepsilon_1$  for all  $\eta \geq \eta_*$  by the choice of  $t_*$ . From (10.12), we conclude  $\|\Delta m_t(z_*)\| \leq CM(t)$ , which contradicts  $\|\Delta m_t(z_*)\| = 2CM(t)$ . Thus,  $\eta_* \leq \inf\{\text{Im } z : z \in \mathbb{D}_{\text{bdd}}, \text{Re } z = \tau\}$ .

As  $\tau$  was arbitrary, this yields  $\|\Delta m_t(z)\| \leq 2CM(t)$  for all  $z \in \mathbb{D}_{\text{bdd}}$ , which proves part (a) of Proposition 10.1 up to (10.12). Since  $\tilde{\xi}_1(z) \sim 1$  for  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$  and  $\tilde{\xi}_2(z) \sim 1$  for  $z \in \mathbb{D}_{\text{nocusp}} \cap \mathbb{D}_{\text{bdd}}$ , we also obtain part (b) and (c) from the definition of  $M$  in (10.11).

Hence, it suffices to show (10.12) to complete the proof of Proposition 10.1. In order to prove (10.12), we use Lemma 10.3 with  $\Theta(\eta) = \Theta_t(\text{Re } z + i\eta)$ ,  $\eta \geq \eta_* := \text{Im } z$ ,  $d = |t|$ , and  $\xi_1, \xi_2$  and  $\tilde{\xi}_1, \tilde{\xi}_2$  are chosen as in (10.6) of Lemma 10.2 and (10.5), respectively. As  $\|\Delta m_t(\text{Re } z + i\eta)\| \leq \varepsilon_1$  for all  $\eta \geq \text{Im } z$ , we conclude that (10.8) is satisfied with  $d = |t|$  due to (10.6).

We first consider  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$ . If  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$  then  $\text{Re } z + i\eta \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$  and  $\xi_1(\text{Re } z + i\eta) = \xi_2(\text{Re } z + i\eta) = 1$  for all  $\eta \geq \eta_*$  and assumption 2) of Lemma 10.3 is always fulfilled. Since  $\|\Delta m_t(\text{Re } z + i\eta)\| \leq 2\eta^{-1}$  as remarked above and  $t \neq 0$ , the condition in (10.9) is met for some sufficiently large  $\eta > 0$ . Hence, by Lemma 10.3, there is  $C \sim 1$  such that  $|\Theta_t(z)| \leq CM(t)$ . Possibly increasing  $C \sim 1$  and using  $|t| \leq t_* \sim 1$  yield  $\|\Delta m_t(z)\| \leq CM(t)$  due to  $\|\Delta m_t\| \lesssim |\Theta_t| + |t|$  from Lemma 10.2.

For each  $z \in \mathbb{D}_{\text{bdd}} \setminus \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$ , due to (10.7), we have  $\xi_1(z_\delta) \sim 1$  and  $\xi_2(z_\delta) \sim 1$  for  $z_\delta := \text{Re } z + i\delta_*$ , where  $\delta_* \sim 1$  is as in Lemma 10.2. Hence, we conclude  $|\Theta_t(z_\delta)| \leq CM(t)$  as for  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$ . For each  $z \in \mathbb{D}_{\text{bdd}} \setminus \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$ , the validity of assumption 1) or assumption 2) of Lemma 10.3 can be read off from (10.7). Lemma 10.3, thus, implies  $|\Theta_t(z)| \leq CM(t)$ . As before, we conclude  $\|\Delta m_t(z)\| \leq CM(t)$  from Lemma 10.2. This completes the proof of (10.12) and, hence, the one of Proposition 10.1.  $\square$

*Proof of Lemma 10.2.* We remark that a straightforward computation starting from (2.3) and (10.2) yields

$$B[\Delta m_t] = A[\Delta m_t, \Delta m_t] + K[\Delta^S, \Delta^a, \Delta m_t] + T[\Delta^S, \Delta^a], \tag{10.13}$$

where  $B := \text{Id} - C_m S$ ,  $A[x, y] := (mS[x]y + yS[x]m)/2$  are defined as in (6.23),

$\Delta^S := S_t - S$ ,  $\Delta^a := a_t - a$  and

$$\begin{aligned}
 K[\Delta^S, \Delta^a, \Delta m_t] &= \frac{1}{2}(m\Delta^S[\Delta m_t]\Delta m_t + \Delta m_t\Delta^S[\Delta m_t]m + m\Delta^S[m]\Delta m_t \\
 &\quad + \Delta m_t\Delta^S[m]m) - \frac{1}{2}(m\Delta^a\Delta m_t + \Delta m_t\Delta^a m), \\
 T[\Delta^S, \Delta^a] &= m\Delta^S[m]m - m\Delta^a m.
 \end{aligned}$$

In the following, we will split  $\mathbb{D}_{\text{bdd}}$  into two regimes and choose  $l$  and  $b$  according to the regime. In both cases, we use the definitions

$$\Theta := \Theta_t = \frac{\langle l, \Delta m_t \rangle}{\langle l, b \rangle}, \quad r = r_t := Q[\Delta m_t], \quad Q := \text{Id} - \frac{\langle l, \cdot \rangle}{\langle l, b \rangle} b. \tag{10.14}$$

In particular,  $\Delta m_t = \Theta b + r$ . We denote by  $\rho(z)$  the harmonic extension of  $\rho$ , i.e.,  $\rho(z) = \langle \text{Im } m(z) \rangle / \pi$ .

If  $z$  is close to a regular edge or close to an almost cusp point then  $\Delta m_t(z)$  is governed by a quadratic or cubic equation for  $\Theta_t$ , respectively, where  $l$  and  $b$  are a left and a right eigenvector of  $B$ , respectively. If  $z$  is in the bulk or away from  $\text{supp } \rho$  then  $\Delta m_t(z)$  can be controlled by  $\Theta_t$  with  $l = b = \mathbf{1}$  and  $\Theta_t$  is the solution of a scalar-valued linear equation. Note that in the bulk and away from  $\text{supp } \rho$  the choice  $l = b = \mathbf{1}$  is arbitrary, in fact the splitting  $\Delta m_t = \Theta_t b + r$  is artificial since the stability operator does not have a distinguished “bad” direction that needs to be treated separately. We still use this formalism in order to treat all three cases uniformly for the sake of brevity. For a similar reason we will always write the equation for  $\Theta_t$  as a cubic equation, sometimes by adding and subtracting apparently superfluous (and negligible) terms.

*Case 1:* We first assume that  $z \in \mathbb{D}_{\text{bdd}}$  satisfies  $\rho(z) \geq \rho_*$  for some  $\rho_* \sim 1$  or  $\text{dist}(z, \text{supp } \rho) \geq \delta$  for some  $\delta \sim 1$ , i.e.,  $z \in \mathbb{D}_{\text{bulk}}^{\rho_*} \cup \mathbb{D}_{\text{out}}^\delta$ . This implies that  $B$  is invertible and  $\|B^{-1}\| \lesssim 1$  due to (4.1),  $\|S\|_{2 \rightarrow \|\cdot\|} \lesssim 1$ ,  $\|m(z)\| \lesssim 1$  and Lemma B.2 (ii). In this case, we choose  $l = b = \mathbf{1}$  and apply  $QB^{-1}$  to (10.13) to obtain

$$\begin{aligned}
 r &= QB^{-1}(A[\Delta m_t, \Delta m_t] + K[\Delta^S, \Delta^a, \Delta m_t] + T[\Delta^S, \Delta^a]) \\
 &= \mathcal{O}(|\Theta|^2 + \|r\| \|\Delta m_t\| + |t|),
 \end{aligned}$$

where we used that  $\|m\| \lesssim 1$  on  $\mathbb{D}_{\text{bdd}}$  as well as  $\|\Delta^S\| + \|\Delta^a\| \lesssim |t|$ . Shrinking  $\varepsilon_1 \sim 1$ , using  $\|\Delta m_t\| \leq \varepsilon_1$  and absorbing  $\|r\| \|\Delta m_t\|$  into the left-hand side yield  $\|r\| \lesssim |\Theta|^2 + |t|$ . Thus,  $\|\Delta m_t\| \lesssim |\Theta| + |t|$ . Hence, applying  $B^{-1}$  and  $\langle \cdot \rangle$  to (10.13) and using  $\langle r \rangle = 0$  as well as  $\|\Delta m_t\| \lesssim |\Theta| + |t|$ , we find  $\xi_2 \in \mathbb{C}$  such that  $|\xi_2| \lesssim 1 = \tilde{\xi}_2$  and

$$\Theta = -\xi_2 \Theta^2 + \mathcal{O}(|t||\Theta| + |t|) = -\xi_2 \Theta^2 + \mathcal{O}(|t|).$$

Adding and subtracting  $\Theta^3$  on the left-hand side as well as setting  $\xi_1 := 1 - \Theta^2$  show (10.6) in Case 1 for sufficiently small  $\varepsilon_1 \sim 1$  as  $|\Theta| \lesssim \|\Delta m_t\| \leq \varepsilon_1$  implies  $|\xi_1| \sim 1 = \tilde{\xi}_1$ . This completes the proof of (10.7a) for  $z \in \mathbb{D}_{\text{bulk}} \cup \mathbb{D}_{\text{out}}$ .



*Case 2:* We now prove (10.6) for  $z \in \mathbb{D}_{\text{bdd}}$  satisfying  $\rho(z) \leq \rho_*$  and  $\text{dist}(z, \text{supp } \rho) \leq \delta$  with sufficiently small  $\rho_* \sim 1$  and  $\delta \sim 1$ . For any  $\varepsilon_* \sim 1$ , we find  $\delta \sim 1$  such that  $\rho(z)^{-1} \text{Im } z \leq \varepsilon_*$  for all  $z \in \mathbb{H}$  satisfying  $\text{dist}(z, \text{supp } \rho) \leq \delta$  due to (5.26) and the 1/3-Hölder continuity of  $z \mapsto \rho(z)^{-1} \text{Im } z$  by Lemma 5.4 (ii). Therefore, using  $\rho(z) \leq \rho_*$ , we see that Lemma 5.1 and Corollary 5.2 are applicable for sufficiently small  $\rho_* \sim 1$  and  $\delta \sim 1$ . They yield  $l, b \in \mathcal{A}$  which we use to define  $\Theta$  and  $r$  as in (10.14), i.e.,  $\Delta m_t = \Theta b + r$  and  $\Theta = \langle l, \Delta m_t \rangle / \langle l, b \rangle$ . In order to derive (10.6), we now follow the proof of Lemma 6.2 applied to (10.13) instead of (6.10). Here,  $\Delta^a$  and  $\Delta^S$  play the role of  $e$ . In fact, by Lemma 5.1 and Corollary 5.2, the first two bounds in (6.12) are fulfilled. Owing to  $\|m\| \lesssim 1$ , the third bound in (6.12) is trivially satisfied. Instead of the last two bounds in (6.12), we use

$$\|T[\Delta^S, \Delta^a]\| \lesssim \|\Delta^S\| + \|\Delta^a\|, \quad \|K[\Delta^S, \Delta^a, \Delta m_t]\| \lesssim (\|\Delta^S\| + \|\Delta^a\|) \|\Delta m_t\|,$$

due to  $\|m\| \lesssim 1$  and  $\|\Delta m_t\| \lesssim 1$ . In fact, the last bound in (6.12) will not hold true for a general  $y \in \mathcal{A}$  but in the proof of Lemma 6.2 it is only used with the special choice  $y = \Delta m_t$ . We choose  $\varepsilon_1 \leq \varepsilon$  for  $\varepsilon$  from Lemma 6.2 and obtain the cubic equation (6.14) from Lemma 6.2 with  $\mu_0 = \langle l, T[\Delta^S, \Delta^a] \rangle$  and  $\|e\|$  replaced by  $|t|$  as  $\|\Delta^S\| + \|\Delta^a\| \lesssim |t|$ . In particular,  $|\mu_0| \lesssim |t|$ . We decompose the error term  $\tilde{e} = \mathcal{O}(|\Theta|^4 + |t||\Theta| + |t|^2)$  from (6.14) into  $\tilde{e} = \tilde{e}_1 \Theta^3 + \tilde{e}_2$  with  $\tilde{e}_1, \tilde{e}_2 \in \mathbb{C}$  satisfying  $\tilde{e}_1 = \mathcal{O}(|\Theta|)$  and  $\tilde{e}_2 = \mathcal{O}(|t||\Theta| + |t|^2)$ . With the notation of Lemma 6.2, the cubic equation (6.14) can be written as

$$(\mu_3 - \tilde{e}_1)\Theta^3 + \mu_2\Theta^2 + \mu_1\Theta = -\mu_0 + \tilde{e}_2 = \mathcal{O}(|t|).$$

Since  $A$  and  $B$  introduced above have the same definitions as in (6.23) and  $\mu_3, \mu_2$  and  $\mu_1$  in (6.15) depend only on  $A$  and  $B$ , Lemma 6.3 yields the expansions of  $\mu_3, \mu_2$  and  $\mu_1$  in (6.24) for sufficiently small  $\rho_* \sim 1$  and  $\delta \sim 1$ . By possibly shrinking  $\varepsilon_1 \sim 1$ , we find  $c \sim 1$  such that  $|\mu_3 - \tilde{e}_1| + |\mu_2| \geq 2c$  as  $|\tilde{e}_1| \lesssim |\Theta| \lesssim \|\Delta m_t\| \leq \varepsilon_1$ . Here, we also used  $|\mu_3| + |\mu_2| \gtrsim \psi + |\sigma|$  by (6.24) as well as (5.35). Consequently, we obtain (10.6), where we introduced

$$\begin{aligned} \xi_2 &:= \left( \mu_2 + (\mu_3 - \tilde{e}_1 - 1)\Theta \right) \mathbf{1}(|\mu_2| \geq c) + \frac{\mu_2}{\mu_3 - \tilde{e}_1} \mathbf{1}(|\mu_2| < c), \\ \xi_1 &:= \mu_1 \mathbf{1}(|\mu_2| \geq c) + \frac{\mu_1}{\mu_3 - \tilde{e}_1} \mathbf{1}(|\mu_2| < c). \end{aligned}$$

Hence, we have  $|\xi_2| \sim |\mu_2|$  and  $|\xi_1| \sim |\mu_1|$  for sufficiently small  $\varepsilon_1 \sim 1$  as  $|\tilde{e}_1| \lesssim |\Theta|$  and  $|\Theta| \lesssim \|\Delta m_t\| \leq \varepsilon_1$ . This completes the proof of (10.6) in Case 2.

It remains to show the scaling relations in (10.7) for  $z \in \mathbb{D}_{\text{bdd}}$  satisfying  $\rho(z) \leq \rho_*$  and  $\text{dist}(z, \text{supp } \rho) \leq \delta$  in order to complete the proof of Lemma 10.2. Starting from  $|\xi_1| \sim |\mu_1|$  and  $|\xi_2| \sim |\mu_2|$  proven in Case 2, we conclude as in the proof of (10.6) in [3] that

$$|\xi_1| \sim \rho(z)^2 + |\sigma(z)|\rho(z) + \rho(z)^{-1} \text{Im } z, \quad |\xi_2| \sim \rho(z) + |\sigma(z)|,$$

where  $\sigma$  is defined as in (5.12). Here,  $\xi_1$  and  $\xi_2$  play the role of  $\pi_1$  and  $\pi_2$ , respectively, in [3]. Their definitions differ slightly but this does not affect the straightforward estimates. Note that the proof in [3] relies on the expansions of  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  from (8.33) in [3]. These are the exact analogues of (6.24), where  $\rho$  plays the role of  $\alpha$  from [3].

Note that according to Remark 7.3 the harmonic extension  $\rho(z)$  for  $z \in \mathbb{H}$  in the vicinity of the singularities has the same scaling behavior as in Corollary A.1 of [3]. Similarly, the proof of (10.7) in [3] yields

$$|\sigma(\beta)| \sim |\sigma(\alpha)| \sim (\alpha - \beta)^{1/3}, \quad |\sigma(\tau_0)| \lesssim \rho(\tau_0)^2, \quad (10.15)$$

where  $\alpha, \beta \in (\partial \text{supp } \rho) \setminus P_m$  satisfy  $\beta < \alpha$  and  $(\beta, \alpha) \cap \text{supp } \rho = \emptyset$  and  $\tau_0 \in \text{supp } \rho \setminus \partial \text{supp } \rho$  is a local minimum of  $\rho$  and  $\rho(\tau_0) \leq \rho_*$ . Here, we use Lemma 7.16 above and  $|\sigma| \sim \widehat{\Delta}^{1/3}$  by Theorem 7.7 (ii) (b) instead of Lemma 9.17 in [3] and Lemma 7.14 above instead of Lemma 9.2 in [3]. We then follow the proof of Proposition 4.3 in [2] and use the 1/3-Hölder continuity of  $\sigma$  proven in Lemma 5.5 (i). This yields the missing scaling relations in (10.7). We remark that  $\Theta_t$  constructed above is not continuous in  $\text{Im } z$  due to the separation into two cases. However, there is only one transition between Case 1 and Case 2 for  $z \in \mathbb{D}_{\text{bdd}}$  when  $\text{Im } z$  is varied while  $\text{Re } z$  is kept fixed. Therefore, we obtain a continuous version of  $\Theta_t$  by a simple interpolation between these two cases in the vicinity of this transition point. We leave the details of this interpolation argument to the reader. This completes the proof of Lemma 10.2.  $\square$

REMARK 10.4 (Scaling of coefficients). The proof of Lemma 10.2 can equally well be carried out under Assumption 4.5 instead of the flatness condition in (10.1). In particular, it shows that in the setting of Theorem 7.2, there are  $\delta_* \sim 1, \rho_* \sim 1$  and  $c_* \sim 1$  such that the following comparison relations hold for  $z \in I_\theta + i[0, \eta_*]$ :

- If  $z$  satisfies the conditions for (10.5a) or (10.5c) with  $\omega \in [c_*\Delta, \Delta/2]$ , then we have

$$\rho(z)^2 + |\sigma(z)|\rho(z) + \rho(z)^{-1}\text{Im } z \sim \widetilde{\xi}_1(z), \quad \rho(z) + |\sigma(z)| \lesssim \widetilde{\xi}_2(z).$$

- If  $z$  satisfies the conditions for (10.5c) with  $\omega \in [-\delta_*, c_*\Delta]$  or (10.5b) or (10.5d) with  $\rho(\tau_0) \leq \rho_*$ , then we have

$$\rho(z)^2 + |\sigma(z)|\rho(z) + \rho(z)^{-1}\text{Im } z \sim \widetilde{\xi}_1(z), \quad \rho(z) + |\sigma(z)| \sim \widetilde{\xi}_2(z).$$

*Proof of Lemma 10.3.* By dividing the cubic inequality through  $d$  and considering  $\frac{\Theta}{d^{1/3}}$  instead of  $\Theta$ , we may assume that  $d = 1$ . We fix  $\varepsilon \in (0, 1)$  sufficiently small. First we prove the lemma under assumption 1). Owing to the smallness of  $\frac{1}{\xi_1^3} + \frac{\xi_2}{\xi_1^2}$  at  $\eta^*$  as well as the monotonicity of  $\widetilde{\xi}_1$  and  $\frac{\widetilde{\xi}_2}{\xi_2}$  there are  $0 < \eta_1, \eta_2 < \eta^*$

with the following properties: (i)  $\tilde{\xi}_2 \geq \varepsilon^4 \tilde{\xi}_1^2$  on  $[\eta_*, \eta_1]$ ; (ii)  $\tilde{\xi}_2 \leq \varepsilon^4 \tilde{\xi}_1^2$  on  $[\eta_1, \eta^*]$ ; (iii)  $\varepsilon \tilde{\xi}_1 \leq 1$  on  $[\eta_*, \eta_2]$ ; (iv)  $\varepsilon \tilde{\xi}_1 \geq 1$  on  $[\eta_2, \eta^*]$ . Here the intervals  $[\eta_*, \eta_2]$  and  $[\eta_*, \eta_1]$  may be empty. We will now assume the bound  $|\Theta| \lesssim \min\{1, \frac{1}{\tilde{\xi}_2^{1/2}}, \frac{1}{\tilde{\xi}_1}\}$  at the initial value  $\eta^*$  and bootstrap it down to  $\eta_*$ . Now we distinguish two cases:

*Case 1* ( $\eta_1 \geq \eta_2$ ): On  $[\eta_1, \eta^*]$  we have  $\varepsilon \tilde{\xi}_1 \geq 1$  and  $\tilde{\xi}_2 \leq \varepsilon^4 \tilde{\xi}_1^2$ . Thus, by the cubic inequality

$$|\Theta| \lesssim \min\left\{1, \frac{1}{\tilde{\xi}_1^{1/2}}\right\} \quad \text{implies} \quad |\Theta| \lesssim \frac{1}{\tilde{\xi}_1} \lesssim \min\left\{\varepsilon, \frac{\varepsilon^2}{\tilde{\xi}_2^{1/2}}\right\}.$$

In particular, there is a gap in the values of  $|\Theta|$  and by continuity all values lie below the gap on  $[\eta_1, \eta^*]$ .

The interval  $[\eta_*, \eta_1]$  is split again,  $[\eta_*, \eta_1] = [\eta_*, \eta_3] \cup [\eta_3, \eta_1]$ , where  $\eta_3$  is chosen such that (i)  $\tilde{\xi}_2 \varepsilon^2 \geq 1$  on  $[\eta_3, \eta_1]$ ; (ii)  $\tilde{\xi}_2 \varepsilon^2 \leq 1$  on  $[\eta_*, \eta_3]$ . Here one or both of these intervals may be empty. Using  $\tilde{\xi}_2 \geq \varepsilon^4 \tilde{\xi}_1^2$  we see that on  $[\eta_3, \eta_1]$  the bound

$$|\Theta| \lesssim \min\left\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^3 \tilde{\xi}_1}\right\} \quad \text{implies} \quad |\Theta| \lesssim \frac{1}{\varepsilon^{3/2} \tilde{\xi}_2^{1/2}} \lesssim \min\left\{\frac{1}{\varepsilon^{1/2}}, \frac{1}{\varepsilon^{7/2} \tilde{\xi}_1}\right\}.$$

Again the gap in the values of  $|\Theta|$  allows us to infer from the bound  $|\Theta| \lesssim \min\{1, \frac{1}{\tilde{\xi}_2^{1/2}}, \frac{1}{\tilde{\xi}_1}\}$  at  $\eta_1$  that  $|\Theta|$  satisfies the same bound on  $[\eta_3, \eta_1]$  up to an  $\varepsilon$ -dependent multiplicative constant.

Finally, on  $[\eta_*, \eta_3]$  we have  $\tilde{\xi}_2 \leq \varepsilon^{-2}$  and  $\tilde{\xi}_1^2 \leq \varepsilon^{-4} \tilde{\xi}_2 \leq \varepsilon^{-6}$ . Using the cubic inequality this immediately implies  $|\Theta| \lesssim_\varepsilon 1 \lesssim_\varepsilon \min\{1, \frac{1}{\tilde{\xi}_2^{1/2}}, \frac{1}{\tilde{\xi}_1}\}$ . Here and in the following, the notation  $\lesssim_\varepsilon$  indicates that the implicit constant in the bound is allowed to depend on  $\varepsilon$ .

*Case 2* ( $\eta_1 \leq \eta_2$ ): On  $[\eta_2, \eta^*]$  we have  $\varepsilon \tilde{\xi}_1 \geq 1$  and  $\tilde{\xi}_2 \leq \varepsilon^4 \tilde{\xi}_1^2$ . So this regime is treated exactly as in the beginning of *Case 1*. On  $[\eta_*, \eta_2]$  we have  $\varepsilon \tilde{\xi}_1 \leq 1$  and  $\tilde{\xi}_2 \leq \tilde{\xi}_2(\eta_2) \leq \varepsilon^4 \tilde{\xi}_1(\eta_2)^2 = \varepsilon^2$ , which implies  $|\Theta| \lesssim_\varepsilon 1 \lesssim_\varepsilon \min\{1, \frac{1}{\tilde{\xi}_2^{1/2}}, \frac{1}{\tilde{\xi}_1}\}$ .

Now we prove the lemma under assumption 2). In this case we choose  $0 < \eta_1 < \eta^*$  such that (i)  $\varepsilon \tilde{\xi}_1 \geq 1$  on  $[\eta_1, \eta^*]$ ; (ii)  $\varepsilon \tilde{\xi}_1 \leq 1$  on  $[\eta_*, \eta_1]$ . Here the interval  $[\eta_*, \eta_1]$  may be empty.

On  $[\eta_1, \eta^*]$  the bound

$$\begin{aligned} |\Theta| \lesssim 1 \quad \text{implies} \quad \tilde{\xi}_1 |\Theta| &\lesssim 1 + \tilde{\xi}_1^{1/2} |\Theta|^2 \lesssim \varepsilon^{-1/2} + \varepsilon^{1/2} \tilde{\xi}_1 |\Theta| \\ &\text{implies} \quad |\Theta| \lesssim \frac{1}{\sqrt{\varepsilon} \tilde{\xi}_1} \leq \sqrt{\varepsilon}. \end{aligned}$$

From the gap in the values of  $|\Theta|$  and its continuity we infer  $|\Theta| \lesssim \min\{\sqrt{\varepsilon}, \frac{1}{\sqrt{\varepsilon} \tilde{\xi}_1}\}$ . On  $[\eta_*, \eta_1]$  we use  $\tilde{\xi}_1 \leq \varepsilon^{-1}$  and  $|\xi_2| \lesssim \tilde{\xi}_1^{1/2} \leq \varepsilon^{-1/2}$  to conclude  $|\Theta| \lesssim_\varepsilon 1 \lesssim_\varepsilon \min\{1, \frac{1}{\tilde{\xi}_1}\}$ . This finishes the proof of the lemma. □

LEMMA 10.5 (Hölder continuity of  $\sigma$  and  $\psi$  with respect to  $a$  and  $S$ ). *Let  $T \subset \mathbb{R}$  contain 0. For each  $t \in T$ , we assume that the linear operator  $S_t: \mathcal{A} \rightarrow \mathcal{A}$  satisfies*

$$c_1 \langle x \rangle \mathbf{1} \leq S_t[x] \leq c_2 \langle x \rangle \mathbf{1} \tag{10.16}$$

for all  $x \in \overline{\mathcal{A}}_+$  and some  $c_2 > c_1 > 0$ . Moreover, let  $a_t = a_t^* \in \mathcal{A}$  be self-adjoint such that  $S_t$  and  $a_t$  satisfy (10.1) with  $a := a_{t=0}$  and  $S := S_{t=0}$ . Let  $m_t$  be the solution to (10.2) and  $\rho(z) := \langle \text{Im } m_0(z) \rangle / \pi$  for  $z \in \mathbb{H}$ .

If  $\sigma_t$  and  $\psi_t$  are defined according to (5.12), where  $m$  is replaced by  $m_t$ , then there are  $\rho_* \sim 1$  and  $t_* \sim 1$  such that

$$|\sigma_t(z_1) - \sigma_0(z_1)| \lesssim |t|^{1/3}, \quad |\psi_t(z_2) - \psi_0(z_2)| \lesssim |t|^{1/3}$$

for all  $t \in [-t_*, t_*] \cap T$  and all  $z_1, z_2 \in \mathbb{D}_{\text{bdd}} \cap \{z \in \mathbb{H} : |z| \leq c_6\}$  satisfying  $\rho(z_1) \leq \rho_*$  and  $\rho(z_2) + \rho(z_2)^{-1} \text{Im } z_2 \leq \rho_*$ . Here,  $c_6 > 0$  is also considered a model parameter.

*Proof.* We choose  $t_*$  as in Proposition 10.1 and conclude from this result that  $\|m_t(z)\| \leq k_3$  for all  $t \in [-t_*, t_*] \cap T$ , all  $z \in \mathbb{D}_{\text{bdd}}$  and some  $k_3 \sim 1$ . Hence, owing to (10.1), (10.16) and Lemma 4.8 (ii), the conditions of Assumptions 4.5 are met on  $\mathbb{D}_{\text{bdd}} \cap \{z \in \mathbb{H} : |z| \leq c_6\}$ . Hence, from the proof of Lemma 5.4, it can be read off that, after reducing  $\rho_* \sim 1$  and  $t_* \sim 1$  if necessary,  $\mathcal{M}^{(2)} := \{m_t(z_1) : t \in [-t_*, t_*] \cap T\}$  and  $\mathcal{M}^{(3)} := \{m_t(z_2) : t \in [-t_*, t_*] \cap T\}$  satisfy the conditions of Remark 5.6 (ii) and (iii), respectively, uniformly for any  $z_1, z_2 \in \mathbb{D}_{\text{bdd}} \cap \{z \in \mathbb{H} : |z| \leq c_6\}$  such that  $\rho(z_1) \leq \rho_*$  and  $\rho(z_2) + \rho(z_2)^{-1} \text{Im } z_2 \leq \rho_*$ . Therefore, the lemma is a consequence of Remark 5.6 (ii) and (iii) as well as Proposition 10.1 (a).  $\square$

REMARK 10.6. Combining Lemma 5.5 and Lemma 10.5, we obtain that  $m$ ,  $\sigma$  and  $\psi$  are jointly Hölder continuous in all three variables  $(z, a, S)$  in the following sense. Suppose that  $m$  solves the MDE for some data pair  $(a, S)$  satisfying Assumptions 4.5 on some  $I$  for some  $\eta_* \in (0, 1]$  and consider a one-parameter family of data pairs  $(a_t, S_t)$ ,  $t \in T$ , as described in Lemma 10.5. Then  $m = m_t(z)$ , as well as of  $\sigma_t(z_1)$  and  $\psi_t(z_2)$  are uniformly 1/3-Hölder continuous functions of  $t \in [-t_*, t_*] \cap T$  as well as  $z \in \mathbb{H}_{I_\theta, \eta_*}$ ,  $z_1 \in \{\zeta \in \mathbb{H}_{I_\theta, \eta_*} : \rho(\zeta) \leq \rho_*\}$  and  $z_2 \in \{\zeta \in \mathbb{H}_{I_\theta, \eta_*} : \rho(\zeta) + \rho(\zeta)^{-1} \text{Im } \zeta \leq \rho_*\}$ , respectively, for sufficiently small  $t_* \sim 1$  and  $\rho_* \sim 1$ .

REMARK 10.7 (Scaling of  $\sigma$ ). Let Assumptions 4.5 hold true for some interval  $I$  and  $\eta_* \in (0, 1]$ . Let  $\theta \in (0, 1]$ .

- (i) As in the proof of (10.15) in the proof of Lemma 10.2, we obtain that

$$|\sigma(\tau_0)| \sim |\sigma(\tau_1)| \sim (\tau_1 - \tau_0)^{1/3},$$

if  $\tau_0, \tau_1 \in \text{supp } \rho \cap I_\theta$  satisfy  $\tau_0 < \tau_1$  and  $(\tau_0, \tau_1) \cap \text{supp } \rho = \emptyset$ . Furthermore, there is  $\rho_* \sim 1$  such that

$$|\sigma(\tau_0)| \lesssim \rho(\tau_0)^2,$$

if  $\tau_0 \in \text{supp } \rho \cap I_\theta$  is a local minimum of  $\rho$  satisfying  $\rho(\tau_0) \leq \rho_*$ .

- (ii) Owing to the 1/3-Hölder continuity of  $\sigma$  from Lemma 5.5 (i), we conclude that there is  $\varepsilon \sim 1$  such that  $|\sigma(\tau)| \sim (\tau_1 - \tau_0)^{1/3}$  for all  $\tau \in \text{supp } \rho \cap I_\theta$  satisfying  $\min\{|\tau - \tau_0|, |\tau - \tau_1|\} \leq \varepsilon(\tau_1 - \tau_0)$  for some  $\tau_0, \tau_1 \in \text{supp } \rho \cap I_\theta$  such that  $\tau_0 < \tau_1$  and  $(\tau_0, \tau_1) \cap \text{supp } \rho = \emptyset$ .
- (iii) If  $\tau \in I_\theta$  satisfies the assumptions of (ii) as well as  $\rho(\tau) > 0$  then we write  $\Delta := \tau_1 - \tau_0$  and conclude from (ii) and Lemma 5.7 that

$$\|\partial_\tau m(\tau)\| \lesssim \frac{1}{\rho(\tau)(\rho(\tau) + \Delta^{1/3})}.$$

A STIELTJES TRANSFORMS OF POSITIVE OPERATOR-VALUED MEASURES

In this appendix, we will show some results about the Stieltjes transform of a positive operator-valued measure on  $\mathcal{A}$ .

We first prove Lemma 3.1 by generalizing existing proofs in the matrix algebra setup. Since we have not found the general version in the literature, we provide a proof here for the convenience of the reader. In the proof of Lemma 3.1, we will use that a von Neumann algebra is always isometrically isomorphic as a Banach space to the dual space of a Banach space. In our setup, this Banach space and the identification are simple to introduce which we will explain now. Analogously to  $L^2$  defined in Section 4, we define  $L^1$  to be the completion of  $\mathcal{A}$  when equipped with the norm  $\|x\|_1 := \langle (x^*x)^{1/2} \rangle = \langle |x| \rangle$  for  $x \in \mathcal{A}$ . Moreover, we extend  $\langle \cdot \rangle$  to  $L^1$  and remark that  $xy \in L^1$  for  $x \in \mathcal{A}$  and  $y \in L^1$ . It is well-known (e.g. [42, Theorem 2.18]) that the dual space  $(L^1)'$  of  $L^1$  can be identified with  $\mathcal{A}$  via the isometric isomorphism

$$\mathcal{A} \rightarrow (L^1)', \quad x \mapsto \psi_x, \quad \psi_x : L^1 \rightarrow \mathbb{C}, \quad y \mapsto \langle xy \rangle. \tag{A.1}$$

We stress that the existence of this isomorphism requires the state  $\langle \cdot \rangle$  to be normal.

*Proof of Lemma 3.1.* From (3.5), we conclude that

$$\lim_{\eta \rightarrow \infty} i\eta \langle x, h(i\eta)x \rangle = -\langle x, x \rangle$$

for all  $x \in \mathcal{A}$ . Hence,  $z \mapsto \langle x, h(z)x \rangle$  is the Stieltjes transform of a unique finite positive measure  $v_x$  on  $\mathbb{R}$  with  $v_x(\mathbb{R}) = \|x^*x\|_1$  (see e.g. Theorem 3.5 in [23]). For any  $x \in \mathcal{A}$ , we can find  $x_1, \dots, x_4 \in \overline{\mathcal{A}}_+$  such that  $x = x_1 - x_2 + ix_3 - ix_4$ . We define

$$\varphi_B(x) := v_{\sqrt{x_1}}(B) - v_{\sqrt{x_2}}(B) + iv_{\sqrt{x_3}}(B) - iv_{\sqrt{x_4}}(B) \tag{A.2}$$

for  $B \in \mathcal{B}$ . This definition is independent of the representation of  $x$ . Indeed, for fixed  $x \in \mathcal{A}$ , any representation  $x = x_1 - x_2 + ix_3 - ix_4$  with  $x_1, \dots, x_4 \in \overline{\mathcal{A}}_+$  defines a complex measure  $\varphi \cdot (x)$  through  $B \mapsto \varphi_B(x)$  on  $\mathbb{R}$  via (A.2). However,

extending  $h$  to the lower half-plane by setting  $h(z) := h(\bar{z})^*$  for  $z \in \mathbb{C}$  with  $\text{Im } z < 0$ , the Stieltjes transform of  $\varphi \cdot (x)$  is given by

$$\int_{\mathbb{R}} \frac{\varphi d\tau(x)}{\tau - z} = \langle \sqrt{x_1}, h(z)\sqrt{x_1} \rangle - \langle \sqrt{x_2}, h(z)\sqrt{x_2} \rangle + i\langle \sqrt{x_3}, h(z)\sqrt{x_3} \rangle - i\langle \sqrt{x_4}, h(z)\sqrt{x_4} \rangle = \langle h(z)x \rangle$$

for all  $z \in \mathbb{C} \setminus \mathbb{R}$ . This formula shows that the Stieltjes transform of  $\varphi \cdot (x)$  is independent of the decomposition  $x = x_1 - x_2 + ix_3 - ix_4$ . Hence,  $\varphi_B(x)$  is independent of this representation for all  $B \in \mathcal{B}$  since the Stieltjes transform uniquely determines even a complex measure. A similar argument also implies that, for fixed  $B \in \mathcal{B}$ ,  $\varphi_B$  defines a linear functional on  $\mathcal{A}$ .

Since  $v_{\sqrt{y}}(\mathbb{R}) = \langle y \rangle$  for  $y \in \overline{\mathcal{A}}_+$ , we obtain for  $x = (\text{Re } x)_+ - (\text{Re } x)_- + i(\text{Im } x)_+ - i(\text{Im } x)_- \in \mathcal{A}$

$$|\varphi_B(x)| \leq v_{\sqrt{(\text{Re } x)_+}}(\mathbb{R}) + v_{\sqrt{(\text{Re } x)_-}}(\mathbb{R}) + v_{\sqrt{(\text{Im } x)_+}}(\mathbb{R}) + v_{\sqrt{(\text{Im } x)_-}}(\mathbb{R}) \leq \langle (\text{Re } x)_+ + (\text{Re } x)_- + (\text{Im } x)_+ + (\text{Im } x)_- \rangle \leq 2\|x\|_1,$$

where we used that  $(\text{Re } x)_+ + (\text{Re } x)_- = |\text{Re } x|$  and  $(\text{Im } x)_+ + (\text{Im } x)_- = |\text{Im } x|$ . Therefore,  $\varphi_B$  extends to a bounded linear functional on  $L^1$  as  $\mathcal{A}$  is a dense linear subspace of  $L^1$ . Using the isomorphism in (A.1), for each  $B \in \mathcal{B}$ , there exists a unique  $v(B) \in \mathcal{A}$  such that

$$\varphi_B(x) = \langle v(B)x \rangle$$

for all  $x \in \mathcal{A}$ . For  $y \in \mathcal{A}$ , we conclude  $v_y(B) = v_{\sqrt{yy^*}}(B) = \varphi_B(yy^*) = \langle y, v(B)y \rangle \geq 0$ , where we used that  $v_y = v_{\sqrt{yy^*}}$  since their Stieltjes transforms coincide by  $\langle y, h(z)y \rangle = \langle \sqrt{yy^*}, h(z)\sqrt{yy^*} \rangle$  due to the trace property of  $\langle \cdot \rangle$ . Since  $\langle v(B)y \rangle \geq 0$  for all  $y \in \overline{\mathcal{A}}_+$ , we have  $v(B) \in \overline{\mathcal{A}}_+$  for all  $B \in \mathcal{B}$ . Moreover,  $v_x = \langle x, v(\cdot)x \rangle$ , in particular,  $\langle x, v(\mathbb{R})x \rangle = v_x(\mathbb{R}) = \langle x, x \rangle$ , for all  $x \in \mathcal{A}$ . The polarization identity yields that  $v$  is an  $\overline{\mathcal{A}}_+$ -valued measure on  $\mathcal{B}$  satisfying (3.6) and  $v(\mathbb{R}) = \mathbf{1}$ . This completes the proof of Lemma 3.1.  $\square$

LEMMA A.1 (Stieltjes transform inherits Hölder regularity). *Let  $v$  be an  $\overline{\mathcal{A}}_+$ -valued measure on  $\mathbb{R}$  and  $h: \mathbb{H} \rightarrow \mathcal{A}$  be its Stieltjes transform, i.e.,  $h$  satisfies (3.6) for all  $z \in \mathbb{H}$ . Let  $f: I \rightarrow \overline{\mathcal{A}}_+$  be a  $\gamma$ -Hölder continuous function on an interval  $I \subset \mathbb{R}$  with  $\gamma \in (0, 1)$  and  $f$  be a density of  $v$  on  $I$  with respect to the Lebesgue measure, i.e.,*

$$\|f(\tau_1) - f(\tau_2)\| \leq C_0|\tau_1 - \tau_2|^\gamma, \quad v(A) = \int_A f(\tau) d\tau$$

for all  $\tau_1, \tau_2 \in I$ , some  $C > 0$  and for all Borel sets  $A \subset I$ . Moreover, we assume that  $\|f(\tau)\| \leq C_1$  for all  $\tau \in I$ . Let  $\theta \in (0, 1]$ .

Then, for  $z_1, z_2 \in \mathbb{H}$  satisfying  $\text{Re } z_1, \text{Re } z_2 \in I$  and  $\text{dist}(\text{Re } z_k, \partial I) \geq \theta$ ,  $k = 1, 2$ , we have

$$\|h(z_1) - h(z_2)\| \leq \left( \frac{21C_0}{\gamma(1-\gamma)} + \frac{4\|v(\mathbb{R})\|}{\theta^{1+\gamma}} + \frac{14C_1}{\gamma\theta^\gamma} \right) |z_1 - z_2|^\gamma. \tag{A.3}$$

Furthermore, for  $z_1, z_2 \in \mathbb{H}$  satisfying  $\text{dist}(z_k, \text{supp } v) \geq \theta$ ,  $k = 1, 2$ , we have

$$\|h(z_1) - h(z_2)\| \leq \frac{2\|v(\mathbb{R})\|}{\theta^2} |z_1 - z_2|^\gamma. \tag{A.4}$$

We omit the proof of Lemma A.1 since it is very similar to the one of Lemma A.7 in [3].

B POSITIVITY-PRESERVING, SYMMETRIC OPERATORS ON  $\mathcal{A}$

LEMMA B.1. Let  $T: \mathcal{A} \rightarrow \mathcal{A}$  be a positivity-preserving, symmetric operator.

(i) If  $T[a] \leq C\langle a \rangle \mathbf{1}$  for some  $C > 0$  and all  $a \in \overline{\mathcal{A}}_+$  then  $\|T\|_2 \leq C$ . Moreover,  $\|T\|_2$  is an eigenvalue of  $T$  and there is  $x \in \overline{\mathcal{A}}_+ \setminus \{0\}$  such that  $T[x] = \|T\|_2 x$ .

(ii) We assume  $\|T\|_2 = 1$  and that there are  $c, C > 0$  such that

$$c\langle a \rangle \mathbf{1} \leq T[a] \leq C\langle a \rangle \mathbf{1} \tag{B.1}$$

for all  $a \in \mathcal{A}_+$ . Then 1 is an eigenvalue of  $T$  with a one-dimensional eigenspace. There is a unique  $x \in \mathcal{A}_+$  satisfying  $T[x] = x$  and  $\|x\|_2 = 1$ . Moreover,  $x$  is positive definite,

$$cC^{-1/2} \mathbf{1} \leq x \leq C \mathbf{1}. \tag{B.2}$$

Furthermore, the spectrum of  $T$  has a gap of size  $\theta := c^6 / (2(c^3 + 2C^2)C^2)$ , i.e.,

$$\text{Spec}(T) \subset [-1 + \theta, 1 - \theta] \cup \{1\}. \tag{B.3}$$

Lemma B.1 is the analogue of Lemma 4.8 in [4]. Here, we explain how to generalize it to the context of von Neumann algebras. In the proof of Lemma B.1, we will use the following lemma.

LEMMA B.2. Let  $T: \mathcal{A} \rightarrow \mathcal{A}$  be a linear map.

(i) If  $T$  is positivity-preserving such that  $T[a] \leq C\langle a \rangle \mathbf{1}$  for all  $a \in \mathcal{A}_+$  and some  $C > 0$  then  $\|T\| \leq \|T\|_{2 \rightarrow \cdot, \cdot} \leq 2C$ .

(ii) If  $T - \omega \text{Id}$  is invertible on  $\mathcal{A}$  for some  $\omega \in \mathbb{C} \setminus \{0\}$  and  $\|(T - \omega \text{Id})^{-1}\|_2 < \infty$ ,  $\|T\|_{2 \rightarrow \cdot, \cdot} < \infty$  then we have

$$\|(T - \omega \text{Id})^{-1}\| \leq |\omega|^{-1} (1 + \|T\|_{2 \rightarrow \cdot, \cdot} \|(T - \omega \text{Id})^{-1}\|_2).$$

We include the short proof of Lemma B.2 for the reader's convenience. In fact, the first part is obtained as in (4.2) of [4] and the second part as in (5.28) of [3].

Throughout the following arguments, we will use that  $|\langle a \rangle| \leq \langle |a| \rangle \leq \|a\|_2$  for all  $a \in \mathcal{A}$ .

*Proof of Lemma B.2.* Let  $a \in \mathcal{A}$  be self-adjoint, i.e.,  $a = a^*$ . Thus,  $a = a_+ - a_-$  is the sum of its positive and negative part,  $a_+, a_- \in \overline{\mathcal{A}}_+$ . We conclude

$$T[a] \leq T[a_+] + T[a_-] \leq C\langle a_+ + a_- \rangle \leq C\|a\|_2$$

since  $a_+ + a_- = |a|$ . Hence,  $\|T[a]\| \leq C\|a\|_2$  as  $T[a] \geq -C\|a\|_2$  is shown similarly. For a general  $a \in \mathcal{A}$ , we obtain  $\|T[a]\| \leq 2C\|a\|_2$ . As  $\|a\|_2 \leq \|a\|$  this completes the proof of part (i).

For the proof of (ii), we take an arbitrary  $x \in \mathcal{A}$ . We set  $y := (T - \omega\text{Id})^{-1}[x]$ . From the definition of the resolvent, we conclude  $\omega y = T[y] - x$ . This yields

$$\|y\| \leq |\omega|^{-1}(\|T\|_{2 \rightarrow \|\cdot\|} \|y\|_2 + \|x\|) \leq |\omega|^{-1}(1 + \|T\|_{2 \rightarrow \|\cdot\|} \|(T - \omega\text{Id})^{-1}\|_2) \|x\|,$$

where we used  $\|x\|_2 \leq \|x\|$  in the last step. Since  $x$  was arbitrary, we have completed the proof of (ii). □

*Proof of Lemma B.1.* For the proof of (i), we remark that Lemma B.2 (i) implies  $\|T\|_2 \leq \|T\|_{2 \rightarrow \|\cdot\|} \leq 2C$ . Since  $T$  is positivity-preserving, we have  $T[b] \in \mathcal{A}_{\text{sa}}$  for all  $b \in \mathcal{A}_{\text{sa}}$ . For  $a = x + iy \in \mathcal{A}$  with  $x, y \in \mathcal{A}_{\text{sa}}$ , we have that

$$\|T[a]\|_2^2 = \|T[x]\|_2^2 + \|T[y]\|_2^2 \leq \|T|_{\mathcal{A}_{\text{sa}}}\|_2^2(\|x\|_2^2 + \|y\|_2^2) = \|T|_{\mathcal{A}_{\text{sa}}}\|_2^2\|a\|_2^2.$$

Hence,  $\|T|_{\mathcal{A}_{\text{sa}}}\|_2 = \|T\|_2$ . If  $a = a^*$  is self-adjoint then decomposing  $a = a_+ - a_-$  in its positive and negative part yields

$$T[a] \leq T[a_+] + T[a_-] \leq C\langle a_+ + a_- \rangle \leq C\|a\|_2,$$

where we used  $a_+ + a_- = |a|$ . Similarly,  $T[a] \geq -C\|a\|_2$  and, thus,  $\|T\|_2 = \|T|_{\mathcal{A}_{\text{sa}}}\|_2 \leq C$ .

Without loss of generality, we assume  $\|T\|_2 = 1$  in the remainder of the proof. Since  $\|T|_{\mathcal{A}_{\text{sa}}}\|_2 = \|T\|_2$ , we conclude that 1 is contained in the spectrum of  $T: L^2_{\text{sa}} \rightarrow L^2_{\text{sa}}$ , where  $L^2_{\text{sa}} := \overline{\mathcal{A}_{\text{sa}}}^{\|\cdot\|_2}$ , due to the variational principle for the spectrum of self-adjoint operators and  $|\langle b, T[b] \rangle| \leq \langle |b|, T[|b|] \rangle$  for all  $b \in \mathcal{A}_{\text{sa}}$ . This last inequality can be checked easily by decomposing  $b = b_+ - b_-$  into positive and negative part.

Hence, due to the symmetry of  $T$ , there is a sequence  $(y_n)_n$  of approximating eigenvectors in  $\mathcal{A}_{\text{sa}}$ , i.e.,  $y_n \in \mathcal{A}_{\text{sa}}$ ,  $\|y_n\|_2 = 1$  and  $T[y_n] - y_n$  converges to 0 in  $L^2$  for  $n \rightarrow \infty$ . We set  $x_n := |y_n|$ . By using  $\|T|_{L^2_{\text{sa}}}\|_2 = 1$  and  $\langle b, T[b] \rangle \leq \langle |b|, T[|b|] \rangle$  for all  $b \in \mathcal{A}_{\text{sa}}$ , we obtain  $\|T[x_n] - x_n\|_2^2 \leq 2\|y_n\|_2 \|T[y_n] - y_n\|_2$  and, thus,

$$\lim_{n \rightarrow \infty} \|T[x_n] - x_n\|_2 = 0. \tag{B.4}$$

Since the unit ball in the Hilbert space  $L^2$  is relatively sequentially compact in the weak topology, we can assume by possibly replacing  $(x_n)_n$  by a subsequence that there is  $x \in L^2$  such that  $x_n \rightharpoonup x$  weakly in  $L^2$ . From  $T[x_n] \leq C\langle x_n \rangle \mathbf{1}$ , we conclude

$$x_n \leq (\text{Id} - T)[x_n] + C\langle x_n \rangle \mathbf{1}.$$



Multiplying this by  $\sqrt{x_n}$  from the left and the right, applying  $\langle \cdot \rangle$  and using  $\|x_n\|_2 = 1$  yield

$$1 \leq \langle x_n, (\text{Id} - T)[x_n] \rangle + C\langle x_n \rangle^2. \tag{B.5}$$

Taking the limit  $n \rightarrow \infty$  in (B.5), we obtain

$$\langle x \rangle \geq C^{-1/2} \tag{B.6}$$

due to (B.4). Hence,  $x \neq 0$  and we can replace  $x$  by  $x/\|x\|_2$  and  $x_n$  by  $x_n/\|x\|_2$ . For any  $b \in L^2$ , we have

$$\langle b, (\text{Id} - T)[x] \rangle = \lim_{n \rightarrow \infty} \langle b, (\text{Id} - T)[x_n] \rangle = 0$$

due to  $x_n \rightharpoonup x$  and (B.4). Hence,  $T[x] = x$ . Since  $\|T\|_{2 \rightarrow \|\cdot\|} \leq 2C$ , we have  $T[b] \in \mathcal{A}$  for all  $b \in L^2$  and thus  $x = T[x] \in \mathcal{A}$ . Owing to  $x_n \rightharpoonup x$  and  $x_n \in \overline{\mathcal{A}}_+$ , we obtain  $x \in \overline{\mathcal{A}}_+$ . This completes the proof of (i).

We start the proof of (ii) by using (B.1) with  $a = x$  which immediately yields the upper bound in (B.2) since  $\langle x \rangle \leq \|x\|_2 = 1$ . As  $\langle x \rangle \geq C^{-1/2}$  by (B.6), the first inequality in (B.1) then yields the lower bound in (B.2).

In order to prove the spectral gap, (B.3), we remark that  $\|T\|_{2 \rightarrow \|\cdot\|} \leq 2C$  due to the upper bound in (B.1) and Lemma B.2 (i). Hence, by Lemma B.2 (ii), the spectrum of  $T$  as an operator on  $\mathcal{A}$  is contained in the union of  $\{0\}$  and the spectrum of  $T$  as an operator on  $L^2$ . Therefore, we will consider  $T$  as an operator on  $L^2$  in the following and exclusively study its spectrum as an operator on  $L^2$ . Hence, to prove the spectral gap, it suffices to establish a lower bound on  $\langle y, (\text{Id} \pm T)[y] \rangle$  for all self-adjoint  $y \in \mathcal{A}$  satisfying  $\|y\|_2 = 1$  and  $\langle x, y \rangle = 0$ . Fix such  $y \in \mathcal{A}$ . Since  $y$  is self-adjoint we have

$$y = \lim_{N \rightarrow \infty} y^N, \quad y^N := \sum_{k=1}^N \lambda_k^N p_k^N \tag{B.7}$$

for some  $\lambda_k^N \in \mathbb{R}$  and  $p_k^N \in \mathcal{A}$  orthogonal projections such that  $p_k^N p_l^N = p_k^N \delta_{k,l}$ . Here, the convergence  $y^N \rightarrow y$  is with respect to  $\|\cdot\|$ . We can assume that  $\|y^N\|_2 = 1$  for all  $N$  as well as  $\langle p_k^N \rangle > 0$  for all  $k$  and  $\langle p_1^N + \dots + p_N^N \rangle = 1$  for all  $N$ .

We will now reduce estimating  $\langle y, (\text{Id} \pm T)[y] \rangle$  to estimating a scalar product on  $\mathbb{C}^N$ . On  $\mathbb{C}^N$ , we consider the scalar product  $\langle \cdot, \cdot \rangle_N$  induced by the probability measure  $\pi(A) = \sum_{k \in A} \langle p_k^N \rangle$  on  $[N]$ , i.e.,

$$\langle \lambda, \mu \rangle_N = \sum_{k=1}^N \overline{\lambda_k} \mu_k \langle p_k^N \rangle$$

for  $\lambda = (\lambda_k)_{k=1}^N, \mu = (\mu_k)_{k=1}^N \in \mathbb{C}^N$ . The norm on  $\mathbb{C}^N$  and the operator norm on  $\mathbb{C}^{N \times N}$  induced by  $\langle \cdot, \cdot \rangle_N$  are denoted by  $\|\cdot\|_N$  and  $\|\cdot\|$ , respectively.

Moreover,  $\text{Id}_N$  is the identity map on  $\mathbb{C}^N$ . With this notation, we obtain from (B.7) that

$$\begin{aligned} \langle y, (\text{Id} \pm T)[y] \rangle &= \lim_{N \rightarrow \infty} \sum_{k,l=1}^N \lambda_k^N \lambda_l^N \langle p_k^N, (\text{Id} \pm T)[p_l^N] \rangle \\ &= \lim_{N \rightarrow \infty} \langle \lambda^N, (\text{Id}_N \pm S^N)[\lambda^N] \rangle_N, \end{aligned}$$

where we introduced  $\lambda^N = (\lambda_k^N)_{k=1}^N \in \mathbb{C}^N$  and the  $N \times N$  symmetric matrix  $S^N$  viewed as an integral operator on  $([N], \pi)$  with the kernel  $s_{kl}^N$  given by

$$s_{kl}^N = \frac{\langle p_k^N, T[p_l^N] \rangle}{\langle p_k^N \rangle \langle p_l^N \rangle}.$$

Since  $\|y^N\|_2 = 1$ , we have  $\|\lambda^N\|_N = 1$ . By the flatness of  $T$ , we have

$$c \leq s_{kl}^N \leq C. \tag{B.8}$$

In the following, we will omit the  $N$ -dependence of  $\lambda_k$ ,  $s_{kl}$  and  $p_k$  from our notation. By the definition of  $\langle \cdot, \cdot \rangle_N$ , we have

$$\langle \lambda, S\lambda \rangle_N = \sum_{k,l=1}^N \lambda_k \langle p_k \rangle s_{kl} \langle p_l \rangle \lambda_l = \langle y^N, T[y^N] \rangle.$$

Let  $s \in \mathbb{C}^N$  be the Perron-Frobenius eigenvector of  $S$  satisfying  $Ss = \|S\|s$ ,  $\|s\|_N = 1$ . From (B.8), we conclude

$$c \leq \langle e, Se \rangle_N \leq \|S\| = \langle s, Ss \rangle_N \leq \|T\|_2 = 1, \tag{B.9}$$

where  $e = (1, \dots, 1) \in \mathbb{C}^N$ . Since  $\|s\|_N = 1$  and  $c \leq \|S\|$ , we have

$$\max_i s_i = \frac{(Ss)_i}{\|S\|} \leq \frac{C}{c} \sum_{k=1}^N s_k \langle p_k \rangle \leq \frac{C}{c} \left( \sum_{k=1}^N \langle p_k \rangle \right)^{1/2} \left( \sum_{k=1}^N s_k^2 \langle p_k \rangle \right)^{1/2} = \frac{C}{c}.$$

Therefore, applying Lemma 5.7 in [3] with  $\mathfrak{X} = \{1, \dots, N\}$  equipped with the normalized counting measure,  $T = S$ ,  $h = s$  yields

$$\text{Spec}(S) \subset \left[ -\|S\| + \frac{c^3}{C^2}, \|S\| - \frac{c^3}{C^2} \right] \cup \{\|S\|\},$$

where we used that  $\inf_{k,l} s_{k,l} \geq c$  by (B.8) and  $\text{Gap}(S)$  is the difference between the largest and second largest of eigenvalue of the matrix  $|S|$ . We decompose  $\lambda = (1 - \|w\|_N^2)^{1/2} s + w$  with  $w \perp s$  and obtain

$$|\langle \lambda, S\lambda \rangle_N| \leq \|S\| (1 - \|w\|_N^2) + \left( \|S\| - \frac{c^3}{C^2} \right) \|w\|_N^2 \leq 1 - \frac{c^3}{C^2} \|w\|_N^2, \tag{B.10}$$

where we used  $\|S\| \leq 1$  in the last step. Hence, it remains to estimate  $\|w\|_N$ . Recalling  $T[x] = x$ , we set  $\tilde{x} = (\langle xp_k \rangle / \langle p_k \rangle)_{k=1}^N$  and compute

$$\langle x, y^N \rangle = \sum_{k=1}^N \lambda_k \langle xp_k \rangle = \langle \tilde{x}, \lambda \rangle_N.$$

Since the left-hand side goes to  $\langle x, y \rangle = 0$  for  $N \rightarrow \infty$ , we can assume that  $|\langle \tilde{x}, \lambda \rangle_N| \leq \sqrt{\varepsilon}/2$  for any fixed  $\varepsilon \sim 1$  and all sufficiently large  $N$ . As  $\tilde{x}_k \geq c/\sqrt{C}$  by (B.2), we obtain

$$\begin{aligned} (1 - \|w\|_N^2) \frac{c^2}{C} \left( \sum_{k=1}^N s_k \langle p_k \rangle \right)^2 &\leq (1 - \|w\|_N^2) \langle \tilde{x}, s \rangle_N^2 \\ &= (\langle \tilde{x}, \lambda \rangle_N - \langle \tilde{x}, w \rangle_N)^2 \\ &\leq 2\|\tilde{x}\|_N^2 \|w\|_N^2 + \varepsilon. \end{aligned} \tag{B.11}$$

Now, we use  $c \leq \langle s, Ss \rangle_N$  from (B.9) to get

$$c \leq \langle s, Ss \rangle_N = \sum_{k,l=1}^N s_k s_{kl} s_l \langle p_k \rangle \langle p_l \rangle \leq C \left( \sum_{k=1}^N s_k \langle p_k \rangle \right)^2.$$

By plugging this and  $\|\tilde{x}\|_N^2 \leq \|x\|^2 \sum_{k=1}^N \langle p_k \rangle = 1$  into (B.11), solving the resulting estimate for  $\|w\|_N^2$  and choosing  $\varepsilon = c^3/(2C^2)$ , we obtain

$$\|w\|_N^2 \geq \frac{c^3}{2(c^3 + 2C^2)}.$$

Therefore, from (B.10), we conclude

$$|\langle \lambda, S\lambda \rangle_N| \leq 1 - \frac{c^6}{2(c^3 + 2C^2)C^2}$$

uniformly for all sufficiently large  $N \in \mathbb{N}$ . We thus obtain that

$$\langle y, (\text{Id} \pm T)[y] \rangle \geq \frac{c^6}{2(c^3 + 2C^2)C^2}$$

if  $y \perp x$  and  $\|y\|_2 = 1$ . We conclude (B.3), which completes the proof of the lemma.  $\square$

LEMMA B.3. *If  $T: \mathcal{A} \rightarrow \mathcal{A}$  is a positivity-preserving operator such that  $\|T\|_2 < 1$  and  $\|T\|_{2 \rightarrow \|\cdot\|} < \infty$  then  $\text{Id} - T$  is invertible as a bounded operator on  $\mathcal{A}$  and  $(\text{Id} - T)^{-1}$  is positivity-preserving with*

$$(\text{Id} - T)^{-1}[x^*x] \geq x^*x \tag{B.12}$$

for all  $x \in \mathcal{A}$ .

*Proof.* Since  $\|T\|_2 < 1$ ,  $\text{Id} - T$  is invertible on  $L^2$  and we conclude the invertibility of  $\text{Id} - T$  on  $\mathcal{A}$  from Lemma B.2 (ii).

Moreover, for  $y \in \mathcal{A}$  with  $\|y^*y\|_2 < 1$ , we expand the inverse as a Neumann series using  $\|T\|_2 < 1$  and obtain

$$(\text{Id} - T)^{-1}[y^*y] = y^*y + \left( \sum_{k=1}^{\infty} T^k[y^*y] \right) \geq y^*y.$$

The series converges with respect to  $\|\cdot\|_2$ . In the last inequality, we used that  $T^k$  is a positivity-preserving operator for all  $k \in \mathbb{N}$ . Hence, by rescaling a general  $x \in \mathcal{A}$ , we see that  $(\text{Id} - T)^{-1}$  is a positivity-preserving operator on  $\mathcal{A}$  which satisfies (B.12).  $\square$

C NON-HERMITIAN PERTURBATION THEORY

Let  $B_0: \mathcal{A} \rightarrow \mathcal{A}$  be a bounded operator with  $\|B_0\| + \|B_0\|_2 \leq C_1$ . We assume that  $B_0$  has an isolated, single eigenvalue  $\beta_0$  satisfying the following conditions:

- There is  $\varepsilon > 0$  such that  $\text{Spec}_2(B_0) \cap D_\varepsilon(\beta_0) = \{\beta_0\}$  and

$$\sup_{\omega \in \partial D_\varepsilon(\beta_0)} (\|(B_0 - \omega \text{Id})^{-1}\|_2 + \|(B_0 - \omega \text{Id})^{-1}\| + \|(B_0^* - \bar{\omega} \text{Id})^{-1}\|) \leq C_2. \tag{C.1}$$

- Furthermore, the spectral projection

$$P_0 := -\frac{1}{2\pi i} \oint_{\partial D_\varepsilon(\beta_0)} (B_0 - \omega \text{Id})^{-1} d\omega$$

has rank 1.

Let  $b_0$  and  $l_0$  be the normalized,  $\|b_0\|_2 = \|l_0\|_2 = 1$ , right and left eigenvectors of  $B_0$  associated to  $\beta_0$ , respectively, i.e.,  $B_0[b_0] = \beta_0 b_0$  and  $B_0^*[l_0] = \bar{\beta}_0 l_0$ . In that situation, we have

$$P_0 = \frac{\langle l_0, \cdot \rangle}{\langle l_0, b_0 \rangle} b_0. \tag{C.2}$$

We assume that  $\|b_0\|, \|l_0\| \leq C_3$ .

Moreover, we set  $Q_0 := \text{Id} - P_0$ , i.e.,  $P_0$  and  $Q_0$  are the spectral projections corresponding to  $\beta_0$  and  $\text{Spec}(B_0) \setminus \{\beta_0\}$ . Note that they are not orthogonal projections in general.

In the following lemma, we established that the perturbation

$$B = B_0 + E$$

has similar spectral properties than  $B_0$  if  $E: \mathcal{A} \rightarrow \mathcal{A}$  is a bounded operator of sufficiently small norm.

LEMMA C.1. *In the above situation, there is  $c > 0$ , depending only on  $\varepsilon, C_1, C_2$  and  $C_3$ , such that if  $\|E\|_2 + \|E\| + \|E^*\| \leq c$  then the following holds true:*

- (i) *We have  $\text{Spec}_2(B) \cap D_\varepsilon(\beta_0) = \{\beta\}$ , where  $\beta$  is an eigenvalue of  $B = B_0 + E$ . The spectral projection  $P$  of  $B$  corresponding to  $\beta$ , which is defined through*

$$P := -\frac{1}{2\pi i} \oint_{\partial D_\varepsilon(\beta_0)} (B - \omega \text{Id})^{-1} d\omega,$$

*has rank 1.*

- (ii) *The definitions  $b := P[b_0]$  and  $l := P^*[l_0]$  yield eigenvectors of  $B$  and  $B^*$  corresponding to  $\beta$  and  $\bar{\beta}$ , respectively. Moreover, we have*

$$b = b_0 + b_1 + b_2 + \mathcal{O}(\|E\|^3), \quad l = l_0 + l_1 + l_2 + \mathcal{O}(\|E\|^3), \quad (\text{C.3})$$

*where we introduced*

$$\begin{aligned} b_1 &= -Q_0(B_0 - \beta_0 \text{Id})^{-1} E[b_0], \\ b_2 &= Q_0(B_0 - \beta_0 \text{Id})^{-1} E(B_0 - \beta_0 \text{Id})^{-1} Q_0 E[b_0] \\ &\quad - Q_0(B_0 - \beta_0 \text{Id})^{-2} E P_0 E[b_0] - P_0 E Q_0 (B_0 - \beta_0 \text{Id})^{-2} E[b_0], \\ l_1 &= -Q_0^*(B_0^* - \bar{\beta}_0 \text{Id})^{-1} E^*[l_0], \\ l_2 &= Q_0^*(B_0^* - \bar{\beta}_0 \text{Id})^{-1} E^*(B_0^* - \bar{\beta}_0 \text{Id})^{-1} Q_0^* E^*[l_0] \\ &\quad - Q_0^*(B_0^* - \bar{\beta}_0 \text{Id})^{-2} E^* P_0^* E^*[l_0] - P_0^* E^* Q_0^*(B_0^* - \bar{\beta}_0 \text{Id})^{-2} E^*[l_0]. \end{aligned}$$

*In particular, we have  $b_i, l_i = \mathcal{O}(\|E\|^i)$  for  $i = 1, 2$ . Furthermore, we obtain*

$$\begin{aligned} \beta \langle l, b \rangle &= \beta_0 \langle l_0, b_0 \rangle + \langle l_0, E[b_0] \rangle \\ &\quad - \langle l_0, E B_0 (B_0 - \beta_0 \text{Id})^{-2} Q_0 E[b_0] \rangle + \mathcal{O}(\|E\|^3). \end{aligned} \quad (\text{C.4})$$

*The implicit constants in the error terms depend only on  $\varepsilon, C_1, C_2$  and  $C_3$ .*

*Proof.* In this proof, the difference  $B - \omega$  with an operator  $B$  and a scalar  $\omega$  is understood as  $B - \omega \text{Id}$ . We first prove that

$$P = P_0 + P_1 + P_2 + \mathcal{O}(\|E\|^3), \quad (\text{C.5})$$

where we defined

$$\begin{aligned} P_1 &:= -\frac{Q_0}{B_0 - \beta_0} E P_0 - P_0 E \frac{Q_0}{B_0 - \beta_0}, \\ P_2 &:= P_0 E \frac{Q_0}{B_0 - \beta_0} E \frac{Q_0}{B_0 - \beta_0} + \frac{Q_0}{B_0 - \beta_0} E P_0 E \frac{Q_0}{B_0 - \beta_0} \\ &\quad + \frac{Q_0}{B_0 - \beta_0} E \frac{Q_0}{B_0 - \beta_0} E P_0 - \frac{Q_0}{(B_0 - \beta_0)^2} E P_0 E P_0 \\ &\quad - P_0 E \frac{Q_0}{(B_0 - \beta_0)^2} E P_0 - P_0 E P_0 E \frac{Q_0}{(B_0 - \beta_0)^2}. \end{aligned}$$

The analytic functional calculus yields that

$$\begin{aligned}
 P &= -\frac{1}{2\pi i} \oint_{\Gamma} \frac{d\omega}{B - \omega} \\
 &= \frac{1}{2\pi i} \oint_{\Gamma} \left( -\frac{1}{B_0 - \omega} + \frac{1}{B_0 - \omega} E \frac{1}{B_0 - \omega} \right. \\
 &\quad \left. - \frac{1}{B_0 - \omega} E \frac{1}{B_0 - \omega} E \frac{1}{B_0 - \omega} \right) d\omega + \mathcal{O}(\|E\|^3),
 \end{aligned}
 \tag{C.6}$$

where  $\Gamma$  is a closed path that encloses only  $\beta$  and  $\beta_0$  both with winding number +1 but no other element of the spectra of  $B$  and  $B_0$ . Integrating the first summand in the integrand of (C.6) yields  $P_0$ . In the second and third summand, we expand  $\text{Id} = P_0 + Q_0$  in the numerators. Applying an analogue of the residue theorem yields  $P_1$  and  $P_2$  for the second and third summand, respectively. For example, for the second summand, we obtain

$$P_1 = \frac{1}{2\pi i} \oint_{\Gamma} \frac{1}{B_0 - \omega} E \frac{1}{B_0 - \omega} d\omega = -\frac{Q_0}{B_0 - \beta_0} EP_0 - P_0 E \frac{Q_0}{B_0 - \beta_0}.$$

The other two combinations of  $P_0, Q_0$  vanish. Using a similar expansion for the third term, we get (C.5).

Starting from (C.5) as well as observing  $b_i = P_i[b_0]$  and  $l_i = P_i^*[l_0]$  for  $i = 1, 2$ , the relations (C.3) are a direct consequence of the definitions  $b = P[b_0]$  and  $l = P^*[l_0]$  and (C.2).

We will show below that

$$BP = B_0P_0 + B_1 + B_2 + \mathcal{O}(\|E\|^3),
 \tag{C.7}$$

where we defined

$$\begin{aligned}
 B_1 &:= P_0EP_0 - \beta_0 \left( \frac{Q_0}{B_0 - \beta_0} EP_0 + P_0E \frac{Q_0}{B_0 - \beta_0} \right), \\
 B_2 &:= \beta_0 \left( P_0E \frac{Q_0}{B_0 - \beta_0} E \frac{Q_0}{B_0 - \beta_0} + \frac{Q_0}{B_0 - \beta_0} EP_0E \frac{Q_0}{B_0 - \beta_0} \right. \\
 &\quad \left. + \frac{Q_0}{B_0 - \beta_0} E \frac{Q_0}{B_0 - \beta_0} EP_0 \right) - \frac{B_0Q_0}{(B_0 - \beta_0)^2} EP_0EP_0 \\
 &\quad - P_0E \frac{B_0Q_0}{(B_0 - \beta_0)^2} EP_0 - P_0EP_0E \frac{B_0Q_0}{(B_0 - \beta_0)^2}.
 \end{aligned}$$

Now, we obtain (C.4) by applying (C.3) as well as (C.7) to  $\beta \langle l, b \rangle = \langle l, BPb \rangle$ . In order to prove (C.7), we use the analytic functional calculus with  $\Gamma$  as defined above to obtain

$$\begin{aligned}
 BP &= -\frac{1}{2\pi i} \oint_{\Gamma} \frac{\omega d\omega}{B - \omega} \\
 &= \frac{1}{2\pi i} \oint_{\Gamma} \omega \left( -\frac{1}{B_0 - \omega} + \frac{1}{B_0 - \omega} E \frac{1}{B_0 - \omega} \right. \\
 &\quad \left. - \frac{1}{B_0 - \omega} E \frac{1}{B_0 - \omega} E \frac{1}{B_0 - \omega} \right) d\omega + \mathcal{O}(\|E\|^3).
 \end{aligned}$$

Proceeding similarly as in the proof of (C.5) yields (C.7) and thus completes the proof of Lemma C.1.  $\square$

D CHARACTERIZATION OF  $\text{supp } \rho$

The following lemma gives equivalent characterizations of  $\text{supp } \rho$  in terms of  $m$ . Note  $\text{supp } \rho = \text{supp } v$  due to the faithfulness of  $\langle \cdot \rangle$ . We denote the disk of radius  $\varepsilon > 0$  centered at  $z \in \mathbb{C}$  by  $D_\varepsilon(z) := \{w \in \mathbb{C} : |z - w| < \varepsilon\}$ .

LEMMA D.1 (Behaviour of  $m$  on  $\mathbb{R} \setminus \text{supp } \rho$ ). *Let  $m$  be the solution of the Dyson equation (2.3) for a data pair  $(a, S) \in \mathcal{A}_{\text{sa}} \times \Sigma$  with  $\|a\| \leq k_0$  and  $S[x] \leq k_1 \langle x \rangle \mathbf{1}$  for all  $x \in \overline{\mathcal{A}}_+$  and some  $k_0, k_1 > 0$ . Then, for any fixed  $\tau \in \mathbb{R}$ , the following statements are equivalent:*

(i) *There is  $c > 0$  such that*

$$\limsup_{\eta \downarrow 0} \eta \|\text{Im } m(\tau + i\eta)\|^{-1} \geq c.$$

(ii) *There are  $C > 1$  and  $N \subset (0, 1]$  with an accumulation point 0 such that*

$$\begin{aligned} \|m(z)\| \leq C, \quad \|m(z)^{-1}\| \leq C, \\ C^{-1} \langle \text{Im } m(z) \rangle \mathbf{1} \leq \text{Im } m(z) \leq C \langle \text{Im } m(z) \rangle \mathbf{1}, \quad \|F(z)\|_2 \leq 1 - C^{-1} \end{aligned} \tag{D.1}$$

for all  $z \in \tau + iN$ . (The definition of  $F$  was given in (3.4).)

(iii) *There is  $m = m^* \in \mathcal{A}$  such that*

$$\lim_{\eta \downarrow 0} \|m(\tau + i\eta) - m\| = 0. \tag{D.2}$$

Moreover, there is  $C > 0$  such that  $\|m\| \leq C$  and  $\|(\text{Id} - C_m S)^{-1}\| \leq C$ .

(iv) *There are  $\varepsilon > 0$  and an analytic function  $f : D_\varepsilon(\tau) \rightarrow \mathcal{A}$  such that  $f(z) = m(z)$  for all  $z \in D_\varepsilon(\tau) \cap \mathbb{H}$  and  $f(z) = f(\bar{z})^*$  for all  $z \in D_\varepsilon(\tau)$ . In particular,  $f(z) = f(z)^*$  for  $z \in D_\varepsilon(\tau) \cap \mathbb{R}$ .*

*In other words,  $m$  can be analytically extended to a neighbourhood of  $\tau$ .*

(v) *There is  $\varepsilon > 0$  such that  $\text{dist}(\tau, \text{supp } \rho) = \text{dist}(\tau, \text{supp } v) \geq \varepsilon$ .*

(vi) *There is  $c > 0$  such that*

$$\liminf_{\eta \downarrow 0} \eta \|\text{Im } m(\tau + i\eta)\|^{-1} \geq c.$$

All constants in (i) – (vi) depend effectively on each other as well as possibly  $k_0, k_1$  and an upper bound on  $|\tau|$ . For example, in the implication (iii)  $\Rightarrow$  (v),  $\varepsilon$  in (v) can be chosen to depend only on  $k_1$  and  $C$  in (iii).

We remark that  $m$  in (iii) above is invertible and satisfies (2.3) at  $z = \tau$ . As a direct consequence of the equivalence of (i) and (v), we spell out the following simple characterization of  $\text{supp } \rho$ .

**COROLLARY D.2** (Characterization of  $\text{supp } \rho$ ). *Under the conditions of Lemma D.1, we have*

$$\lim_{\eta \downarrow 0} \eta \|\text{Im } m(\tau + i\eta)\|^{-1} = 0. \tag{D.3}$$

if and only if  $\tau \in \text{supp } \rho (= \text{supp } v)$ .

**REMARK D.3.** In the proof of Lemma D.1, the condition  $S[x] \leq k_1 \langle x \rangle \mathbb{1}$  for all  $x \in \overline{\mathcal{A}}_+$  is only used to guarantee the following two weaker consequences: First, this condition implies  $\|S\|_{2 \rightarrow \|\cdot\|} \leq 2k_1$ . Moreover, this condition yields, by Lemma B.1 (i), that  $F = F(\tau + i\eta)$  has an eigenvector  $f \in \overline{\mathcal{A}}_+$  corresponding to  $\|F\|_2$ ,  $Ff = \|F\|_2 f$ , for any fixed  $\tau \in \mathbb{R} \setminus \text{supp } \rho$  and any  $\eta \in (0, 1]$ . If both of these consequences are verified, then the condition  $S[x] \leq k_1 \langle x \rangle \mathbb{1}$  may be dropped from Lemma D.1 without any changes in the proof.

**LEMMA D.4** (Quantitative implicit function theorem). *Let  $X, Y, Z$  be Banach spaces,  $U \subset X$  and  $V \subset Y$  open subsets with  $0 \in U, V$ . Let  $\Phi: U \times V \rightarrow Z$  be continuously Fréchet-differentiable map such that the derivative  $\partial_1 \Phi(0, 0)$  with respect to the first variable has a bounded inverse in the origin and  $\Phi(0, 0) = 0$ . Let  $\delta > 0$  such that  $B_\delta^X \subset U$ ,  $B_\delta^Y \subset V$  and*

$$\sup_{(x,y) \in B_\delta^X \times B_\delta^Y} \|\text{Id}_X - (\partial_1 \Phi(0, 0))^{-1} \partial_1 \Phi(x, y)\| \leq \frac{1}{2}, \tag{D.4}$$

where  $B_\delta^X$  and  $B_\delta^Y$  denote the  $\delta$ -ball around 0 in  $X$  and  $Y$ , respectively. We also assume that

$$\|(\partial_1 \Phi(0, 0))^{-1}\| \leq C_1, \quad \sup_{(x,y) \in B_\delta^X \times B_\delta^Y} \|\partial_2 \Phi(x, y)\| \leq C_2$$

for some constants  $C_1, C_2$ , where  $\partial_2$  denotes the derivative of  $\Phi$  with respect to the second variable. Then there is a constant  $\varepsilon > 0$ , depending only on  $\delta, C_1$  and  $C_2$ , and a unique function  $f: B_\varepsilon^Y \rightarrow B_\delta^X$  such that  $\Phi(f(y), y) = 0$  for all  $y \in B_\varepsilon^Y$ . Moreover,  $f$  is continuously Fréchet-differentiable and if  $\Phi(x, y) = 0$  for some  $(x, y) \in B_\delta^X \times B_\varepsilon^Y$  then  $x = f(y)$ . If  $\Phi$  is analytic then  $f$  will be analytic.

*Proof.* Following the standard proof of the implicit function theorem (see e.g. Theorem 3.4.10 in [33]) on Banach spaces and using the additional assumptions of Lemma D.4 yields this quantitative version of the implicit function theorem. We leave the details to the reader.  $\square$

For  $x, y \in \mathcal{A}$  and  $\omega \in \mathbb{C}$ , we define

$$\Phi_x(y, \omega) := (\text{Id} - C_x S)[y] - \omega x^2 - \frac{\omega}{2}(xy + yx) - \frac{1}{2}(xS[y]y + yS[y]x). \tag{D.5}$$



We remark that  $\Phi_{m(z)}(m(z + \omega) - m(z), \omega) = 0$  for all  $z \in \mathbb{H}$  and  $z + \omega \in \mathbb{H}$  (see (6.9)).

*Proof of Lemma D.1.* Lemma B.2 (i) yields  $\|S\|_{2 \rightarrow \cdot} \lesssim 1$  due to  $S[x] \leq k_1 \langle x \rangle \mathbf{1}$  for all  $x \in \overline{\mathcal{A}}_+$ . Therefore,  $\|a\| \lesssim 1$  and  $\|S\| \leq \|S\|_{2 \rightarrow \cdot} \lesssim 1$  imply that  $\text{supp } v = \text{supp } \rho$  is bounded, i.e.,  $\sup\{|\tau| : \tau \in \text{supp } \rho\} \lesssim 1$  by (2.5a).

First, we assume that (i) holds true. We set  $N := \{\eta \in (0, 1] : \eta \|\text{Im } m(\tau + i\eta)\|^{-1} \geq c/2\}$ . By assumption,  $N$  is nonempty and has 0 as an accumulation point. In particular, we have

$$\|\text{Im } m(z)\| \leq \frac{2\eta}{c}, \quad \eta \mathbf{1} \lesssim \text{Im } m(z) \lesssim \frac{\eta}{c} \mathbf{1} \tag{D.6}$$

for all  $z \in \tau + iN$ . The first bound is a direct consequence of the definition of  $N$ . The second bound follows from (2.4) and the bounded support of  $v$ . Moreover, the first bound immediately implies the third bound. By averaging the two last bounds in (D.6) and using  $\text{Im } m(\tau + i\eta) \lesssim \eta$  for  $\eta \in N$ , we obtain the third and fourth estimates in (D.1). In particular,  $\rho(z) \sim \|\text{Im } m(z)\|$  for  $z \in \tau + iN$ . Owing to (2.4), for any  $z \in \mathbb{H}$  and  $x, y \in L^2$ , we have

$$\begin{aligned} |\langle x, m(z)y \rangle| &\leq \frac{1}{2} \int_{\mathbb{R}} \frac{\langle x, v(d\tau)x \rangle + \langle y, v(d\tau)y \rangle}{|\tau - z|} \\ &\lesssim \frac{1}{\eta} (\langle x, \text{Im } m(z)x \rangle + \langle y, \text{Im } m(z)y \rangle) \\ &\leq \frac{2}{c} (\|x\|_2^2 + \|y\|_2^2). \end{aligned}$$

Here, we used that  $v$  has a bounded support and (2.4) in the second step and the first bound in (D.6) in the last step. This proves the first bound in (D.1). The second estimate in (D.1) is a consequence of (2.3) as well as  $\|a\| \lesssim 1$ ,  $\|S\| \leq \|S\|_{2 \rightarrow \cdot} \lesssim 1$  and the first bound in (D.1). We recall the definitions of  $q = q(z)$  and  $u = u(z)$  in (3.1). Owing to Lemma 4.8 (i), the bounds in (D.1) yield

$$\|q\| \lesssim 1, \quad \|q^{-1}\| \lesssim 1, \quad \text{Im } u \sim \langle \text{Im } u \rangle \mathbf{1} \sim \rho \mathbf{1} \tag{D.7}$$

uniformly for all  $z \in \tau + iN$ . Thus, for all  $x \in \overline{\mathcal{A}}_+$  and  $z = \tau + i\eta$  and  $\eta \in N$ ,  $F = F(z)$  satisfies  $F[x] \lesssim \langle x \rangle \mathbf{1}$  due to  $S[x] \lesssim \langle x \rangle \mathbf{1}$ . Hence, Lemma B.1 (i) yields the existence of an eigenvector  $f \in \overline{\mathcal{A}}_+$ , i.e.,  $Ff = \|F\|_2 f$ . By taking the imaginary part of (3.3) and then the scalar product with  $f$  as well as using the symmetry of  $F$ , we get

$$1 - \|F\|_2 = \eta \frac{\langle f, qq^* \rangle}{\langle f, \text{Im } u \rangle} \sim \eta \|\text{Im } m(z)\|^{-1} \gtrsim c \tag{D.8}$$

for  $z = \tau + i\eta$  and  $\eta \in N$  (compare (4.5)). Here, we also used  $f \in \overline{\mathcal{A}}_+$ , (D.7),  $\rho(z) \sim \|\text{Im } m(z)\|$  and the definition of  $N$ . This completes the proof of (i)  $\Rightarrow$  (ii).

Next, let (ii) be satisfied. As before, Lemma 4.8 (i) implies (D.7) for all  $z \in \tau + iN$  due to the first four bounds in (D.1). Thus, inspecting the proofs of Lemma 4.8 (iii) and Proposition 4.1 and using  $\|S\|_{2 \rightarrow \| \cdot \|} \lesssim 1$  via Lemma B.2 (ii) yield

$$\|(\text{Id} - C_{m(z)}S)^{-1}\| \lesssim 1 \tag{D.9}$$

uniformly for all  $z \in \tau + iN$ . Thus, we can apply the implicit function theorem, Lemma D.4, to  $\Psi_\eta(\Delta, \omega) := \Phi_{m(\tau+i\eta)}(\Delta, \omega)$  ( $\Phi$  has been defined in (D.5)) for each  $\eta \in N$  with  $\omega \in \mathbb{C}$ . Since  $\Psi_\eta(0, 0) = 0$  for all  $\eta \in N$ , there are  $\varepsilon > 0$  and unique analytic functions  $\Delta_\eta: D_\varepsilon(0) \rightarrow B_\delta^{\mathcal{A}}$  by Lemma D.4 such that  $\Psi_\eta(\Delta_\eta(\omega), \omega) = 0$  for all  $\omega \in D_\varepsilon(0)$  and all  $\eta \in N$ . We now explain why  $\varepsilon$  can be chosen uniformly for all  $\eta \in N$ . By (D.1) and (D.9), there are bounds on  $m(z)$  and  $(\text{Id} - C_{m(z)}S)^{-1}$  which hold uniformly for  $z \in \tau + iN$ . Hence, it is easy to find  $\delta > 0$  such that (D.4) holds true uniformly for all  $\eta \in N$ . These uniform bounds yield the uniformity of  $\varepsilon$ . Since 0 is an accumulation point of  $N$ , there is  $\eta_0 \in N$  such that  $\eta_0 < \varepsilon$ . We set  $z := \tau + i\eta_0$ . An easy computation using (2.3) at spectral parameters  $z$  and  $z + \omega$  shows  $\Psi_{\eta_0}(m(\omega + z) - m(z), \omega) = 0$  for all  $\omega \in \mathbb{C}$  such that  $\omega + z \in \mathbb{H}$ . Owing to the continuity of  $m$ , we find  $\varepsilon' \in (0, \varepsilon)$  such that  $m(\omega + z) - m(z) \in B_\delta^{\mathcal{A}}$  for all  $\omega \in D_{\varepsilon'}(0)$ . Thus, by the uniqueness of  $\Delta_{\eta_0}$  (cf. Lemma D.4),  $\Delta_{\eta_0}(\omega) = m(\omega + z) - m(z)$  for all  $\omega \in D_{\varepsilon'}(0)$ . As  $\Delta_{\eta_0}$  and  $m(\cdot + z)$  are analytic, owing to the identity theorem, we obtain  $\Delta_{\eta_0}(\omega) + m(z) = m(\omega + z)$  for all  $\omega \in D_\varepsilon(0)$  satisfying  $\omega + z \in \mathbb{H}$ . Using  $\eta_0 < \varepsilon$ , we set  $m := \Delta_{\eta_0}(-i\eta_0) + m(z)$ . For this choice of  $m$ , the continuity of  $\Delta_{\eta_0}(\omega)$  for  $\omega \rightarrow -i\eta_0$  and  $\Delta_{\eta_0}(\omega) + m(z) = m(\omega + z)$  yield (D.2). It remains to show that  $m$  is self-adjoint. Since (D.7) holds true under (ii) as we have shown above, we obtain

$$\eta \|\text{Im } m(z)\|^{-1} \sim 1 - \|F\|_2 \geq C^{-1}$$

for  $z = \tau + i\eta$  and  $\eta \in N$  as in (D.8). Thus,  $\liminf_{\eta \downarrow 0} \|\text{Im } m(\tau + i\eta)\| \leq 0$ . Hence, we obtain  $\text{Im } m = 0$ , i.e.,  $m = m^*$ . This completes the proof of (ii)  $\Rightarrow$  (iii).

If (iii) holds true then  $\text{Id} - C_m S$  has a bounded linear inverse on  $\mathcal{A}$  for  $m$ . Hence, we can apply the implicit function theorem, Lemma D.4, to  $\Phi_m(\Delta, \omega) = 0$  (see (D.5) for the definition of  $\Phi$ ) as  $\Phi_m(0, 0) = 0$  and  $\partial_1 \Phi_m(0, 0) = \text{Id} - C_m S$ . It is easy to see that there is  $\delta > 0$  such that (D.4) is satisfied. Therefore, there are  $\varepsilon > 0$  and an analytic function  $\Delta: D_\varepsilon(0) \rightarrow B_\delta^{\mathcal{A}}$  such that  $\Phi_m(\Delta(\omega), \omega) = 0$  for all  $\omega \in D_\varepsilon(0)$ . In particular,  $f: D_\varepsilon(\tau) \rightarrow \mathcal{A}$ ,  $f(w) := \Delta(w - \tau) + m$  is analytic. From (D.2) and (2.3), we see that  $m$  is invertible and satisfies (2.3) at  $z = \tau$ . Thus, a straightforward computation using (2.3) at  $z = \tau$  and at  $z = \tau + i\eta$  yields  $\Phi_m(m(\tau + i\eta) - m, i\eta) = 0$  for all  $\eta \in (0, \varepsilon]$ . Therefore,  $m(\tau + i\eta) = \Delta(i\eta) + m = f(\tau + i\eta)$  for all  $\eta \in (0, \eta_*]$  and some  $\eta_* \in (0, \varepsilon]$  due to the uniqueness part of Lemma D.4 and (D.2). Since  $m$  and  $f$  are analytic on  $D_\varepsilon(\tau) \cap \mathbb{H}$ , the identity theorem implies  $m(z) = f(z)$  for all  $z \in D_\varepsilon(\tau) \cap \mathbb{H}$ . A simple computation shows  $\Phi_m(\Delta(\bar{\omega})^*, \omega) = \Phi_m(\Delta(\bar{\omega}), \bar{\omega})^* = 0$  for all  $\omega \in D_\varepsilon(0)$  as  $m = m^*$ . Hence,  $\Delta(\omega) = \Delta(\bar{\omega})^*$  for all  $\omega \in D_\varepsilon(0)$  by the uniqueness part of

Lemma D.4. Thus,  $f(w) = f(\bar{w})^*$  for all  $w \in D_\varepsilon(\tau)$  and  $f(w) = f(w)^*$  for all  $w \in D_\varepsilon(\tau) \cap \mathbb{R}$ . This proves (iii)  $\Rightarrow$  (iv). Clearly, (iv) implies (v) by (2.4). If the statement in (v) holds true then  $\text{dist}(\tau, \text{supp } \rho) \geq \varepsilon$ . In particular, by (3.7), we have

$$\liminf_{\eta \downarrow 0} \eta \|\text{Im } m(\tau + i\eta)\|^{-1} \geq \liminf_{\eta \downarrow 0} \text{dist}(\tau + i\eta, \text{supp } \rho)^2 \geq \varepsilon^2$$

for all  $\eta > 0$ . Here, we used (3.7) in the first step. This immediately implies (vi) with  $c = \varepsilon^2$ . Moreover, (i) is immediate from (vi).

Inspecting the proofs of the implications above shows the additional statement about the effective dependence of the constants in (i) – (vi). In particular, the application of the implicit function theorem, Lemma D.4, in the proof of (iv) shows that  $\varepsilon$  can be chosen to depend only on  $k_1$  and  $C$  from (iii). This completes the proof of Lemma D.1.  $\square$

## REFERENCES

- [1] O. Ajanki, L. Erdős, and T. Krüger, *Singularities of solutions to quadratic vector equations on the complex upper half-plane*, Comm. Pure Appl. Math. 70 (2017), no. 9, 1672–1705.
- [2] O. Ajanki, L. Erdős, and T. Krüger, *Universality for general Wigner-type matrices*, Probab. Theory Related Fields 169 (2017), no. 3-4, 667–727.
- [3] O. Ajanki, L. Erdős, and T. Krüger, *Quadratic vector equations on complex upper half-plane*, Mem. Amer. Math. Soc. 261 (2019), no. 1261, v+133. MR 4031100
- [4] O. Ajanki, L. Erdős, and T. Krüger, *Stability of the matrix Dyson equation and random matrices with correlations*, Probab. Theory Related Fields 173 (2019), no. 1-2, 293–373. MR 3916109
- [5] A. B. Aleksandrov and V. V. Peller, *Operator Lipschitz functions*, Russian Math. Surveys 71 (2016), no. 4, 605–702.
- [6] J. Alt, *Singularities of the density of states of random Gram matrices*, Electron. Commun. Probab. 22 (2017), paper no. 63, 13 pp.
- [7] J. Alt, L. Erdős, T. Krüger, and Yu. Nemish, *Location of the spectrum of Kronecker random matrices*, Ann. Inst. Henri Poincaré Probab. Stat. 55 (2019), no. 2, 661–696. MR 3949949
- [8] J. Alt, L. Erdős, T. Krüger, and D. Schröder, *Correlated random matrices: Band rigidity and edge universality*, Ann. Probab. 48 (2020), no. 2, 963–1001. MR 4089499
- [9] J. Alt, L. Erdős, and T. Krüger, *Local law for random Gram matrices*, Electron. J. Probab. 22 (2017), paper no. 25, 41 pp.

- [10] G. W. Anderson and O. Zeitouni, *A CLT for a band matrix model*, Probab. Theory Related Fields 134 (2006), no. 2, 283–338.
- [11] Z. D. Bai and J. W. Silverstein, *Exact separation of eigenvalues of large-dimensional sample covariance matrices*, Ann. Probab. 27 (1999), no. 3, 1536–1555. MR 1733159
- [12] F. Bekerman, T. Leblé, and S. Serfaty, *CLT for fluctuations of  $\beta$ -ensembles with general potential*, Electron. J. Probab. 23 (2018), paper no. 115, 31 pp. MR 3885548
- [13] F. A. Berezin, *Some remarks on the Wigner distribution*, Theoret. Math. Phys. 17 (1973), 1163–1171.
- [14] R. Bhatia, *Matrix analysis*, Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997. MR 1477662
- [15] G. Cipolloni, L. Erdős, T. Krüger, and D. Schröder, *Cusp universality for random matrices, II: The real symmetric case*, Pure Appl. Anal. 1 (2019), no. 4, 615–707. MR 4026551
- [16] T. Claeys, I. Krasovsky, and A. Its, *Higher-order analogues of the Tracy-Widom distribution and the Painlevé II hierarchy*, Comm. Pure Appl. Math. 63, no. 3, 362–412.
- [17] T. Claeys, A. B. J. Kuijlaars, K. Liechty, and D. Wang, *Propagation of singular behavior for Gaussian perturbations of random matrices*, Comm. Math. Phys. 362 (2018), no. 1, 1–54.
- [18] P. Deift, T. Kriecherbauer, and K. T. R. McLaughlin, *New results on the equilibrium measure for logarithmic potentials in the presence of an external field*, J. Approx. Theory 95 (1998), no. 3, 388–475.
- [19] L. Erdős, T. Krüger, and D. Schröder, *Cusp Universality for Random Matrices, I: Local Law and the Complex Hermitian Case*, Commun. Math. Phys. 378 (2020), no. 2, 1203–1278.
- [20] L. Erdős, T. Krüger, and D. Schröder, *Random matrices with slow correlation decay*, Forum Math. Sigma 7 (2019), paper no. e8, 89 pp. MR 3941370
- [21] L. Erdős and H.-T. Yau, *A dynamical approach to random matrix theory*, Courant Lecture Notes in Mathematics, vol. 28, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.
- [22] A. Garg, L. Gurvits, R. Mendes de Oliveira, and A. Wigderson, *A deterministic polynomial time algorithm for non-commutative rational identity testing*, 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS) (2016), 109–117.

- [23] J. B. Garnett, *Bounded analytic functions*, first ed., Graduate Texts in Mathematics, vol. 236, Springer, New York, 2007. MR 2261424
- [24] V. L. Girko, *Theory of stochastic canonical equations: Volumes I and II*, Mathematics and Its Applications, Springer Netherlands, 2012.
- [25] A. Guionnet, *Large deviations upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices*, Ann. Inst. H. Poincaré Probab. Statist. 38 (2002), no. 3, 341–384.
- [26] U. Haagerup, H. Schultz, and S. Thorbjørnsen, *A random matrix approach to the lack of projections in  $C_{\text{red}}^*(\mathbb{F}_2)$* , Adv. Math. 204 (2006), no. 1, 1–83. MR 2233126
- [27] U. Haagerup and S. Thorbjørnsen, *A new application of random matrices:  $\text{Ext}(C_{\text{red}}^*(\mathbb{F}_2))$  is not a group*, Ann. of Math. (2) 162 (2005), no. 2, 711–775. MR 2183281
- [28] Y. He, A. Knowles, and R. Rosenthal, *Isotropic self-consistent equations for mean-field random matrices*, Probab. Theory Related Fields 171 (2018), no. 1-2, 203–249. MR 3800833
- [29] J. W. Helton, T. Mai, and R. Speicher, *Applications of realizations (aka linearizations) to free probability*, J. Funct. Anal. 274 (2018), no. 1, 1–79. MR 3718048
- [30] J. W. Helton, R. Rashidi Far, and R. Speicher, *Operator-valued semicircular elements: Solving a quadratic matrix equation with positivity constraints*, Int. Math. Res. Not. IMRN (2007), no. 22, Art. ID rnm086.
- [31] A. M. Khorunzhy and L. A. Pastur, *On the eigenvalue distribution of the deformed Wigner ensemble of random matrices*, Spectral operator theory and related topics, Adv. Soviet Math., 19, Amer. Math. Soc., Providence, RI, 1994, pp. 97–127.
- [32] A. Knowles and J. Yin, *Anisotropic local laws for random matrices*, Probab. Theory Related Fields 169 (2017), no. 1-2, 257–352. MR 3704770
- [33] S. G. Krantz and H. R. Parks, *The Implicit Function Theorem – History, Theory, and Applications*, Birkhäuser Basel, 2002.
- [34] T. Mai, R. Speicher, and M. Weber, *Absence of algebraic relations and of zero divisors under the assumption of full non-microstates free entropy dimension*, Adv. Math. 304 (2017), 1080–1107.
- [35] T. Mai, R. Speicher, and S. Yin, *The free field: realization via unbounded operators and Atiyah property*, arXiv:1905.08187, 2019.
- [36] J. A. Mingo and R. Speicher, *Free probability and random matrices*, Fields Institute Monographs, Springer New York, 2017.

- [37] L. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs, vol. 171, American Mathematical Society, Providence, RI, 2011.
- [38] V. I. Paulsen, *Completely bounded maps and operator algebras*, Cambridge Studies in Advanced Mathematics, vol. 78, Cambridge University Press, 2002.
- [39] D. Shlyakhtenko, *Random Gaussian band matrices and freeness with amalgamation*, Int. Math. Res. Not. IMRN (1996), no. 20, 1013–1025.
- [40] D. Shlyakhtenko and P. Skoufranis, *Freely independent random variables with non-atomic distributions*, Trans. Amer. Math. Soc. 367 (2015), no. 9, 6267–6291.
- [41] R. Speicher, *Combinatorial theory of the free product with amalgamation and operator-valued free probability theory*, Mem. Amer. Math. Soc. 132 (1998), no. 627.
- [42] M. Takesaki, *Theory of operator algebras I*, Encyclopaedia of mathematical sciences, no. 124, Springer, 1979.
- [43] D. Voiculescu, *The analogues of entropy and of Fisher’s information measure in free probability theory. II*, Invent. Math. 118 (1994), no. 3, 411–440. MR 1296352
- [44] D. Voiculescu, *Operations on certain non-commutative operator-valued random variables*, Astérisque (1995), no. 232, 243–275, Recent advances in operator algebras (Orléans, 1992). MR 1372537
- [45] D. Voiculescu, *The analogues of entropy and of Fisher’s information measure in free probability theory. V. Noncommutative Hilbert transforms*, Invent. Math. 132 (1998), no. 1, 189–227. MR 1618636
- [46] D. Voiculescu, *The coalgebra of the free difference quotient and free probability*, Int. Math. Res. Not. IMRN (2000), no. 2, 79–106.
- [47] D. Voiculescu, *Free entropy*, Bull. Lond. Math. Soc. 34 (2002), no. 3, 257–278.
- [48] F. J. Wegner, *Disordered system with  $n$  orbitals per site:  $n = \infty$  limit*, Physical Review B 19 (1979).

Johannes Alt  
Section of Mathematics  
University of Geneva  
2-4 Rue du Lièvre  
CH-1211 Genève 4  
Switzerland  
johannes.alt@unige.ch

László Erdős  
IST Austria  
Am Campus 1  
A-3400 Klosterneuburg  
Austria  
lerdos@ist.ac.at

Torben Krüger  
Department of Mathematical Sciences  
University of Copenhagen  
Universitetsparken 5  
DK-2100 København  
Denmark  
tk@math.ku.dk

