

CORRELATED RANDOM MATRICES: BAND RIGIDITY AND EDGE UNIVERSALITY

BY JOHANNES ALT¹, LÁSZLÓ ERDŐS^{2,*}, TORBEN KRÜGER³ AND DOMINIK SCHRÖDER^{2,**}

¹University of Geneva, johannes.alt@unige.ch

²IST Austria, *lerdos@ist.ac.at; **dschroed@ist.ac.at

³University of Bonn, torben.krueger@uni-bonn.de

We prove edge universality for a general class of correlated real symmetric or complex Hermitian Wigner matrices with arbitrary expectation. Our theorem also applies to internal edges of the self-consistent density of states. In particular, we establish a strong form of band rigidity which excludes mismatches between location and label of eigenvalues close to internal edges in these general models.

1. Introduction. Spectral statistics of large random matrices exhibit a remarkably robust universality pattern; the local distribution of eigenvalues is independent of details of the matrix ensemble up to symmetry type. In the bulk of the spectrum, this was first observed by Wigner and formalized by Dyson and Mehta [41] who also computed the correlation functions of the Gaussian ensembles in the 1960s. At the spectral edges, the correct statistics were identified by Tracy and Widom both in the Gaussian unitary ensemble (GUE) and the Gaussian orthogonal ensemble (GOE) [49, 50] in the mid 1990s.

Beyond Gaussian ensembles, the first actual proofs of universality for Wigner matrices took different paths in the bulk and at the edge. While in the bulk, only limited progress was made until a decade ago; the first fairly general edge universality proof by Soshnikov [46] appeared shortly after [49, 50]. The main reason is that edge statistics is accessible via an ingenious but laborious extension of the classical moment method of Wigner. In contrast, the bulk universality required fundamentally new tools based on resolvents and the analysis of the Dyson Brownian motion developed in a series of work [17, 18, 22, 23, 26, 28]. This method, called the *three-step strategy*, is summarized in [27]. In certain cases, parallel results [47, 48] were obtained via the *four moment comparison theorem*.

Despite its initial success [46], the moment method for edge universality seems limited when it comes to generalizations beyond Wigner matrices with i.i.d. entries; the resolvent approach is much more flexible. Its primary goal is to establish *local laws*, that is, proving that the local eigenvalue density on scales slightly above the eigenvalue spacing becomes deterministic as the dimension of the matrix tends to infinity. Refined versions of the local law even identify resolvent matrix elements with a spectral parameter very close to the real axis. In contrast to the bulk, at the spectral edge this information can be boosted to detect individual eigenvalue statistics by comparison with the Gaussian ensemble. These ideas have led to the proof of the Tracy–Widom edge universality for Wigner matrices with high moment conditions [28]; see also [47] with vanishing third moment. Finally, a necessary and sufficient condition on the entry distributions was found in [40] following an almost optimal necessary condition in [7]. Direct resolvent comparison methods have been used to prove Tracy–Widom universality for *deformed Wigner matrices*, that is, matrices with a deterministic diagonal expectation, [37], even in a certain sparse regime [39]. The extension of this approach to sample

Received December 2018; revised May 2019.

MSC2010 subject classifications. 60B20, 15B52.

Key words and phrases. Local law, spectral universality, Tracy–Widom distribution.

covariance matrices with a diagonal population covariance matrix at extreme edges [38] has resolved a long standing conjecture in the statistics literature. Tracy–Widom universality for general population covariance matrices, including internal edges, was established in [33].

The next level of generality is to depart from the i.i.d. case. While the resolvent method for proving local laws can handle *generalized Wigner ensemble*, that is, matrices $H = (h_{ab})$ with merely stochastic variance profile $\sum_b \mathbf{Var} h_{ab} = 1$, varying variances cannot be simultaneously matched with a GUE/GOE ensemble so the direct comparison does not work. The problem was resolved in [13] with a general approach that also covered invariant β -ensembles. While Dyson Brownian motion did not play a direct role in [13], the proof used the addition of a small Gaussian component and the concept of local ergodicity of the Gibbs state; ideas developed originally in [23, 24] in the context of bulk universality.

A fully dynamical approach to edge universality, following an earlier development in the bulk based on the *three-step strategy*, has recently been given in [35]. In general, the first step within any three-step strategy is the local law providing a priori bounds. The second step is the fast relaxation to equilibrium of the Dyson Brownian motion that proves universality for Gaussian divisible ensembles. The third step is a perturbative comparison argument to remove the small Gaussian component. Recent advances in the bulk have crystallized that the only model dependent step in this strategy is the first one. The other two steps have been formulated as very general “black-box” tools whose only input is the local law; see [25, 34–36]. Using the three-step approach and [35], edge universality for sparse matrices was proved in [30] and for correlated Gaussian matrices with a quite specific two-scale correlation structure in [1]. All of these edge universality results only cover the *extremal edges* of the spectrum, while the self-consistent (deterministic) density of states ϱ may be supported on several intervals.

Multiple interval support becomes ubiquitous for *Wigner-type* matrices [3], that is, matrices with independent entries and general expectation and variance profile. A prerequisite for Tracy–Widom universality, the square root singularity in the density, even at the *internal edges*, is a universal phenomenon for a very large class of random matrices since it is inherent to the underlying *Dyson equation*. This was demonstrated for Wigner-type matrices in [2] and here we extend it for correlated random matrices with a general correlation structure. We remark that a second singularity type, the *cubic root cusp*, is also possible; the corresponding analysis of the Dyson equation is given in [5], while the optimal local law and the universal spectral statistics are proven in [16, 20].

In the current paper, we show that the eigenvalue statistics at the spectral edges of ϱ follow the Tracy–Widom distribution, assuming only a mild decay of correlation between entries, but otherwise no special structure. More specifically, we assume that the covariance of entries exhibits a power law decay with exponent $s > 12$ in the general setting, and with exponent $s \geq 2$ in the Gaussian setting. In the Gaussian case with two scale correlation structure, the decay exponent $s > 2$ has already been achieved in [1].

We can handle any internal edge as well. In the literature, internal edge universality for matrices of Wigner-type has first been established for deformed GUE ensembles [45] which critically relied on contour integral methods, only available for Gaussian models in the Hermitian symmetry class. A similar method handled extreme eigenvalues of deformed GUE [14, 31]. A more general approach for internal edges has been given in [33] that could handle any deformed Wigner matrices with general expectation, as long as the variance profile is constant, by comparing it with the corresponding Gaussian model. Our method requires neither constant variance nor independence of the matrix elements.

The proof of our general form of edge universality at all internal edges follows the three-step strategy and uses the recent paper [35] for the second step and well-established canonical arguments for the third step that will be summarized. The backbone of the work is thus

the first step, an optimal local law at the spectral edges, the proof of which has two well-separated components: a probabilistic and a deterministic one. The probabilistic component is insensitive to the location in the spectrum and follows directly from [21]. Here, we present a compact and practically self-contained proof of the deterministic component of the local law that can be followed without consulting previous works; we only rely on some general results from functional analysis proven in [4] and some minor technicality on the Dyson equation from [5]. First, we develop a detailed shape analysis of the self-consistent density ϱ near the regular edges, generalizing the previous bulk result from [4] and the singularity analysis in the independent case from [2]. This extension of the singularity analysis to the noncommutative matrix Dyson equation constitutes the most involved part of the proof and contains the main technical novelty of the present work. Second, we prove a strong version of the local law that excludes eigenvalues in the internal gaps. Third, we establish a topological rigidity phenomenon for the *bands*, the connected components that constitute the support of ϱ .

Band rigidity is a new phenomenon for the Dyson equation and it asserts that the number of eigenvalues within each band exactly matches the mass that ϱ predicts for that band. The topological nature of band rigidity guarantees that this mass remains constant along the deformations of the model as long as the gaps between the bands remain open. A similar rigidity (also called “exact separation of eigenvalues”) has first been established for sample covariance matrices in [8] and it also played a key role in the Tracy–Widom universality proof at internal edges in [33]. Note that band rigidity is a much stronger concept than the customary rigidity in random matrix theory [28] that allows for an uncertainty in the location of N^ϵ eigenvalues. In other words, there is no mismatch whatsoever between location and label of the eigenvalues near the internal edges along the matrix Dyson Brownian motion, the label of the eigenvalue uniquely determines to which spectral band it belongs.

Our result highlights a key difference between Wigner-type matrix models and invariant β -ensembles. For self-consistent densities with multiple support intervals (the so-called *multicut* regime), the number of particles (eigenvalues) close to some support interval fluctuates for invariant ensembles with general potentials [12]. As a consequence, internal edge universality results (see, e.g., [9, 43]) require a stochastic relabeling of eigenvalues.

Our setup is a general $N \times N$ random matrix $H = H^*$ with a slowly decaying correlation structure and arbitrary expectation, under the very same general conditions as the recent bulk universality result from [21]. The starting point is to find the deterministic approximation of the resolvent $G(z) = (H - z)^{-1}$ with a complex spectral parameter z in the upper half-plane. This approximation is given by the solution $M = M(z)$ to the *Matrix Dyson Equation (MDE)*; see (1) below. The resolvent $G(z)$ approximately satisfies the MDE with an additive perturbation term which was already shown to be sufficiently small in [21]. This fact, combined with a careful stability and shape analysis of the MDE in Section 4 imply that G is indeed close to M . In order to prove edge universality, we use a correlated Ornstein–Uhlenbeck process H_t which adds a small Gaussian component of size t to the original matrix model, while preserving expectation and covariance. We prove that the resolvent satisfies the optimal local law uniformly along the flow and appeal to the recent result from [35] to prove edge universality for H_t whenever $t \gg N^{-1/3}$. In the final step, we perform a resolvent comparison together with our band rigidity to show that the eigenvalue correlation functions of H_t match those of H as long as $t \ll N^{-1/6}$ which yields the desired edge universality.

After presenting our main results in Section 2, we then prove the optimal local law in Section 3. Section 4 contains the analysis of the MDE. Both types of rigidity are shown in Section 5. Section 6 is devoted to the proof of edge universality.

Notation. If for some constants $c, C > 0$ it holds that $f \leq Cg$ or $cg \leq f \leq Cg$, then we write $f \lesssim g$ and $f \sim g$, respectively. These constants c, C may depend on some basic parameters which we call model parameters later. We use the relations \lesssim and \sim also for Hermitian matrices for which they are meant in the sense of quadratic forms. We denote vectors by bold-faced lower case Roman letters $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$ and matrices by upper case Roman letters $A, B \in \mathbb{C}^{N \times N}$. The standard scalar product and Euclidean norm on \mathbb{C}^N will be written as $\langle \mathbf{x}, \mathbf{y} \rangle$ and $\|\mathbf{x}\|$, while we also write $\langle A, B \rangle := N^{-1} \text{Tr } A^* B$ for the scalar product of matrices, and $\langle A \rangle := N^{-1} \text{Tr } A$. The usual operator norm induced by the vector norm $\|\cdot\|$ will be denoted by $\|A\|$, while the Hilbert–Schmidt (or Frobenius) norm will be denoted by $\|A\|_{\text{hs}} := \sqrt{\langle A, A \rangle}$. The operator norms induced on linear maps $\mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$ by $\|\cdot\|_{\text{hs}}$ and $\|\cdot\|$ are denoted by $\|\cdot\|_{\text{sp}}$ and $\|\cdot\|$, respectively. The identity matrix in $\mathbb{C}^{N \times N}$ is indicated by I and the identity mapping on $\mathbb{C}^{N \times N}$ by Id . For random variables X, Y, \dots , we denote the joint cumulant by $\kappa(X, Y, \dots)$. For integers n , we define $[n] := \{1, \dots, n\}$.

2. Main results. We consider correlated real symmetric and complex Hermitian random matrices of the form

$$H = A + W, \quad \mathbf{E}W = 0$$

with deterministic $A \in \mathbb{C}^{N \times N}$ and sufficiently fast decaying correlations among the matrix elements of W . The matrix entries $w_{ab} = w_\alpha$ are often labeled by double indices $\alpha = (a, b) \in [N]^2$. The randomness W is scaled in such a way that $\sqrt{N}w_\alpha$ are random variables of order one.¹ This requirement ensures that the size of the spectrum of H is kept of order 1, as N tends to infinity. Our first aim is to prove that the resolvent $G = G(z) = (H - z)^{-1}$ is well approximated by the solution $M = M(z)$ to the *Matrix Dyson equation (MDE)*

$$(1) \quad I + (z - A + \mathfrak{S}[M])M = 0, \quad \Im M := \frac{M - M^*}{2i} > 0, \quad \mathfrak{S}[R] := \mathbf{E}W R W$$

for $z \in \mathbb{H} := \{z \in \mathbb{C} | \Im z > 0\}$ in a neighborhood around the edges of the spectrum. We suppress the dependence of G and M , and similarly of many other quantities, on the spectral parameter z in our notation. Estimates on z -dependent quantities are always meant uniformly for z in some specified domain. From the solution M , we define $\varrho: \mathbb{H} \rightarrow \mathbb{R}$ and extend it to the real line

$$(2) \quad \varrho(z) := \frac{1}{\pi} \Im \langle M(z) \rangle, \quad z \in \mathbb{H}, \quad \varrho(\tau) := \lim_{\eta \searrow 0} \varrho(\tau + i\eta), \quad \tau \in \mathbb{R}.$$

By [4], Proposition 2.2, the limit in (2) exists and ϱ is a Hölder continuous function on $\mathbb{H} \cup \mathbb{R}$ under Assumptions (A) and (E) below. The *self-consistent density of states* is the restriction of ϱ to \mathbb{R} which approximates the density of eigenvalues of H increasingly well as N tends to infinity. Its support, $\text{supp } \varrho \subset \mathbb{R}$, is called the *self-consistent spectrum*. We remark that ϱ on \mathbb{H} is the harmonic extension of $\varrho|_{\mathbb{R}}$. We now list our main assumptions, which are identical to those from [21], apart from the additional Assumption (G), which was automatically satisfied in [21], that is, in the bulk regime (cf. Remark 2.3 below). All constants in Assumptions (A)–(G) and Definition 2.4 are called *model parameters*.

ASSUMPTION (A) (Bounded expectation). There exists some constant C such that $\|A\| \leq C$ for all N .

ASSUMPTION (B) (Finite moments). For all $q \in \mathbb{N}$, there exists a constant μ_q such that $\mathbf{E}|\sqrt{N}w_\alpha|^q \leq \mu_q$ for all α .

¹In some previous works, as in [21], the convention $H = A + W/\sqrt{N}$ with order one w_α was used.

ASSUMPTION (CD) (Polynomially decaying metric correlation structure). For the $k = 2$ point correlation, we assume

$$(3a) \quad |\kappa(f_1(\sqrt{N}W), f_2(\sqrt{N}W))| \leq C_2 \frac{\sqrt{\mathbf{E}|f_1(\sqrt{N}W)|^2} \sqrt{\mathbf{E}|f_2(\sqrt{N}W)|^2}}{1 + d(\text{supp } f_1, \text{supp } f_2)^s},$$

for some $s > 12$ and all square integrable functions f_1, f_2 . For $k \geq 3$, we assume a decay condition of the form

$$(3b) \quad |\kappa(f_1(\sqrt{N}W), \dots, f_k(\sqrt{N}W))| \leq C_k \prod_{e \in E(T_{\min})} |\kappa(e)|,$$

where T_{\min} is the minimal spanning tree in the complete graph on the vertices $1, \dots, k$ with respect to the edge length $\text{dist}(\{i, j\}) = d(\text{supp } f_i, \text{supp } f_j)$, that is, the tree for which the sum of the lengths $\text{dist}(e)$ is minimal, and $\kappa(\{i, j\}) = \kappa(f_i, f_j)$. Here, d is the standard Euclidean metric on the index space $[N]^2$ and $\text{supp } f \subset [N]^2$ denotes the set indexing all entries in $\sqrt{N}W$ that f genuinely depends on, and $C_k < \infty$ are some absolute constants.

REMARK 2.1. All results in this paper and their proofs hold verbatim if Assumption (CD) is replaced by the more general assumptions (C), (D) from [21]. In particular, the metric structure imposed on the index space $[N]^2$ is not essential and, in the Gaussian case it is sufficient to assume that the decay exponent satisfies $s \geq 2$, cf. [21], Example 2.12. For details, the reader is referred to [21], Section 2.1.

ASSUMPTION (E) (Flatness). There exist constants $0 < c < C$ such that $c\langle T \rangle \leq \mathcal{S}[T] \leq C\langle T \rangle$ for any positive semidefinite matrix T .

ASSUMPTION (F) (Fullness). There exists a constant $\lambda > 0$ such that $N\mathbf{E}|\text{Tr } BW|^2 \geq \lambda \text{Tr } B^2$ for any deterministic matrix B of the same symmetry class (either real symmetric or complex Hermitian) as H .

ASSUMPTION (G) (Bounded self-consistent Green function). There exist constants $\omega_*, M_* > 0$ such that

$$\sup_z \|M(z)\| \leq M_*,$$

where the supremum is taken over all $z \in \mathbb{H}$ with $|\Re z - \tau_0| \leq \omega_*$ and $0 < \Im z \leq 1$.

REMARK 2.2. Assumption (E) is an effective mean field condition that provides upper and lower bounds on the variances of the entries of W . In fact, it is equivalent to $\mathbf{E}|\langle \mathbf{x}, W\mathbf{y} \rangle|^2 \sim 1/N$ for all normalized $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$. Assumption (F) is equivalent to $\mathcal{S} - \lambda S_G$ remaining positivity preserving, where S_G is the self-energy operator of a full GUE/GOE matrix.

REMARK 2.3. The boundedness of $\|M\|$ is automatically satisfied in the spectral bulk. At the edges, however, the boundedness cannot be guaranteed under Assumptions (A)–(E) but has to be verified for each concrete model (see [5], Section 9, for a large class of models for which $\|M\|$ is guaranteed to be bounded).

Our main technical result is an optimal local law at *regular edges* $\tau_0 \in \partial \text{supp } \varrho$ asserting that $G(z) = (H - z)^{-1}$ is well approximated by $M(z)$ in the $N \rightarrow \infty$ limit. Around such an edge, we consider the domain of spectral parameters $z = \tau + i\eta$ whose imaginary part $\Im z = \eta$ is slightly larger than $1/N$, that is, in the spectral domain

$$(4) \quad \mathbb{D}_\gamma^\delta := \{z \in \mathbb{D}^\delta \mid \Im z \geq N^{-1+\gamma}\}, \quad \mathbb{D}^\delta := \{\tau + i\eta \mid |\tau - \tau_0| \leq \delta, 0 < \eta \leq 1\}$$

for any $\gamma, \delta > 0$.

DEFINITION 2.4 (Regular edge). We call an edge $\tau_0 \in \partial \text{supp } \varrho$ *regular* if the limit

$$(5) \quad \lim_{\text{supp } \varrho \ni \tau \rightarrow \tau_0} \frac{\varrho(\tau)}{\sqrt{|\tau - \tau_0|}} = \frac{\gamma_{\text{edge}}^{3/2}}{\pi}$$

exists for some *slope parameter* γ_{edge} that satisfies $0 < c_* \leq \gamma_{\text{edge}} \leq c^* < \infty$ for some constants c_*, c^* .

REMARK 2.5. We remark that there are several equivalent characterizations of *regular edges*. We chose (5) here because it highlights that the essential prerequisite for Tracy–Widom universality is a local square-root singularity. According to the classification result from [5], it follows that (5) is equivalent² to assuming that the gap in $\text{supp } \varrho$ adjacent to τ_0 is of size $\gtrsim 1$.

THEOREM 2.6 (Edge local law). *Let Assumptions (A)–(E) and (G) be satisfied for some regular edge $\tau_0 \in \partial \text{supp } \varrho$. Then for any $D, \gamma, \epsilon > 0$ and sufficiently small $\delta > 0$, there exists some $C < \infty$ depending only on these and the model parameters such that with $G = G(z)$ and $M = M(z)$ we have the isotropic local law*

$$(6a) \quad \mathbf{P}\left(\left|\langle \mathbf{x}, (G - M)\mathbf{y} \rangle\right| \leq N^\epsilon \|\mathbf{x}\| \|\mathbf{y}\| \left(\sqrt{\frac{\varrho}{N\Im z}} + \frac{1}{N\Im z}\right) \text{ in } \mathbb{D}_\gamma^\delta\right) \geq 1 - CN^{-D}$$

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

$$(6b) \quad \mathbf{P}\left(\left\|B(G - M)\right\| \leq N^\epsilon \frac{\|B\|}{N\Im z} \text{ in } \mathbb{D}_\gamma^\delta\right) \geq 1 - CN^{-D}$$

for all deterministic matrices $B \in \mathbb{C}^{N \times N}$. Moreover, at a distance at least $N^{-2/3+\epsilon}$ away from the self-consistent spectrum, we have the improved averaged local law for any $\epsilon > 0$

$$(6c) \quad \mathbf{P}\left(\left\|B(G - M)\right\| \leq \frac{N^\epsilon \|B\|}{N \text{dist}(z, \text{supp } \varrho)} \text{ in } \left\{z \in \mathbb{D}^\delta \mid \frac{\text{dist}(z, \text{supp } \varrho)}{N^{-2/3+\epsilon}} \geq 1\right\}\right) \geq 1 - CN^{-D}$$

with C also depending on ϵ .

COROLLARY 2.7 (No eigenvalues outside the support of the self-consistent density). *Under the assumptions of Theorem 2.6, we have for any $\epsilon, D > 0$ and sufficiently small $\delta > 0$*

$$\mathbf{P}(\exists \lambda \in \text{Spec } H \mid |\tau_0 - \lambda| \leq \delta, \text{dist}(\lambda, \text{supp } \varrho) \geq N^{-2/3+\epsilon}) \leq_{\epsilon, D} N^{-D},$$

where $\leq_{\epsilon, D}$ means a bound up to some multiplicative constant $C = C(\epsilon, D)$.

COROLLARY 2.8 (Delocalization). *Under the assumptions of Theorem 2.6, it holds for a ℓ^2 -normalized eigenvector \mathbf{u} corresponding to an eigenvalue λ of H close to the edge τ_0 that*

$$\sup_{\|\mathbf{x}\|=1} \mathbf{P}\left(\left|\langle \mathbf{x}, \mathbf{u} \rangle\right| \geq \frac{N^\epsilon}{\sqrt{N}} \mid H\mathbf{u} = \lambda\mathbf{u}, \|\mathbf{u}\| = 1, |\tau_0 - \lambda| \leq \delta\right) \leq_{\epsilon, D} N^{-D}$$

for any $\epsilon, D > 0$ and sufficiently small $\delta > 0$.

²In fact, in [5], Section 7.6, it is proven that if the self-consistent spectrum $\text{supp } \varrho$ has a macroscopic gap next to some $\tau_0 \in \partial \text{supp } \varrho$, then ϱ has a square root behavior at τ_0 . Together with Theorem 4.1 later, this shows that regular edges in the sense of (5) are precisely those $\tau_0 \in \partial \text{supp } \varrho$ which are adjacent to macroscopic gaps.

COROLLARY 2.9 (Band rigidity and eigenvalue rigidity). *Under the assumptions of Theorem 2.6, the following holds. For any $\epsilon, D > 0$, there exists some $C < \infty$ such that for any $\tau \in \mathbb{R} \setminus \text{supp } \varrho$ with $\text{dist}(\tau, \text{supp } \varrho) \geq \epsilon$ the number of eigenvalues less than τ is with high probability deterministic, that is, that*

$$(7a) \quad \mathbf{P}\left(|\text{Spec } H \cap (-\infty, \tau)| = N \int_{-\infty}^{\tau} \varrho(x) dx\right) \geq 1 - CN^{-D}.$$

We also have the following strong form of eigenvalue rigidity in a neighborhood of a regular edge τ_0 . Let $\lambda_1 \leq \dots \leq \lambda_N$ be the ordered eigenvalues of H and denote the index of the N -quantile close to energy $\tau \in \text{int}(\text{supp } \varrho)$ by $k(\tau) := \lceil N \int_{-\infty}^{\tau} \varrho(x) dx \rceil$. It then holds that

$$(7b) \quad \mathbf{P}\left(\sup_{\tau} |\lambda_{k(\tau)} - \tau| \geq \min\left\{\frac{N^\epsilon}{N|\tau - \tau_0|^{1/2}}, \frac{N^\epsilon}{N^{2/3}}\right\}\right) \leq_{\epsilon, D} N^{-D}$$

for any $\epsilon, D > 0$ and sufficiently small $\delta > 0$, where the supremum is taken over all $\tau \in \text{supp } \varrho$ such that $|\tau - \tau_0| \leq \delta$.

REMARK 2.10 (Integer mass). Note that (7a) entails the nontrivial fact that for $\tau \notin \text{supp } \varrho$, $N \int_{-\infty}^{\tau} \varrho(x) dx$ is always an integer; see Proposition 5.1 below. Moreover, it then trivially implies that $N \int_a^b \varrho(x) dx$ is an integer for each spectral band $[a, b]$, that is, connected component of $\text{supp } \varrho$. Finally, (7a) also shows that the number of eigenvalues in each band is given by this integer with overwhelming probability. This is in sharp contrast to invariant β -ensembles where no such mechanism is present. For example, for an odd number of particles in a symmetric double-well potential, $N \int_{-\infty}^0 \varrho(x) dx = N/2$ is a half-integer.

The main application of the optimal local law from Theorem 2.6 is edge universality, as stated in the following theorem, generalising several previous edge universality results listed in the Introduction. For definiteness, we only state and prove the result for regular right edges. The corresponding statement for left edges can be proven along the same lines.

THEOREM 2.11 (Edge universality). *Under the Assumptions (A)–(G) the following statement holds true. Assume that $\tau_0 \in \partial \text{supp } \varrho$ is a right regular edge of ϱ with slope parameter γ_{edge} as in Definition 2.4. The integer (see Remark 2.10) $i_0 := N \int_{-\infty}^{\tau_0} \varrho(x) dx$ labels the largest eigenvalue λ_{i_0} close to the band edge τ_0 with high probability. Furthermore, for test functions $F : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$ such that $\|F\|_\infty + \|\nabla F\|_\infty \leq C < \infty$ we have*

$$\begin{aligned} & \left| \mathbf{E}F(\gamma_{\text{edge}}N^{2/3}(\lambda_{i_0} - \tau_0), \dots, \gamma_{\text{edge}}N^{2/3}(\lambda_{i_0-k} - \tau_0)) \right. \\ & \left. - \mathbf{E}F(N^{2/3}(\mu_N - 2), \dots, N^{2/3}(\mu_{N-k} - 2)) \right| \lesssim N^{-c} \end{aligned}$$

for some $c = c_k > 0$. Here, μ_1, \dots, μ_N are the eigenvalues of a standard GUE/GOE matrix, depending on the symmetry class of H .

From Theorem 2.11, we can immediately conclude that the eigenvalues of H near the regular edges follow the Tracy–Widom distribution. We remark that the direct analogue of Theorem 2.11 does not hold true for invariant β -ensembles with a *multicut* density. This is due to the fact that the number of particles close to a band of the self-consistent density, commonly known as the *filling fraction*, is known to be a fluctuating quantity for general classes of potentials. We refer the reader to [11] for a description of this phenomenon, to [42, 44] for non-Gaussian linear statistics in the multicut regime and to [12] for results on the fluctuations of filling fractions. Variants of Theorem 2.11 which allow for a relabeling of eigenvalues for invariant β -ensembles can be found in [9, 43].

3. Proof of the local law. The proof of a local law consists of three largely separate arguments. The first part concerns the analysis of the *stability operator*

$$(8) \quad \mathcal{B}[R] := R - MS[R]M$$

for $R \in \mathbb{C}^{N \times N}$, and shape analysis of the solution M to (1). The second part is proving that the resolvent G is indeed an approximate solution to (1) in the sense that

$$(9) \quad D := I + (z - A + S[G])G = WG + S[G]G$$

is small. Finally, the third part consists of a bootstrap argument starting in the domain \mathbb{D}_1^δ and iteratively increasing the domain to \mathbb{D}_γ^δ while maintaining the desired bound on $G - M$.

3.1. *Stability.* From (1) and (9), we see that the difference between G and M is described by the relation

$$(10) \quad \mathcal{B}[G - M] = -MD + MSG - M.$$

To prove estimates on $G - M$, we need to analyze \mathcal{B} , the stability operator. Near the edge, we will demonstrate that \mathcal{B} has a very small (in absolute value) simple eigenvalue, that we will denote by β , and it turns out that β is well separated away from the rest of the spectrum of \mathcal{B} . Let P and B denote the corresponding left and right eigenvectors of \mathcal{B} , that is, $\mathcal{B}^*[P] = \bar{\beta}P$ and $\mathcal{B}[B] = \beta B$, and we will specify their normalization later. Note that \mathcal{B} is typically not self-adjoint, so $P \neq B$. Since β is small, \mathcal{B}^{-1} is unstable in the direction of the eigenspace of β . We therefore separate this unstable direction by writing $G - M = \Theta B + \text{Error}$ where

$$(11) \quad \Theta := \frac{\langle P, G - M \rangle}{\langle P, B \rangle}$$

is the key quantity and the error term lies in spectral subspace complementary to B . We will then establish bounds in terms of Θ and D from (10). We note that this separation is not necessary in the bulk regime studied in [21], where the stability operator is bounded in every direction, which explains the additional complexity of the proof of Theorem 2.6 compared to the bulk local law in [21].

The reader should not be confused by the term ‘‘eigenvector’’ in the context of operators $\mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$ as eigenvectors are in fact matrices in this setting, for example, the eigenvectors P and B of \mathcal{B} above are actually matrices in $\mathbb{C}^{N \times N}$.

We begin by collecting some qualitative and quantitative information about the MDE and its stability operator, which will be proven in Section 4.5 below. We note that (i) was first obtained in [29] and (ii) goes back to [4].

PROPOSITION 3.1 (Stability of MDE and properties of the solution). *The following hold true under Assumption (A), (E) and (G) for some $\tau_0 \in \mathbb{R}$:*

(i) *The MDE (1) has a unique solution $M = M(z)$ for all $z \in \mathbb{H}$ and, moreover, the map $z \mapsto M(z)$ is holomorphic.*

(ii) *The holomorphic function $\langle M \rangle : \mathbb{H} \rightarrow \mathbb{H}$ is the Stieltjes transform of a compactly supported probability measure with continuous density $\varrho : \mathbb{R} \rightarrow [0, \infty)$ given by (2). Moreover, ϱ is real analytic on the open set $\{\varrho > 0\}$.*

If $\tau_0 \in \partial \text{supp } \varrho$ is a regular edge, then there is $\delta_ \sim 1$ such that, for all $z \in \mathbb{H}$ satisfying $|z - \tau_0| \leq \delta_*$, we have*

(iii) *The harmonic extension of the self-consistent density of states scales like*

$$\varrho(z) \sim \begin{cases} \sqrt{\kappa + \eta} & \text{if } \tau \in \text{supp } \varrho, \\ \eta / \sqrt{\kappa + \eta} & \text{if } \tau \notin \text{supp } \varrho, \end{cases}$$

where $\tau = \Re z$, $\eta = \Im z$ and $\kappa := |\tau - \tau_0|$.

(iv) *There exist $P, B \in \mathbb{C}^{N \times N}$ left and right eigenvectors of \mathcal{B} such that*

$$\begin{aligned} \|\mathcal{B}^{-1}\|_{\text{sp}} &\lesssim (|\tau - \tau_0| + \eta)^{-1/2}, & \|\mathcal{B}^{-1}\mathcal{Q}\|_{\text{sp}} + \|B\| + \|P\| &\lesssim 1, \\ |\beta| &\sim \sqrt{|\tau - \tau_0| + \eta}, & |\langle P, B \rangle| &\sim 1, & |\langle P, M\mathcal{S}[B]B \rangle| &\sim 1, \end{aligned}$$

where $\mathcal{Q} := 1 - \mathcal{P}$ and $\mathcal{P} := \langle P, \cdot \rangle B / \langle P, B \rangle$ are spectral projections of \mathcal{B} .

We now design a suitable norm following [21]. For cumulants of matrix elements $\kappa(w_{ab}, w_{cd})$, we use the shorthand notation $\kappa(ab, cd)$. We also use the shorthand notation $\kappa(\mathbf{x}b, cd)$ for the $\mathbf{x} = (x_a)_{a \in [N]}$ -weighted linear combination $\sum_a x_a \kappa(ab, cd)$ of such cumulants. We use the notation that replacing an index in a scalar quantity by a dot (\cdot) refers to the corresponding vector, for example, A_{\cdot} is a shorthand notation for the vector $(A_{ab})_{b \in [N]}$. We fix two vectors \mathbf{x}, \mathbf{y} and some large integer K and define the sets

$$\begin{aligned} I_0 &:= \{\mathbf{x}, \mathbf{y}\} \cup \{e_a, P_a^* | a \in [N]\}, \\ I_{k+1} &:= I_k \cup \{M\mathbf{u} | \mathbf{u} \in I_k\} \\ &\quad \cup \{\kappa_c((M\mathbf{u})a, \cdot), \kappa_d((M\mathbf{u})a, \cdot) | \mathbf{u} \in I_k, a, b \in [N]\}, \end{aligned}$$

where $\kappa_c + \kappa_d = \kappa$ is a decomposition of κ according to the Hermitian symmetry.³ Due to (3a) such a decomposition exists in a way that the operator norms of the matrices $\|\kappa_d(\mathbf{x}a, \cdot b)\|$ and $\|\kappa_c(\mathbf{x}a, \cdot b)\|$, indexed by (a, b) , are bounded uniformly in \mathbf{x} with $\|\mathbf{x}\| \leq 1$. We now define the norm

$$\begin{aligned} \|R\|_* &:= \|R\|_*^{K, \mathbf{x}, \mathbf{y}} := \sum_{0 \leq k < K} N^{-k/2K} \|R\|_{I_k} + N^{-1/2} \max_{\mathbf{u} \in I_K} \frac{\|R_{\cdot \mathbf{u}}\|}{\|\mathbf{u}\|}, \\ \|R\|_I &:= \max_{\mathbf{u}, \mathbf{v} \in I} \frac{|R_{\mathbf{u}\mathbf{v}}|}{\|\mathbf{u}\| \|\mathbf{v}\|}. \end{aligned}$$

We note that the sets I_k and thereby also the norm $\|\cdot\|_*$ depend implicitly on the spectral parameter z via M and P .

REMARK 3.2. Compared to [21], the sets I_k contain some additional vectors generated by the vectors of the form P_a^* in I_0 . This addition is necessary to control the spectral projection \mathcal{P} in the $\|\cdot\|_*$ -norm. We note, however, that the precise form of the sets I_k were not important for the proofs in [21]. It was only used that these sets contain deterministic vectors, and that their cardinality grows at most as some finite power $|I_k| \lesssim N^{C_k}$ of N .

In terms of this norm, we obtain the following easy estimate on $G - M$ in terms of its projection Θ onto the unstable direction of the stability operator \mathcal{B} .

PROPOSITION 3.3. *For sufficiently small δ and fixed z such that $\|G - M\|_* \lesssim N^{-3/K}$, there are deterministic matrices R_1, R_2 with norm $\lesssim 1$ such that*

$$(12a) \quad G - M = \Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD] + \mathcal{E}, \quad \|\mathcal{E}\|_* \lesssim N^{2/K} (|\Theta|^2 + \|D\|_*^2),$$

with an error term \mathcal{E} , where Θ , defined in (11), satisfies the approximate quadratic equation

$$(12b) \quad \xi_1 \Theta + \xi_2 \Theta^2 = \mathcal{O}(N^{2/K} \|D\|_*^2 + |\langle R_1 D \rangle| + |\langle R_2 D \rangle|)$$

³If h_{ab} is strongly correlated with h_{cd} then, by Hermitian symmetry, it is also strongly correlated with $h_{dc} = \overline{h_{cd}}$. Therefore, it is natural to split the covariance into a *direct* and *cross* contribution. The precise splitting $\kappa = \kappa_c + \kappa_d$ is chosen via an optimization problem. The precise definition is irrelevant for the current proof; see [21], Remark 2.8, for more details.

with

$$|\xi_1| \sim \sqrt{\eta + \kappa}, \quad |\xi_2| \sim 1$$

and any implied constants are uniform in \mathbf{x}, \mathbf{y} and $z \in \mathbb{D}^\delta$.

PROOF. We begin with an auxiliary lemma about the $\|\cdot\|_*$ -norm of some important quantities, the proof of which we defer to the [Appendix](#).

LEMMA 3.4. *Depending only on the model parameters, we have the estimates for any $R \in \mathbb{C}^{N \times N}$,*

$$\begin{aligned} \|MS[R]R\|_* &\lesssim N^{1/2K} \|R\|_*^2, & \|MR\|_* &\lesssim N^{1/2K} \|R\|_*, \\ \|\mathcal{Q}\|_{* \rightarrow *} &\lesssim 1, & \|\mathcal{B}^{-1}\mathcal{Q}\|_{* \rightarrow *} &\lesssim 1. \end{aligned}$$

Decomposing $G - M = \mathcal{P}[G - M] + \mathcal{Q}[G - M]$ and inverting \mathcal{B} in (10) on the range of \mathcal{Q} yield

$$\begin{aligned} G - M &= \Theta B + \mathcal{Q}[G - M] = \Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD] + \mathcal{O}(N^{1/2K} \|G - M\|_*^2) \\ &= \Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD] + \mathcal{O}(N^{3/2K} (|\Theta|^2 + \|D\|_*^2)), \end{aligned}$$

where $\mathcal{O}(\cdot)$ is meant with respect to the $\|\cdot\|_*$ -norm and the second equality followed by iteration, Lemma 3.4 and the assumption on $\|G - M\|_*$. Going back to the original equation (10), we find

$$\begin{aligned} \beta \Theta B + \mathcal{B}\mathcal{Q}[G - M] &= -MD + MS[\Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD]](\Theta B - \mathcal{B}^{-1}\mathcal{Q}[MD]) \\ &\quad + \mathcal{O}(N^{2/K} (|\Theta|^3 + \|D\|_*^3)) \end{aligned}$$

and thus by projecting with \mathcal{P} we arrive at the quadratic equation

$$\begin{aligned} \mu_0 - \mu_1 \Theta + \mu_2 \Theta^2 &= \mathcal{O}(N^{2/K} (|\Theta|^3 + \|D\|_*^3)), \\ \mu_0 &= \langle P, MS[\mathcal{B}^{-1}\mathcal{Q}[MD]]\mathcal{B}^{-1}\mathcal{Q}[MD] - MD \rangle, \\ \mu_1 &= \langle P, MS[B]\mathcal{B}^{-1}\mathcal{Q}[MD] + MS[\mathcal{B}^{-1}\mathcal{Q}[MD]]B \rangle + \beta \langle P, B \rangle, \\ \mu_2 &= \langle P, MS[B]B \rangle. \end{aligned}$$

We now proceed by analyzing the coefficients in this quadratic equation. We estimate the quadratic term in μ_0 directly by $N^{2/K} \|D\|_*^2$, while we write the linear term as $\langle R_1 D \rangle$ for the deterministic $R_1 := -M^* P$ with $\|R_1\| \lesssim 1$. For the linear coefficient μ_1 , we similarly find a deterministic matrix R_2 such that $\|R_2\| \lesssim 1$ and $\mu_1 = \langle R_2 D \rangle + \beta \langle P, B \rangle$. Finally, we find from Proposition 3.1(iv) that $|\mu_2| \sim 1$ and $|\beta \langle P, B \rangle| \sim \sqrt{\kappa + \eta}$. By incorporating the $|\Theta|N^{2/K}$ term into ξ_2 , we obtain (12b). Here, δ has to be chosen sufficiently small such that Proposition 3.1 is applicable. \square

3.2. *Probabilistic bound.* We now collect the averaged and isotropic bound on D from [21]. We first introduce a commonly used (see, e.g., [19]) notion of high-probability bound.

DEFINITION 3.5 (Stochastic domination). If

$$X = (X^{(N)}(u) | N \in \mathbb{N}, u \in U^{(N)}) \quad \text{and} \quad Y = (Y^{(N)}(u) | N \in \mathbb{N}, u \in U^{(N)})$$

are families of nonnegative random variables indexed by N , and possibly some parameter u , then we say that X is stochastically dominated by Y , if for all $\epsilon, D > 0$ we have

$$\sup_{u \in U^{(N)}} \mathbf{P}[X^{(N)}(u) > N^\epsilon Y^{(N)}(u)] \leq N^{-D}$$

for large enough $N \geq N_0(\epsilon, D)$. In this case, we use the notation $X \prec Y$.

It can be checked (see [19], Lemma 4.4) that \prec satisfies the usual arithmetic properties, for example, if $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then also $X_1 + X_2 \prec Y_1 + Y_2$ and $X_1 X_2 \prec Y_1 Y_2$. To formulate the result compactly, we also introduce the notation

$$(13) \quad \begin{aligned} |R| \prec \Lambda \text{ in } \mathbb{D} &\iff \|R\|_*^{K,\mathbf{x},\mathbf{y}} \prec \Lambda \text{ unif. in } \mathbf{x}, \mathbf{y} \text{ and } z \in \mathbb{D}, \\ |R|_{\text{av}} \prec \Lambda \text{ in } \mathbb{D} &\iff \frac{|\langle BR \rangle|}{\|B\|} \prec \Lambda \text{ unif. in } B \text{ and } z \in \mathbb{D} \end{aligned}$$

for random matrices $R = R(z)$ and a deterministic control parameter $\Lambda = \Lambda(z)$, where $B, \mathbf{x}, \mathbf{y}$ are deterministic matrices and vectors. We also define an isotropic high-moment norm, already used in [21], for $p \geq 1$ and a random matrix R ,

$$\|R\|_p := \sup_{\mathbf{x}, \mathbf{y}} \frac{(\mathbf{E}|\langle \mathbf{x}, R\mathbf{y} \rangle|^p)^{1/p}}{\|\mathbf{x}\| \|\mathbf{y}\|}.$$

PROPOSITION 3.6 (Bound on the error). *Under the Assumptions (A)–(E), there exists a constant C such that for any fixed vectors \mathbf{x}, \mathbf{y} and matrices B and spectral parameters $z \in \mathbb{D}^\delta$, and any $p \geq 1, \epsilon > 0$,*

$$(14a) \quad \frac{\|\langle \mathbf{x}, D\mathbf{y} \rangle\|_p}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq_{\epsilon, p} N^\epsilon \sqrt{\frac{\|\Im G\|_q}{N\Im z}} (1 + \|G\|_q)^C \left(1 + \frac{\|G\|_q}{N^\mu}\right)^{Cp},$$

$$(14b) \quad \frac{\|\langle BD \rangle\|_p}{\|B\|} \leq_{\epsilon, p} N^\epsilon \frac{\|\Im G\|_q}{N\Im z} (1 + \|G\|_q)^C \left(1 + \frac{\|G\|_q}{N^\mu}\right)^{Cp},$$

where $q := Cp^4/\epsilon$. Here, $\mu > 0$ depends on s in Assumption (CD). In particular, if $|G - M| \prec \Lambda \lesssim 1$, then

$$(14c) \quad |D| \prec \sqrt{\frac{\varrho + \Lambda}{N\eta}}, \quad |D|_{\text{av}} \prec \frac{\varrho + \Lambda}{N\eta}.$$

PROOF. This follows from combining [21], Theorem 3.1, the following lemma⁴ from [21], Lemma 4.4, and $\|M\| \leq M_*$. \square

LEMMA 3.7. *Let R be a random matrix and Φ a deterministic control parameter. Then the following implications hold:*

- (i) *If $\Phi \geq N^{-C}$, $\|R\| \leq N^C$ and $|R_{\mathbf{xy}}| \prec \Phi \|\mathbf{x}\| \|\mathbf{y}\|$ for all \mathbf{x}, \mathbf{y} and some C , then $\|R\|_p \leq_{p, \epsilon} N^\epsilon \Phi$ for all $\epsilon > 0, p \geq 1$.*
- (ii) *Conversely, if $\|R\|_p \leq_{p, \epsilon} N^\epsilon \Phi$ for all $\epsilon > 0, p \geq 1$, then $\|R\|_*^{K,\mathbf{x},\mathbf{y}} \prec \Phi$ for any fixed $K \in \mathbb{N}, \mathbf{x}, \mathbf{y} \in \mathbb{C}^N$.*

3.3. Bootstrapping. We now fix $\gamma > 0$ and start with the proof of Theorem 2.6. Phrased in terms of the $\|\cdot\|_*$ -norm, we will prove

$$(15) \quad \begin{aligned} |G - M| &\prec N^{2/K} \left(\sqrt{\frac{\varrho}{N\eta}} + \frac{1}{N\eta} \right) \text{ in } \mathbb{D}, \\ |G - M|_{\text{av}} &\prec N^{2/K} \begin{cases} \frac{1}{N\eta}, & \Re z \in \text{supp } \varrho, \\ \frac{1}{N(\kappa + \eta)} + \frac{N^{2/K}}{(N\eta)^2 \sqrt{\kappa + \eta}}, & \Re z \notin \text{supp } \varrho \end{cases} \text{ in } \mathbb{D}, \end{aligned}$$

⁴Cf. Remark 3.2, where we argue that the proof of [21] about $\|\cdot\|_*$ hold true verbatim in the present case despite the slightly larger sets I_k .

for $\mathbb{D} = \mathbb{D}_\gamma^\delta$ and $K \gg 1/\gamma$, that is, for $K\gamma$ sufficiently large. In order to prove (15), we use the following iteration procedure.

PROPOSITION 3.8. *There exists a constant $\gamma_s > 0$ depending only on K and γ such that (15) for $\mathbb{D} = \mathbb{D}_{\gamma_0}^\delta$ with $\gamma_0 > \gamma$ implies (15) also for $\mathbb{D} = \mathbb{D}_{\gamma_1}^\delta$ with $\gamma_1 := \max\{\gamma, \gamma_0 - \gamma_s\}$.*

PROOF OF (15) FOR $\mathbb{D} = \mathbb{D}_\gamma^\delta$, ASSUMING PROPOSITION 3.8. For $\mathbb{D} = \mathbb{D}_\gamma^\delta$ with $\gamma \geq 1$, we have (15) by [21], Theorem 2.1. For $\gamma < 1$, we iteratively apply Proposition 3.8 starting from \mathbb{D}_1^δ finitely many times until we have shown (15) for $\mathbb{D} = \mathbb{D}_\gamma^\delta$. \square

PROOF OF PROPOSITION 3.8. We now suppose that (15) has been proven for some $\mathbb{D} = \mathbb{D}_{\gamma_0}^\delta$ and aim at proving (15) for $\mathbb{D} = \mathbb{D}_{\gamma_1}^\delta$ for some $\gamma_1 = \gamma_0 - \gamma_s$, $0 < \gamma_s \ll \gamma$. The proof has two stages. First, we will establish the rough bounds

$$(16) \quad |\Theta| \prec N^{-5/K} \quad \text{and} \quad |G - M| \prec N^{-5/K} \quad \text{in } \mathbb{D}_{\gamma_1},$$

and then in the second stage improve upon this bound iteratively until we reach (15) for $\mathbb{D} = \mathbb{D}_{\gamma_1}^\delta$. \square

Rough bound. By (15), Lemma 3.7 and monotonicity of the map $\eta \mapsto \eta \|G(\tau + i\eta)\|_p$ (see, e.g., (77) in [21]), we find $\|G\|_p \leq_{\epsilon,p} N^{\epsilon+\gamma_s} \leq N^{2\gamma_s}$ in $\mathbb{D}_{\gamma_1}^\delta$. As long as $2\gamma_s < \mu$, we thus have

$$\begin{aligned} \|D\|_p \leq_{\epsilon,p} \frac{N^{\epsilon+2C\gamma_s+\gamma_s}}{\sqrt{N\eta}} &\leq \frac{N^{\gamma_s(2+2C)}}{\sqrt{N\eta}}, \\ \|(BD)\|_p \leq_{\epsilon,p} \|B\| \frac{N^{\epsilon+2\gamma_s+2\gamma_s C}}{N\eta} &\leq \|B\| \frac{N^{\gamma_s(3+2C)}}{N\eta}. \end{aligned}$$

We now fix \mathbf{x}, \mathbf{y} and it follows from (12b) that

$$|\xi_1\Theta + \xi_2\Theta^2| \prec \frac{N^{2\gamma_s(3+2C)+2/K}}{N\eta} \quad \text{in } \mathbb{D}_{\gamma_1}$$

and consequently by Lipschitz continuity of the left-hand side with a Lipschitz constant of $\eta^{-2} \leq N^2$, and choosing K, γ_s large, and respectively small enough depending on γ , we find that with high probability $|\xi_1\Theta + \xi_2\Theta^2| \leq N^{-10/K}$ in all of $\mathbb{D}_{\gamma_1}^\delta$. The following lemma translates the bound on $|\xi_1\Theta + \xi_2\Theta^2|$ into a bound on $|\Theta|$.

LEMMA 3.9. *Let $d = d(\eta)$ be a monotonically decreasing function in $\eta \geq 1/N$ and assume $0 \leq d \lesssim N^{-\epsilon}$ for some $\epsilon > 0$. Furthermore, suppose that*

$$|\xi_1\Theta + \xi_2\Theta^2| \lesssim d \quad \text{for all } z \in \mathbb{D}^\delta, \quad \text{and} \quad |\Theta| \lesssim \min\left\{\frac{d}{\sqrt{\kappa+\eta}}, \sqrt{d}\right\}$$

for some z_0 . Then it holds that also $|\Theta| \lesssim \min\{d/\sqrt{\kappa+\eta}, \sqrt{d}\}$ for all $z' \in \mathbb{D}^\delta$ with $\Re z' = \Re z_0$ and $\Im z' < \Im z_0$.

PROOF. This proof is basically identical to the analysis of the solutions to the same approximate quadratic equation, as appeared in various previous works; see, for example, [27],

⁵Strictly speaking, in the very first step we start from $\mathbb{D}^\delta \cap \{\Im z \geq \delta/2\}$ instead of \mathbb{D}_1^δ since, depending on the value of δ , the latter might be empty.

Section 9. In the spectral bulk, this is trivial since then $|\xi_1| \sim \sqrt{\kappa + \eta} \sim 1$. Near a spectral edge, we observe that $(\kappa + \eta)/d$ is monotonically increasing in η . First, suppose that $(\kappa + \eta)/d \gg 1$ from which it follows that $|\Theta| \lesssim d/\sqrt{\kappa + \eta} \lesssim \sqrt{d}$ in the relevant branch determined by the given estimate on Θ at z_0 . Now suppose that below some η -threshold we have $(\kappa + \eta)/d \lesssim 1$. Then we find $|\Theta| \lesssim \sqrt{\kappa + \eta} + \sqrt{d} \lesssim \sqrt{d} \lesssim d/\sqrt{\kappa + \eta}$ and the claim follows also in this regime. \square

Since (16) holds in $\mathbb{D}_{\gamma_0}^\delta$ and $1/N\eta \leq N^{-100/K}$, we know

$$|\Theta| \leq \min\{N^{-10/K}/\sqrt{\kappa + \eta}, N^{-5/K}\}$$

and, therefore, can conclude the rough bound $|\Theta| \prec N^{-5/K}$ in all of $\mathbb{D}_{\gamma_1}^\delta$ by Lemma 3.9 with $d = N^{-10/K}$. Consequently, we have also that

$$\|G - M\|_* \mathbf{1}(\|G - M\|_* \leq N^{-3/K}) \prec N^{-5/K} \quad \text{in } \mathbb{D}_{\gamma_1}.$$

Due to this gap in the possible values for $\|G - M\|_*$, it follows from a standard continuity argument that $\|G - M\|_* \prec N^{-5/K}$ and, therefore, since \mathbf{x}, \mathbf{y} were arbitrary, $|\Theta| \prec N^{-5/K}$ and $|G - M| \prec N^{-5/K}$ in all of $\mathbb{D}_{\gamma_1}^\delta$.

Strong bound. All of the following bounds hold uniformly in the domain $\mathbb{D}_{\gamma_1}^\delta$ which is why we suppress this qualifier. By combining Propositions 3.3 and 3.6, we find for deterministic $0 \leq \theta \leq \Lambda \leq N^{-3/K}$ under the assumptions $|\Theta| \prec \theta, |G - M| \prec \Lambda$, that

$$(17) \quad |G - M| \prec \theta + N^{2/K} \sqrt{\frac{\varrho + \Lambda}{N\eta}}, \quad |\xi_1 \Theta + \xi_2 \Theta^2| \prec N^{2/K} \frac{\varrho + \Lambda}{N\eta}.$$

The bound on $|G - M|$ in (17) is a self-improving bound and we find after iteration that

$$|G - M| \prec \theta + N^{2/K} \left(\frac{1}{N\eta} + \sqrt{\frac{\varrho + \theta}{N\eta}} \right),$$

hence

$$|\xi_1 \Theta + \xi_2 \Theta^2| \prec N^{2/K} \frac{\varrho + \theta}{N\eta} + N^{4/K} \frac{1}{(N\eta)^2}.$$

We now distinguish whether $\Re z$ is inside or outside the spectrum. Inside, we have $\varrho \sim \sqrt{\kappa + \eta}$, so we fix θ and use Lemma 3.9 with $d = N^{2/K}(\sqrt{\kappa + \eta} + \theta)/(N\eta) + N^{4/K}/(N\eta)^2$ to conclude $|\Theta| \prec \min\{d/\sqrt{\kappa + \eta}, \sqrt{d}\}$ from the input assumption $|\Theta| \prec N^{2/K}/N\eta$ in \mathbb{D}_{γ_0} . Iterating this bound, we obtain

$$|\Theta| \prec N^{2/K} \frac{1}{N\eta}, \quad \text{hence } |G - M| \prec N^{2/K} \left(\sqrt{\frac{\varrho}{N\eta}} + \frac{1}{N\eta} \right).$$

By an analogous argument, outside of the spectrum we have an improved bound on Θ

$$|\Theta| \prec N^{2/K} \frac{1}{N(\kappa + \eta)} + N^{4/K} \frac{1}{(N\eta)^2 \sqrt{\kappa + \eta}},$$

because $\varrho \sim \eta/\sqrt{\kappa + \eta}$. Finally, for the claimed bound on $|G - M|_{\text{av}}$ we use (12a) in order to obtain a bound on $|G - M|_{\text{av}}$ in terms of a bound on Θ .

Due to (15), we now have all the ingredients to prove the local law, as well as delocalization of eigenvectors, and the absence of eigenvalues away from the support of ϱ .

PROOF OF THEOREM 2.6, COROLLARY 2.7 AND COROLLARY 2.8. The local law inside the spectrum (6a)–(6b) follows immediately from (15). Now we prove Corollary 2.7. If there

exists an eigenvalue λ with $\text{dist}(\lambda, \text{supp } \varrho) > N^{-2/3+\omega}$, then at, say, $z = \lambda + iN^{-4/5}$, we have $|\langle G - M \rangle| \geq cN^{-1/5}$. On the other hand, we know from the improved local law (15) that with high probability $|\langle G - M \rangle| \leq N^{-1/4}$ and we obtain the claim.

We now turn to the proof of Corollary 2.8. For the eigenvectors \mathbf{u}_k and eigenvalues λ_k of H , we find from the spectral decomposition and the local law with high probability

$$1 \gtrsim \Im \langle \mathbf{x}, G\mathbf{x} \rangle = \eta \sum_k \frac{|\langle \mathbf{x}, \mathbf{u}_k \rangle|^2}{(\tau - \lambda_k)^2 + \eta^2} \geq \frac{|\langle \mathbf{x}, \mathbf{u}_k \rangle|^2}{\eta} \quad \text{for } z = \tau + i\eta$$

for any normalized $\mathbf{x} \in \mathbb{C}^N$, where the last inequality followed assuming that τ is chosen η -close to λ_k . With the choice $\eta = N^{-1+\gamma}$ for arbitrarily small $\gamma > 0$, the claim follows. Note that for this proof only (6a) of Theorem 2.6 was used.

Finally, we establish (6c) and consider $z \in \mathbb{D}^\delta$ with $\text{dist}(\Re z, \text{supp } \varrho) \geq N^{-2/3+\omega}$ and $\mathbf{x}, \mathbf{y}, B$ fixed. As in the proof of [3], Corollary 1.11, the optimal local law (15) implies rigidity up to the edge as formulated in Corollary 2.9. The only difference is that this standard argument proves (7b) only if the supremum is restricted to $\tau \in \text{supp } \varrho$ with $\text{dist}(\tau, \partial \text{supp } \varrho) \geq N^{-2/3+\epsilon}$. The cause for this restriction is a possible mismatch of the labeling of the edge eigenvalues; in other words, the precise location of N^ϵ eigenvalues near an internal gap is not established yet; they may belong to either band adjacent to this gap. This shortcoming will be remedied by the band rigidity in the proof of Corollary 2.9 in Section 5 below. However, for the current argument, the imprecise location of N^ϵ eigenvalues does not matter. In fact, already from this version of rigidity, together with the delocalization of eigenvectors (Corollary 2.8) and the absence of eigenvalues outside of the spectrum by Corollary 2.7 we have, at $z = \tau + i\eta$ (recall that we consider $z \in \mathbb{D}^\delta$ with $\text{dist}(\Re z, \text{supp } \varrho) \geq N^{-2/3+\omega}$),

$$\begin{aligned} \Im \langle \mathbf{x}, G(z)\mathbf{x} \rangle &= \eta \sum_k \frac{|\langle \mathbf{x}, \mathbf{u}_k \rangle|^2}{(\tau - \lambda_k)^2 + \eta^2} < \frac{1}{N} \sum_k \frac{\eta}{(\tau - \lambda_k)^2 + \eta^2} \\ &< \int_{\mathbb{R}} \frac{\eta \varrho(x) dx}{|\tau - x|^2 + \eta^2} \end{aligned}$$

for any normalized vector \mathbf{x} . From the square root behavior of ϱ at the edge and $\kappa(z) \geq N^{-2/3+\omega}$, we can easily infer $\|\Im G\|_* < \eta/\sqrt{\kappa + \eta}$. Therefore, it follows from Proposition 3.6 that $\|D\|_*^2 + |\langle RD \rangle| < 1/(N\sqrt{\kappa + \eta})$ and from (12b) and Lemma 3.9 that $|\Theta| < N^{2/K-1}/(\kappa + \eta)$. Finally, we thus obtain

$$|G - M|_{\text{av}} < \frac{N^{2/K}}{N(\kappa + \eta)} + \frac{N^{2/K}}{N\sqrt{\kappa + \eta}} \lesssim N^{2/K} \frac{1}{N(\kappa + \eta)}$$

from (12a) and (6c) follows. \square

4. Analysis of the matrix Dyson equation. The essential prerequisite for edge universality is the regularity of the edge, that is, the local square root behavior of the self-consistent density ϱ as imposed in Definition 2.4. For the proof of universality via [35], however, it is necessary to first establish that the square-root behavior and the adjacent gap persist in a macroscopic interval. This is achieved in the following main theorem whose proof will be given in Section 4.4 after several preparatory results. In particular, as a second main result of this section, in Theorem 4.2, we will give a sharp estimate on the inverse of the stability operator $\mathcal{B} = \text{Id} - MS[\cdot]M$ which also plays a central role in the proof of the local law in Section 3.

THEOREM 4.1 (Behavior of ϱ close to a square root edge). *Let (A), (E) and (G) be satisfied for some $\tau_0 \in \mathbb{R}$. If $\tau_0 \in \partial \text{supp } \varrho$ is a regular edge, then there are $c \sim 1$ and $\delta_* \sim 1$ such that*

$$\varrho(\tau_0 + \omega) = \begin{cases} c|\omega|^{1/2} + \mathcal{O}(|\omega|) & \text{if } \omega \in [-\delta_*, 0], \\ 0 & \text{if } \omega \in [0, \delta_*]. \end{cases}$$

In this section and, in particular, the previous theorem, the comparison relation \sim is understood with respect to the constants in (A), (E) and (G) as well as in (5).

We now outline the strategy for the proof of Theorem 4.1. First, we will extend M to the real line by showing that it is $1/2$ -Hölder continuous in the vicinity of τ_0 (see Corollary 4.3 below). The Hölder continuity also yields an a priori bound on $\Delta := M(\tau_0 + \omega) - M(\tau_0)$, hence on $\varrho(\tau_0 + \omega) = \pi^{-1} \langle \Im M(\tau_0 + \omega) \rangle = \pi^{-1} \langle \Im \Delta \rangle$ as well, with small $\omega \in \mathbb{R}$. Second, by using this bound, we will verify that Δ is governed by a scalar quantity analogous to Θ from (11) which satisfies a quadratic equation (see Proposition 4.13 below). The fact that $\Im \Delta \geq 0$ will select the correct solution to this quadratic equation and Theorem 4.1 will follow from analyzing the stability of this solution.

The equation for Δ can be obtained from subtracting the MDE at $\tau_0 + \omega$ and τ_0 . It reads as

$$(18) \quad \mathcal{B}[\Delta] = M\mathcal{S}[\Delta]\Delta + \omega M^2 + \omega M\Delta, \quad M = M(\tau_0).$$

To express Δ from (18), it is therefore essential to understand the instabilities of \mathcal{B}^{-1} very precisely. The main difficulty is that near the edge \mathcal{B} has a small eigenvalue that is very sensitive to a delicate balance between \mathcal{S} and M . An additional complication is that \mathcal{B} is nonself-adjoint. Both obstacles are overcome by representing \mathcal{B} in the form $\mathcal{B} = \mathcal{V}(\mathcal{U} - \mathcal{F})\mathcal{V}^{-1}$, where \mathcal{U} is unitary, \mathcal{V} is bounded invertible, \mathcal{F} is self-adjoint and it preserves the cone of positive matrices. Thus a Perron–Frobenius argument can be applied to \mathcal{F} , that is, its norm can be obtained simply by finding its top eigenvector. In this way, we can very precisely determine the size of $M\mathcal{S}[\cdot]M$ and estimate its top eigenvalue without explicitly solving the MDE. This representation of \mathcal{B} (cf. (26) below) with the Perron–Frobenius argument is one of the main results of [4] and the analysis of \mathcal{F} will partly be imported from [4]. We will see that \mathcal{B}^{-1} has precisely one unstable direction and we will obtain the quadratic equation for Θ , the projection of Δ , onto this direction. The sharp estimate on the eigenvalue of the unstable direction will give rise to the following bound on \mathcal{B}^{-1} .

THEOREM 4.2 (Sharp bound on \mathcal{B}^{-1} near a regular edge). *Let (A), (E) and (G) be satisfied for a regular edge $\tau_0 \in \partial \text{supp } \varrho$. Then there is $\delta_* \sim 1$ such that we have*

$$\|\mathcal{B}(z)^{-1}\|_{\text{sp}} + \|\mathcal{B}(z)^{-1}\| \lesssim \frac{1}{\varrho(z) + \eta\varrho(z)^{-1}},$$

for all $z \in \mathbb{H}$ satisfying $|z - \tau_0| \leq \delta_*$, where $\eta = \Im z$.

From the previous theorem, we will immediately conclude the $1/2$ -Hölder continuity stated in the following corollary. The proofs of both statements will be given in Section 4.2 below.

COROLLARY 4.3 (Hölder continuity of M). *Let (A), (E) and (G) be satisfied for a regular edge $\tau_0 \in \mathbb{R}$. Then M is uniformly $1/2$ -Hölder continuous around τ_0 in the sense that there is $\delta_* \sim 1$ such that*

$$\|M(z_1) - M(z_2)\| \lesssim |z_1 - z_2|^{1/2}$$

for all $z_1, z_2 \in \{\tau + i\eta : |\tau - \tau_0| \leq \delta_*, 0 < \eta < \infty\}$. In particular, M has a unique extension to $[\tau_0 - \delta_*, \tau_0 + \delta_*]$.

4.1. *Analysis of the stability operator.* In this section, we will always assume that (A), (E) and (G) are satisfied for some $\tau_0 \in \mathbb{R}$. The main result of this section is the bound on the inverse of the stability operator \mathcal{B} in Proposition 4.4 below. We introduce the *balanced polar decomposition*

$$(19) \quad M = Q^* U Q,$$

where we define

$$(20) \quad \begin{aligned} W &:= (\Im M)^{-1/2} (\Re M) (\Im M)^{-1/2} + i, \\ Q &:= |W|^{1/2} (\Im M)^{1/2}, \quad U := \frac{W}{|W|}. \end{aligned}$$

We remark that W is normal, $|W| := (W^* W)^{1/2}$, U is unitary and $\Im U$ is positive definite. In this context, the balanced polar decomposition first appeared in [4]. We also define

$$(21) \quad S := \text{sign } \Re U, \quad F_U := \varrho^{-1} \Im U, \quad \sigma := \langle S F_U^3 \rangle.$$

The quantities \mathcal{B} , W , Q , U , S , F_U and σ introduced above all depend on z through the z -dependence of M . In the following, we will mostly omit this dependence from our notation.

PROPOSITION 4.4 (General bound on \mathcal{B}^{-1}). *If (A), (E) and (G) are satisfied for some $\tau_0 \in \mathbb{R}$ then, uniformly for all $z \in \mathbb{D}^{\omega_*}$, we have*

$$(22) \quad \|\mathcal{B}(z)^{-1}\|_{\text{sp}} + \|\mathcal{B}(z)^{-1}\| \lesssim \frac{1}{\varrho(z)(\varrho(z) + |\sigma(z)|) + \eta \varrho(z)^{-1}}, \quad \eta = \Im z.$$

This proposition will be shown at the end of the present section. Now, we apply it to show that M is 1/3-Hölder continuous.

COROLLARY 4.5 (1/3-Hölder continuity of M). *Let (A), (E) and (G) be satisfied for some $\tau_0 \in \mathbb{R}$. Then the solution M of the MDE, (1), is uniformly 1/3-Hölder continuous around τ_0 in the sense that, for each $\theta \in (0, \omega_*)$, we have*

$$\|M(z_1) - M(z_2)\| \lesssim_\theta |z_1 - z_2|^{1/3}$$

for all $z_1, z_2 \in \{\tau + i\eta : |\tau - \tau_0| \leq \omega_* - \theta, 0 < \eta < \infty\}$. In particular, $\varrho(z) = \pi^{-1} \langle \Im M(z) \rangle$ is 1/3-Hölder continuous on the same domain.

Before we prove the previous corollary, we state and prove the following lemma. It collects a few basic properties of M , Q and U which will often be used in the following.

LEMMA 4.6 (Properties of M , Q and U). *Uniformly for $z \in \mathbb{D}^{\omega_*}$, we have*

$$(23a) \quad \|M(z)^{-1}\| \sim \|M(z)\| \sim 1,$$

$$(23b) \quad \Im M(z) \sim \langle \Im M(z) \rangle,$$

$$(23c) \quad \|Q(z)\| \sim \|Q(z)^{-1}\| \sim 1,$$

$$(23d) \quad \Im U(z) \sim \langle \Im U(z) \rangle \sim \varrho(z),$$

where $A \lesssim B$ and $A \sim B$ for matrices A, B indicate that $A \leq CB$ and $cB \leq A \leq CB$ for some constants c, C in the sense of quadratic forms.

PROOF OF LEMMA 4.6. The bounds in (23a) and (23b) follow easily from the bound on $\|M\|$ on \mathbb{D}^{ω_*} as well as the flatness of \mathcal{S} (see, e.g., the proof of Proposition 4.2 in [4]).

For the proof of (23c), we use the monotonicity of the square root and (23b) to obtain

$$\begin{aligned} Q^*Q &= (\Im M)^{1/2}(1 + (\Im M)^{-1/2}(\Re M)(\Im M)^{-1}(\Re M)(\Im M)^{-1/2})^{1/2}(\Im M)^{1/2} \\ &\sim \langle \Im M \rangle^{-1/2}(\Im M)^{1/2}((\Im M)^{-1/2}((\Im M)^2 + (\Re M)^2)(\Im M)^{-1/2})^{1/2}(\Im M)^{1/2}. \end{aligned}$$

Thus, employing $(\Re M)^2 + (\Im M)^2 \sim 1$ by (23a) yields (23c) due to (23b).

Owing to (23c), (23d) is a direct consequence of (23b). This completes the proof of Lemma 4.6. \square

In the following, we will use the derivative of M with respect to z several times. For $z \in \mathbb{H}$, we take the derivative of (1) with respect to z . Owing to the invertibility of $\mathcal{B} = \mathcal{B}(z)$, this yields

$$(24) \quad \partial_z M(z) = \mathcal{B}^{-1}[M(z)^2]$$

for $z \in \mathbb{D}^{\omega_*}$.

PROOF OF COROLLARY 4.5. As $\partial_z \Im M(z) = (2i)^{-1}\partial_z M(z)$ due to the analyticity of M , we conclude from (24) and (22) and (23b) that

$$\|\partial_z \Im M(z)\| \lesssim \varrho(z)^{-2} \sim \|\Im M(z)\|^{-2}.$$

This implies that $z \mapsto (\Im M(z))^3$ is Lipschitz-continuous on \mathbb{D}^{ω_*} . Therefore, $\Im M(z)$ is 1/3-Hölder continuous on \mathbb{D}^{ω_*} (see, e.g., Theorem X.1.1 in [10]), and thus M is uniformly 1/3-Hölder continuous on $\{\tau + i\eta : |\tau - \tau_0| \leq \omega_* - \theta, 0 < \eta < \infty\}$ for all $\theta \in (0, \omega_*)$ (see, e.g., Lemma A.7 in [2] as well as Lemma A.1 in [5] for a slightly more general formulation). \square

For the analysis of the stability operator \mathcal{B} defined in (8), we now introduce the Hermitian operator $\mathcal{F} : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$ defined through

$$(25) \quad \mathcal{F} := \mathcal{C}_{Q,Q^*} \mathcal{S} \mathcal{C}_{Q^*,Q}.$$

Here, we used the following notation for operators on $\mathbb{C}^{N \times N}$. For $T_1, T_2 \in \mathbb{C}^{N \times N}$, we define the operator $\mathcal{C}_{T_1, T_2} : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$ through

$$\mathcal{C}_{T_1, T_2}[R] = T_1 R T_2$$

for all $R \in \mathbb{C}^{N \times N}$. We also set $\mathcal{C}_T := \mathcal{C}_{T, T}$. We remark that $\mathcal{C}_{T_1, T_2}^* = \mathcal{C}_{T_1^*, T_2^*}$ and, if T_1 and T_2 are invertible then $\mathcal{C}_{T_1, T_2}^{-1} = \mathcal{C}_{T_1^{-1}, T_2^{-1}}$. The importance of \mathcal{F} for the analysis of \mathcal{B} and its inverse comes from the following consequence of the balanced polar decomposition (19):

$$(26) \quad \mathcal{B} = \text{Id} - \mathcal{C}_M \mathcal{S} = \mathcal{C}_{Q^*, Q} \mathcal{C}_U (\mathcal{C}_U^* - \mathcal{F}) \mathcal{C}_{Q^*, Q}^{-1}.$$

When $\varrho = \varrho(z)$ is small, we will view \mathcal{B} as a perturbation of the operator \mathcal{B}_0 , which we introduce now. We define

$$(27) \quad \begin{aligned} \mathcal{B}_0 &:= \mathcal{C}_{Q^*, Q} (\text{Id} - \mathcal{C}_S \mathcal{F}) \mathcal{C}_{Q^*, Q}^{-1}, \\ \mathcal{E} &:= (\mathcal{C}_{Q^* S Q} - \mathcal{C}_M) \mathcal{S} = \mathcal{C}_{Q^*, Q} (\mathcal{C}_S - \mathcal{C}_U) \mathcal{F} \mathcal{C}_{Q^*, Q}^{-1}, \end{aligned}$$

with U and Q defined in (20), S defined in (21) and \mathcal{F} defined in (25). Note $\mathcal{B}_0 = \text{Id} - \mathcal{C}_{Q^* S Q}$, that is, in the definition of \mathcal{B} , the unitary matrix U in $M = Q^* U Q$ is replaced by S . Thus we have $\mathcal{B} = \mathcal{B}_0 + \mathcal{E}$.

In the following, we will often use (23c) and (23d). In particular, since $I - |\Re U| = I - \sqrt{I - (\Im U)^2} \leq (\Im U)^2 \lesssim \varrho^2$, we also obtain

$$(28) \quad \Re U = S + \mathcal{O}(\varrho^2), \quad \Im U = \mathcal{O}(\varrho), \quad \Re M = Q^* S Q + \mathcal{O}(\varrho^2).$$

Here, the notation $R = \mathcal{O}(\alpha)$ for matrices R and scalars α means that $\|R\| \lesssim \alpha$. Owing to $\mathcal{C}_S - \mathcal{C}_U = \mathcal{O}(\|S - U\|) = \mathcal{O}(\varrho)$, we get

$$(29) \quad \mathcal{E} = \mathcal{O}(\varrho).$$

Here, we use the notation $\mathcal{R} = \mathcal{T} + \mathcal{O}(\alpha)$ for operators \mathcal{R} and \mathcal{T} on $\mathbb{C}^{N \times N}$ and $\alpha > 0$ if $\|\mathcal{R} - \mathcal{T}\| \lesssim \alpha$. By the functional calculus, the normal matrices $U, \Re U, S$ and F_U commute. Hence, $\mathcal{C}_S[F_U] = F_U$.

The MDE, (1), the balanced polar decomposition, $M = Q^* U Q$, and the definition of \mathcal{F} in (25) yield

$$(30) \quad -U^* = Q(z - A)Q^* + \mathcal{F}[U].$$

We take the imaginary part of (30) and use (23c) as well as (23d) to conclude that

$$(31) \quad (\text{Id} - \mathcal{F})[F_U] = \eta \varrho^{-1} Q Q^* = \mathcal{O}(\eta \varrho^{-1}).$$

We also introduce the operator \mathcal{B}_* , and view it as a perturbation of \mathcal{B}_0 , via

$$\begin{aligned} \mathcal{B}_* &:= \text{Id} - \mathcal{C}_{M^*, M} S, \\ \mathcal{E}_* &:= (\mathcal{C}_{Q^* S Q} - \mathcal{C}_{M^*, M}) S = \mathcal{C}_{Q^*, Q} (\mathcal{C}_S - \mathcal{C}_{U^*, U}) \mathcal{F} \mathcal{C}_{Q^*, Q}^{-1}. \end{aligned}$$

Hence, we have $\mathcal{B}_* = \mathcal{B}_0 + \mathcal{E}_*$. Analogously to (29), we conclude from (28) that

$$(32) \quad \mathcal{E}_* = \mathcal{O}(\varrho).$$

In the following, 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 ε around z .

LEMMA 4.7 (Spectral properties of stability operator for small density). *Let $\mathcal{T} \in \{\text{Id} - \mathcal{F}, \text{Id} - \mathcal{C}_S \mathcal{F}, \mathcal{B}_0, \mathcal{B}, \mathcal{B}_*\}$. Then there are $\varrho_* \sim 1$ and $\varepsilon \sim 1$ such that*

$$(33) \quad \|(\mathcal{T} - \omega \text{Id})^{-1}\|_{\text{sp}} + \|(\mathcal{T} - \omega \text{Id})^{-1}\| + \|(\mathcal{T}^* - \omega \text{Id})^{-1}\| \lesssim 1$$

uniformly for all $z \in \mathbb{D}^{\omega_}$ satisfying $\varrho(z) + \eta \varrho(z)^{-1} \leq \varrho_*$ and for all $\omega \in \mathbb{C}$ with $\omega \notin D_\varepsilon(0) \cup D_{1-2\varepsilon}(1)$. Furthermore, there is a single simple (algebraic multiplicity 1) eigenvalue λ in the disk around 0, that is,*

$$(34) \quad \text{Spec}(\mathcal{T}) \cap D_\varepsilon(0) = \{\lambda\} \quad \text{and} \quad \text{rank } \mathcal{P}_\mathcal{T} = 1,$$

where

$$\mathcal{P}_\mathcal{T} := -\frac{1}{2\pi i} \int_{\partial D_\varepsilon(0)} (\mathcal{T} - \omega \text{Id})^{-1} d\omega.$$

PROOF. First, we introduce the bounded operators $\mathcal{V}_t : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$ for $t \in [0, 1]$ interpolating between Id and \mathcal{C}_S by

$$\mathcal{V}_t := (1 - t)\text{Id} + t\mathcal{C}_S.$$

We will perform the proof one by one for the choices $\mathcal{T} = \text{Id} - \mathcal{F}, \text{Id} - \mathcal{V}_t \mathcal{F}, \mathcal{B}_0, \mathcal{B}, \mathcal{B}_*$ in that order. We will first show that the operator $\text{Id} - \mathcal{F}$ has a spectral gap above the single eigenvalue around 0, so for this choice the statements are easy. Then we perform two approximations. First, we interpolate between $\text{Id} - \mathcal{F}$ and $\text{Id} - \mathcal{C}_S \mathcal{F}$ via $\text{Id} - \mathcal{V}_t \mathcal{F}$. This gives Lemma 4.7

for $\mathcal{T} = \mathcal{B}_0$. Then we use perturbation theory to get the results for $\mathcal{T} = \mathcal{B} = \mathcal{B}_0 + \mathcal{O}(\varrho)$ and for $\mathcal{T} = \mathcal{B}_* = \mathcal{B}_0 + \mathcal{O}(\varrho)$. Note that for all these choices of \mathcal{T} the bound $\|\text{Id} - \mathcal{T}\|_{\text{hs} \rightarrow \|\cdot\|} \lesssim 1$ holds due to $\|\mathcal{S}\|_{\text{hs} \rightarrow \|\cdot\|} \lesssim 1$, $\|M\| \lesssim 1$ and (23c). Hence, the invertibility of $\mathcal{T} - \omega\text{Id}$ as an operator on $(\mathbb{C}^{N \times N}, \|\cdot\|)$ and on $(\mathbb{C}^{N \times N}, \|\cdot\|_{\text{hs}})$ are therefore closely related as

$$\|(\mathcal{T} - \omega\text{Id})^{-1}\| \leq |1 - \omega|^{-1} (1 + \|\text{Id} - \mathcal{T}\|_{\text{hs} \rightarrow \|\cdot\|}) \|(\mathcal{T} - \omega\text{Id})^{-1}\|_{\text{sp}}.$$

The proof of this bound is elementary; see, for example, Lemma B.2(ii) of [5]. In particular, it suffices to show (34) and the $\|\cdot\|_{\text{sp}}$ -norm bound

$$(35) \quad \|(\mathcal{T} - \omega\text{Id})^{-1}\|_{\text{sp}} \lesssim 1,$$

for $\omega \notin D_\varepsilon(0) \cup D_{1-2\varepsilon}(1)$ in (33) to establish the lemma. For $\mathcal{T} = \text{Id} - \mathcal{F}$, both of these assertions are true due to the following facts about the operator \mathcal{F} that have been the backbone of the analysis of [4]:

(a) The norm $\|\mathcal{F}\|_{\text{sp}}$ of the Hermitian operator $\mathcal{F}: \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$ is a simple eigenvalue of \mathcal{F} . Moreover, there is a unique, positive definite eigenvector $F \in \mathbb{C}^{N \times N}$ such that $\mathcal{F}[F] = \|\mathcal{F}\|_{\text{sp}} F$ and $\|F\|_{\text{hs}} = 1$. This eigenvector satisfies

$$(36) \quad 1 - \|\mathcal{F}\|_{\text{sp}} = (\Im z) \frac{\langle F, \mathcal{Q}\mathcal{Q}^* \rangle}{\langle F, \Im U \rangle}.$$

In particular, $\|\mathcal{F}\|_{\text{sp}} \leq 1$.

Furthermore, uniformly for all $z \in \mathbb{D}^{\omega*}$, the following properties hold true:

(b) The eigenvector F is bounded from above and below, that is,

$$(37) \quad F \sim 1.$$

(c) The operator \mathcal{F} has a spectral gap $\vartheta \sim 1$, that is,

$$(38) \quad \text{Spec}(\mathcal{F} / \|\mathcal{F}\|_{\text{sp}}) \subset [-1 + \vartheta, 1 - \vartheta] \cup \{1\}.$$

(d) The eigenvector F , $\mathcal{F}F = \|\mathcal{F}\|_{\text{sp}} F$, satisfies

$$(39) \quad F = \|F_U\|_{\text{hs}}^{-1} F_U + \mathcal{O}(\eta\varrho^{-1}).$$

These facts are proven as Lemma 4.7 in [4] using Lemma 4.6 instead of (4.11) and (4.23) in the proof of (4.33) in [4]. The identity (36) follows directly from [4], equation (4.24), since $\mathcal{Q}\mathcal{Q}^* = \mathcal{C}_W[\Im \mathbf{M}]$ and $\Im U = \mathbf{W}^{-2}$, where boldface letters indicate the notation from [4], more precisely \mathbf{W} was defined in [4], equation (4.21b), and $M = \mathbf{M}$. Moreover, the proof of (39) follows from (31) and $\|\mathcal{F}\|_{\text{sp}} = 1 + \mathcal{O}(\eta\varrho^{-1})$ (cf. (36)) by straightforward perturbation theory of the simple isolated eigenvalue $\|\mathcal{F}\|_{\text{sp}}$.

Now we consider the choice $\mathcal{T} = \mathcal{T}_t = \text{Id} - \mathcal{V}_t \mathcal{F}$. Once (35), and with it (33), is established for \mathcal{T}_t , the statement about the single isolated eigenvalue (34) follows. Indeed, assuming (33) for $\mathcal{T} = \mathcal{T}_t$, we obtain that \mathcal{T}_t , and hence, the rank of $\mathcal{P}_{\mathcal{T}_t}$ is a continuous function of t on $[0, 1]$. Hence, the rank of $\mathcal{P}_{\mathcal{T}_t}$ is constant along this interpolation. On the other hand, $\text{rank } \mathcal{P}_{\mathcal{T}_0} = 1$ by Fact (a) above. Therefore, for each $t \in [0, 1]$, $\text{Spec}(\mathcal{T}_t) \cap D_\varepsilon(0)$ consists of precisely one simple eigenvalue. We are thus left with establishing (35) for \mathcal{T}_t . As $\|\mathcal{V}_t\|_{\text{sp}} \leq 1$ and $\|\mathcal{F}\|_{\text{sp}} \leq 1$ the bound (35) is certainly satisfied for $|\omega| \geq 3$. Thus we now assume $|\omega| \leq 3$. In order to conclude (35), we now show a lower bound on $\|((1 - \omega)\text{Id} - \mathcal{V}_t \mathcal{F})[R]\|_{\text{hs}}$ for all normalized, $\|R\|_{\text{hs}} = 1$, elements $R \in \mathbb{C}^{N \times N}$. We decompose R as $R = \alpha F + R^\perp$, where $R^\perp \perp F$ with respect to the Hilbert–Schmidt scalar product on $\mathbb{C}^{N \times N}$ and $\alpha \in \mathbb{C}$. Then

$$(40) \quad \begin{aligned} \|((1 - \omega)\text{Id} - \mathcal{V}_t \mathcal{F})[R]\|_{\text{hs}}^2 &= |\alpha|^2 |\omega|^2 + \mathcal{O}(\eta\varrho^{-1}) \\ &+ \|((1 - \omega)\text{Id} - \mathcal{V}_t \mathcal{F})[R^\perp]\|_{\text{hs}}^2, \end{aligned}$$

because of $\|\mathcal{F}\|_{\text{sp}} = 1 + \mathcal{O}(\eta\varrho^{-1})$, $\mathcal{V}_t[F_U] = F_U$ together with (39), and because the mixed terms are negligible due to

$$\langle F, \mathcal{V}_t \mathcal{F}[R^\perp] \rangle = \langle \mathcal{F} \mathcal{V}_t[F], R^\perp \rangle = \mathcal{O}(\|R^\perp\|_{\text{hs}} \eta \varrho^{-1}).$$

Using the spectral gap $\vartheta \sim 1$ of \mathcal{F} from (38) and $R^\perp \perp F$, we infer (35) from (40) by estimating

$$\|((1 - \omega)\text{Id} - \mathcal{V}_t \mathcal{F})[R^\perp]\|_{\text{hs}}^2 \geq \text{dist}(\omega, D_{1-\vartheta}(1))^2 \|R^\perp\|_{\text{hs}}^2 \geq (\vartheta - 2\varepsilon)^2 (1 - |\alpha|^2),$$

optimizing in α and choosing $\varepsilon \leq \vartheta/3$. This shows the lemma for $\mathcal{T} = \text{Id} - \mathcal{V}_t \mathcal{F}$.

Since \mathcal{B}_0 is related by the similarity transform (27) to $\text{Id} - \mathcal{V}_1 \mathcal{F} = \text{Id} - \mathcal{C}_S \mathcal{F}$ and $\|Q\| \|Q^{-1}\| \lesssim 1$ (cf. (23c)), the operator \mathcal{B}_0 inherits the properties listed in the lemma from $\text{Id} - \mathcal{C}_S \mathcal{F}$. Finally, we can perform analytic perturbation theory for the simple isolated eigenvalue in $D_\varepsilon(0)$ of \mathcal{B}_0 to verify the lemma for $\mathcal{T} = \mathcal{B} = \mathcal{B}_0 + \mathcal{E}$ with $\mathcal{E} = \mathcal{O}(\varrho)$ (cf. (29)) and $\mathcal{T} = \mathcal{B}_* = \mathcal{B}_0 + \mathcal{E}_*$ with $\mathcal{E}_* = \mathcal{O}(\varrho)$ (cf. (32)) if ϱ_* is sufficiently small. This completes the proof of Lemma 4.7. \square

In the following corollary, we use the concepts of *left* and *right eigenvector* of an operator $\mathcal{T}: \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$. We say $V_l \in \mathbb{C}^{N \times N}$ ($V_r \in \mathbb{C}^{N \times N}$) is a left (right) eigenvector of \mathcal{T} corresponding to the eigenvalue $\lambda \in \mathbb{C}$ of \mathcal{T} if $\mathcal{T}^*[V_l] = \bar{\lambda} V_l$ ($\mathcal{T}[V_r] = \lambda V_r$).

COROLLARY 4.8. *Let $z \in \mathbb{D}^{\omega*}$ satisfy $\varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*$ for $\varrho_* \sim 1$ from Lemma 4.7.*

Let β_0 and β be the isolated eigenvalues in $D_\varepsilon(0)$ of \mathcal{B}_0 and \mathcal{B} , respectively, from Lemma 4.7. Furthermore, let $\mathcal{P}_0 = \mathcal{P}_{\mathcal{B}_0}$ and $\mathcal{P} = \mathcal{P}_{\mathcal{B}}$ be the spectral projections corresponding to the isolated eigenvalue of \mathcal{B}_0 and \mathcal{B} , respectively (see (34)). Then with $\mathcal{Q}_0 := \text{Id} - \mathcal{P}_0$ and $\mathcal{Q} := \text{Id} - \mathcal{P}$ we have

$$(41) \quad \|\mathcal{B}^{-1} \mathcal{Q}\| + \|\mathcal{B}^{-1} \mathcal{Q}\|_{\text{sp}} + \|\mathcal{B}_0^{-1} \mathcal{Q}_0\| \lesssim 1.$$

We define $B_0 := \mathcal{P}_0 \mathcal{C}_{\mathcal{Q}^, \mathcal{Q}}[F_U]$ and $P_0 := \mathcal{P}_0^* \mathcal{C}_{\mathcal{Q}, \mathcal{Q}^*}^{-1}[F_U]$. Then B_0 and P_0 are right and left eigenvector of \mathcal{B}_0 corresponding to β_0 and we have*

$$(42a) \quad B_0 = \mathcal{C}_{\mathcal{Q}^*, \mathcal{Q}}[F_U] + \mathcal{O}(\eta\varrho^{-1}), \quad P_0 = \mathcal{C}_{\mathcal{Q}, \mathcal{Q}^*}^{-1}[F_U] + \mathcal{O}(\eta\varrho^{-1}),$$

$$(42b) \quad \beta_0 = \frac{\eta}{\varrho} \frac{\pi}{\langle F_U^2 \rangle} + \mathcal{O}(\eta^2 \varrho^{-2}) = \mathcal{O}(\eta\varrho^{-1}).$$

We also define $B := \mathcal{P}[B_0]$ and $P := \mathcal{P}^[P_0]$. This yields right and left eigenvectors of \mathcal{B} corresponding to β which satisfy*

$$(43a) \quad B = B_0 + \mathcal{O}(\varrho),$$

$$(43b) \quad P = P_0 + \mathcal{O}(\varrho),$$

$$(43c) \quad \beta \langle P, B \rangle = \pi \eta \varrho^{-1} - 2i\varrho\sigma + \mathcal{O}(\varrho^2 + \eta + \eta^2 \varrho^{-2}).$$

Moreover, we have

$$(44) \quad \|B\| \lesssim 1, \quad \|P\| \lesssim 1, \quad |\langle P, B \rangle| \sim 1.$$

The following identity will be used a few times:

$$(45) \quad \langle F_U \mathcal{Q} \mathcal{Q}^* \rangle = \varrho^{-1} \langle \mathfrak{S} M \rangle = \pi.$$

It is obtained by a direct computation starting from the definition of F_U in (21), the balanced polar decomposition, $M = \mathcal{Q}^* U \mathcal{Q}$ and $\varrho(z) = \pi^{-1} \langle \mathfrak{S} M(z) \rangle$.

PROOF OF COROLLARY 4.8. The bounds in (41) are a direct consequence of Lemma 4.7. Using (31) and $\mathcal{C}_S[F_U] = F_U$, we see that

$$(46) \quad \mathcal{B}_0^* \mathcal{C}_{\varrho, \varrho^*}^{-1}[F_U] = \eta \varrho^{-1} I, \quad \mathcal{B}_0 \mathcal{C}_{\varrho^*, \varrho}[F_U] = \mathcal{O}(\eta \varrho^{-1}).$$

The representations of B_0 and P_0 in (42a) follow by simple perturbation theory because β_0 is a nondegenerate isolated eigenvalue. The expression for β_0 in (42b) is seen by taking the scalar product with B_0 in the first identity of (46) as well as using (42a) and (45).

The expansions (43) follow by first-order analytic perturbation theory. Indeed, $B = B_0 + \mathcal{O}(\varrho)$ and $P = P_0 + \mathcal{O}(\varrho)$ as $\mathcal{E} = \mathcal{B} - \mathcal{B}_0 = \mathcal{O}(\varrho)$ due to (29). For the proof of (43c), we first compute $\mathcal{E}[B_0]$. From (42a), we obtain the first equality below:

$$(47) \quad \begin{aligned} \mathcal{E}[B_0] &= \mathcal{C}_{\varrho^*, \varrho}(\mathcal{C}_S - \mathcal{C}_U)\mathcal{F}[F_U] + \mathcal{O}(\eta) \\ &= -2i\varrho \mathcal{C}_{\varrho^*, \varrho}[SF_U^2] + \mathcal{O}(\varrho^2 + \eta). \end{aligned}$$

For the second equality in (47), we used (31), $\|\mathcal{C}_S - \mathcal{C}_U\| = \mathcal{O}(\varrho)$ and $(\mathcal{C}_S - \mathcal{C}_U)[F_U] = 2(\Im U - i\Re U)(\Im U)F_U = -2i\varrho SF_U^2 + \mathcal{O}(\varrho^2)$ due to (28). For the proof of (43c), we start from $\mathcal{B}[B] = \beta B$, $\mathcal{B} = \mathcal{B}_0 + \mathcal{E}$, use (43a), (43b) as well as $\mathcal{E} = \mathcal{O}(\varrho)$ and obtain

$$(48) \quad \beta \langle P, B \rangle = \beta_0 \langle P_0, B_0 \rangle + \langle P_0, \mathcal{E}[B_0] \rangle + \mathcal{O}(\varrho^2).$$

Together with the following two expansions, this yields (43c). We have

$$\begin{aligned} \beta_0 \langle P_0, B_0 \rangle &= \pi \eta \varrho^{-1} + \mathcal{O}(\eta^2 \varrho^{-2}), \\ \langle P_0, \mathcal{E}[B_0] \rangle &= -2i\varrho \langle SF_U^3 \rangle + \mathcal{O}(\varrho^2 + \eta) = -2i\varrho \sigma + \mathcal{O}(\varrho^2 + \eta). \end{aligned}$$

The first expansion is a consequence of $\langle P_0, B_0 \rangle = \langle F_U^2 \rangle + \mathcal{O}(\eta \varrho^{-1})$ due to (42a) and (42b). The second expansion follows from (42a) and (47).

The first two bounds in (44) follow directly from (43a) and (43b) as well as (42a), (23c) and (23d). Moreover, (42a), (43a) and (43b) imply $|\langle P, B \rangle| \sim \langle F_U^2 \rangle \sim 1$ by (23d).

This completes the proof of Corollary 4.8. \square

PROOF OF PROPOSITION 4.4. As in the proof of Lemma 4.7, it suffices to show the bound on $\|\mathcal{B}^{-1}\|_{\text{sp}}$ in (22).

From (26), by using Lemma 4.6, we conclude that

$$\begin{aligned} \|\mathcal{B}^{-1}\|_{\text{sp}} &\lesssim \|(\mathcal{C}_U^* - \mathcal{F})^{-1}\|_{\text{sp}} \lesssim |1 - \|\mathcal{F}\|_{\text{sp}} \langle F, \mathcal{C}_U^*[F] \rangle|^{-1} \\ &\lesssim (1 - \|\mathcal{F}\|_{\text{sp}} + |1 - \langle F, \mathcal{C}_U^*[F] \rangle|)^{-1}. \end{aligned}$$

Here, we applied the rotation-inversion Lemma, Lemma 4.9 in [4], with $\mathcal{T} = \mathcal{F}$ and $\mathcal{U} = \mathcal{C}_U^*$ in the second step. Its conditions are met due to Fact (a) and Fact (c) about \mathcal{F} from the proof of Lemma 4.7.

Owing to (36) as well as (23c) and (23d), we have $1 - \|\mathcal{F}\|_{\text{sp}} \sim \eta \varrho^{-1}$. Therefore, it suffices to show that

$$(49) \quad |1 - \langle F, \mathcal{C}_U^*[F] \rangle| \gtrsim \varrho(\varrho + |\sigma|)$$

when $\eta \varrho^{-1}$ is small. As $1 \geq \langle F \Re U F \Re U \rangle$ due to $\|F\|_{\text{hs}} = 1$, we estimate

$$|1 - \langle F, \mathcal{C}_U^*[F] \rangle| = |1 - \langle FU^*FU^* \rangle| \gtrsim \langle F \Im U F \Im U \rangle + |\langle F \Im U F \Re U \rangle|.$$

Since $\Im U \sim \varrho$ by (23d), the first term on the right-hand side scales like $\sim \varrho^2$. This proves (49) when $\varrho \geq \varrho_*$ for any $\varrho_* \sim 1$ as $|\sigma| \lesssim 1$. If ϱ_* is sufficiently small and $\varrho + \eta \varrho^{-1} \leq \varrho_*$ then we use $\langle F \Im U F \Re U \rangle = \varrho \|F_U\|_{\text{hs}}^{-2} \langle F_U^3 S \rangle + \mathcal{O}(\varrho^3 + \eta)$ by (39) and (28) to conclude (49), and thus (22) in the missing regime. This completes the proof of Proposition 4.4. \square

4.2. *Sharp bound on \mathcal{B}^{-1} and 1/2-Hölder continuity of M .* In this section, we will prove Theorem 4.2 and Corollary 4.3. They will be proven directly after the following proposition, the main result of the present section. It shows that σ introduced in (21) is of order one close to regular edges $\tau_0 \in \partial \text{supp } \varrho$. For the formulation of this proposition, we define

$$(50) \quad \mathcal{A}[R, T] := \frac{1}{2}(MS[R]T + TS[R]M)$$

with $R, T \in \mathbb{C}^{N \times N}$.

PROPOSITION 4.9. *Let (A), (E) and (G) be satisfied for some $\tau_0 \in \mathbb{R}$. If $\tau_0 \in \partial \text{supp } \varrho$ is a regular edge, then the following statements hold true:*

(i) *At $z = \tau_0$, for P and B defined as in Corollary 4.8, we have*

$$|\langle P, \mathcal{A}[B, B] \rangle| \sim 1.$$

(ii) *There is $\delta_* \sim 1$ such that*

$$|\sigma(z)| \sim 1$$

for all $z \in \mathbb{H}$ satisfying $|z - \tau_0| \leq \delta_*$.

Proposition 4.9 immediately implies Theorem 4.2 and Corollary 4.3.

PROOF OF THEOREM 4.2. By Proposition 4.9(ii), there is $\delta_* \sim 1$ such that $|\sigma(z)| \sim 1$ for all $z \in \mathbb{H}$ satisfying $|z - \tau_0| \leq \delta_*$. Therefore, Theorem 4.2 follows directly from Proposition 4.4. \square

PROOF OF COROLLARY 4.3. We proceed exactly as in the proof of Corollary 4.5 but use Theorem 4.2 instead of (22) for all $z \in \mathbb{H}$ such that $|z - \tau_0| \leq \delta_*$, where δ_* is chosen as in Theorem 4.2. \square

The proof of Proposition 4.9 requires three auxiliary lemmas whose proofs are postponed until the end of the section. Some statements in these lemmas will be stated for more general $\tau_0 \in \mathbb{R}$ not only when τ_0 is a regular edge, although we will eventually use them in this case.

We now choose $\theta = \omega_*/2$ in Corollary 4.5 and work on the set $\mathbb{D}^{\omega_*/2}$ in the following. Note that $\mathbb{D}^{\omega_*/2} \subset \mathbb{D}^{\omega_*}$. By Hölder continuity, we can then extend M to $\overline{\mathbb{D}^{\omega_*/2}}$, and we denote the extension by M as well. Moreover, the operators \mathcal{B} and \mathcal{B}_* are defined for all $z \in \overline{\mathbb{D}^{\omega_*/2}}$ and the results about \mathcal{B} and \mathcal{B}_* in Lemma 4.7 hold true on $\overline{\{z \in \mathbb{D}^{\omega_*/2} : \varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*\}}$, where the closure is taken with respect to the Euclidean topology on \mathbb{C} . Lemma 4.11 below shows that this set contains a neighborhood around any point $\tau_0 \in \partial \text{supp } \varrho$.

LEMMA 4.10 (Hölder continuity of eigenvalues and eigenvectors). *Let (A), (E) and (G) be satisfied for some $\tau_0 \in \mathbb{R}$. Let β and β_* be the eigenvalues of \mathcal{B} and \mathcal{B}_* in $D_\varepsilon(0)$, respectively, (cf. Lemma 4.7) and P and B defined as in Corollary 4.8. Then there is $\varrho_* \sim 1$ such that β, β_*, P and B are uniformly 1/3-Hölder continuous functions of z on the domain $\{z \in \mathbb{D}^{\omega_*/2} : \varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*\}$.*

LEMMA 4.11. *Let (A), (E) and (G) hold true for some $\tau_0 \in \mathbb{R}$. Then the following hold true:*

(i) There is $\varrho_* \sim 1$ such that, for the eigenvalue β_* of $\mathcal{B}_* = \text{Id} - \mathcal{C}_{M^*,MS}$ in $D_\varepsilon(0)$ (cf. Lemma 4.7), we have

$$(51) \quad |\beta_*| \sim \eta/\varrho$$

uniformly for $z \in \mathbb{D}^{\omega_*}$ satisfying $\varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*$.

(ii) If $\tau_0 \in \partial \text{supp } \varrho$ and $\varrho_* \sim 1$, then there is $\delta_* \sim 1$ such that $\varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*$ for all $z \in \mathbb{H}$ satisfying $|z - \tau_0| \leq \delta_*$.

Moreover, we have

$$(52) \quad \lim_{\eta \downarrow 0} \eta\varrho(\tau_0 + i\eta)^{-1} = 0.$$

LEMMA 4.12. Let (A), (E) and (G) be satisfied for some $\tau_0 \in \mathbb{R}$. Then there is $\varrho_* \sim 1$ such that, uniformly for all $z \in \mathbb{D}^{\omega_*}$ satisfying $\varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*$, we have

$$(53a) \quad \langle P, \mathcal{A}[B, B] \rangle = \sigma + \mathcal{O}(\varrho + \eta\varrho^{-1}),$$

$$(53b) \quad \langle P, MS[B]B \rangle = \sigma + \mathcal{O}(\varrho + \eta\varrho^{-1}).$$

We remark that (53b) will be used in the next section. We note that owing to Proposition 2.2 in [4] there is a continuous function ϱ on \mathbb{R} such that $\langle M \rangle$ is the Stieltjes transform of ϱ , that is,

$$(54) \quad \langle M(z) \rangle = \int_{\mathbb{R}} \frac{\varrho(\tau)}{\tau - z} d\tau$$

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

PROOF OF PROPOSITION 4.9. In this proof, we will analyze M and $\mathcal{B} = \text{Id} - \mathcal{C}_M \mathcal{S}$ on the real line outside the self-consistent spectrum, that is, we will consider spectral parameters $z = \tau + i\eta$ such that $\tau \in [\tau_0 - \omega_*/2, \tau_0 + \omega_*/2] \setminus \text{supp } \varrho$ and $\eta = 0$. In particular, $\varrho(\tau) = 0$, and thus $M = M^*$ by (23b). Owing to the continuity of M (Corollary 4.5), M satisfies the MDE, (1), also for these spectral parameter z . Moreover, $\varrho(\tau + i\eta) \lesssim \eta / \text{dist}(\tau + i\eta, \text{supp } \varrho)^2$ as $\langle M \rangle$ is the Stieltjes transform of the measure with density ϱ on \mathbb{R} (compare (54)). Thus \mathcal{B} is invertible at $\tau \notin \text{supp } \varrho$ due to Proposition 4.4 as the term $\eta\varrho^{-1}$ has a uniform lower bound for $z = \tau + i\eta$ with $\eta > 0$. In particular, M and β are differentiable with respect to $\omega = \tau - \tau_0$ for $\tau \notin \text{supp } \varrho$. First-order perturbation theory of the isolated eigenvalue β of the nonself-adjoint operator \mathcal{B} yields

$$(55) \quad \begin{aligned} \partial_\omega \beta &= - \frac{\langle P, \mathcal{C}_{\partial_\omega M, MS}[B] \rangle}{\langle P, B \rangle} - \frac{\langle P, \mathcal{C}_{M, \partial_\omega M \mathcal{S}}[B] \rangle}{\langle P, B \rangle} \\ &= - \frac{\langle P, (\partial_\omega M) \mathcal{S}[B]M + MS[B](\partial_\omega M) \rangle}{\langle P, B \rangle}. \end{aligned}$$

For definiteness, we assume in the following that τ_0 is a right edge. Hence, $\omega > 0$. The argument for a left edge works completely analogously.

Owing to the invertibility of \mathcal{B} , the MDE, (1), is differentiable at τ with respect to ω . Similar to (24), we obtain

$$\partial_\omega M = \mathcal{B}^{-1}[M^2] = \frac{\langle P, M^2 \rangle}{\beta \langle P, B \rangle} B + \mathcal{B}^{-1} \mathcal{Q}[M^2].$$

In the second step, we inserted $\mathcal{P} + \mathcal{Q} = \text{Id}$ and employed the definition of $\mathcal{P} = \mathcal{P}_{\mathcal{B}}$ in Corollary 4.8. We insert this into (55) and get from Lemma 4.7 and (41) that

$$\begin{aligned} \partial_{\omega}\beta &= -\frac{\langle P, M^2 \rangle}{\beta \langle P, B \rangle^2} \langle P, B\mathcal{S}[B]M + M\mathcal{S}[B]B \rangle + \mathcal{O}(1) \\ &= \frac{2\langle P, M^2 \rangle}{\beta \langle P, B \rangle^2} \langle P, \mathcal{A}[B, B] \rangle + \mathcal{O}(1). \end{aligned}$$

The bounds in (44) of Corollary 4.8 yield $\|P\| \lesssim 1$, and hence, $|\langle P, M^2 \rangle| \lesssim 1$ by Assumption (G). By (44), we have $|\langle P, B \rangle| \sim 1$ if $\eta > 0$. Thus, as a consequence of the continuity of P and B by Lemma 4.10, the derivative of β^2 is bounded by $|\partial_{\omega}(\beta^2)| \lesssim |\langle P, \mathcal{A}[B, B] \rangle| + |\beta|$. This implies

$$(56) \quad |\beta|^2 \lesssim \max\{|\langle P, \mathcal{A}[B, B] \rangle| \omega, \omega^2\} \lesssim |\langle P, \mathcal{A}[B, B] \rangle| \omega + \omega^2$$

by distinguishing the regimes $|\langle P, \mathcal{A}[B, B] \rangle| > |\beta|$ and $|\langle P, \mathcal{A}[B, B] \rangle| \leq |\beta|$. On the other hand, from (51) and the continuity of β_* , and $\beta_* = \beta$ for $\eta = 0$ (as $M = M^*$) we get

$$|\beta(\tau_0 + \omega)| \sim \lim_{\eta \downarrow 0} \frac{\eta}{\varrho(\tau_0 + \omega + i\eta)} \sim \left(\int_0^{\delta} \frac{\varrho(\tau_0 - \omega')}{(\omega' + \omega)^2} d\omega' \right)^{-1},$$

for some $\delta \sim 1$. From this and (5), we conclude that

$$\liminf_{\omega \downarrow 0} \frac{|\beta(\tau_0 + \omega)|}{\sqrt{\omega}} \sim \limsup_{\omega \downarrow 0} \frac{|\beta(\tau_0 + \omega)|}{\sqrt{\omega}} \sim 1,$$

that is, $|\beta|^2 \sim \omega$ as $\omega \downarrow 0$. Therefore, we find $|\langle P, \mathcal{A}[B, B] \rangle| \gtrsim 1$ at $z = \tau_0$ due to (56). The upper bound follows from $\|P\| \lesssim 1$ and $\|B\| \lesssim 1$ by Corollary 4.8. This completes the proof of (i).

For the proof of (ii), we conclude that $\langle P, \mathcal{A}[B, B] \rangle$ is a uniformly 1/3-Hölder continuous function of z on $\{w \in \mathbb{H} \cup \mathbb{R} : |w - \tau_0| \leq \delta_*\}$ for some $\delta_* \sim 1$ due to Corollary 4.5, Lemma 4.10 and Lemma 4.11(ii). By possibly shrinking $\delta_* \sim 1$, we can thus assume that $|\langle P, \mathcal{A}[B, B] \rangle| \sim 1$ for all $z \in \mathbb{H}$ satisfying $|z - \tau_0| \leq \delta_*$. From Lemma 4.11(ii) and (53a), we conclude that $|\sigma(z)| \sim 1$ for all $z \in \mathbb{H}$ such that $|z - \tau_0| \leq \delta_*$ for some sufficiently small $\delta_* \sim 1$. Hence, we have completed the proof of Proposition 4.9. \square

PROOF OF LEMMA 4.10. This proof was already given in [5], Lemma 5.4; for the convenience of the reader, we only recall the main idea here. All arguments in this proof are performed on $\{z \in \mathbb{D}^{\omega_*/2} : \varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*\}$ with a sufficiently small $\varrho_* \sim 1$. The isolated eigenvalues β and β_* are 1/3-Hölder continuous by the definition of \mathcal{B} and \mathcal{B}_* and Corollary 4.5. A short computation starting from the definition of B in Corollary 4.8 reveals that

$$B = \mathcal{P}\mathcal{P}_0[\varrho^{-1}\mathfrak{S}M], \quad P = -\mathcal{P}^*\mathcal{P}_0^*[\varrho^{-1}\mathfrak{S}M^{-1}].$$

In these representations, the right-hand sides are easily seen to be 1/3-Hölder continuous as follows. The projections \mathcal{P} and \mathcal{P}^* are 1/3-Hölder continuous due to the 1/3-Hölder continuity of \mathcal{B} . Taking the imaginary part of the MDE, (1), and using the 1/3-Hölder continuity of \mathcal{B}_* yield the same regularity for $\varrho^{-1}\mathfrak{S}M$. Consequently, $\varrho^{-1}\mathfrak{S}M^{-1}$, \mathcal{Q} , U and S , and thus \mathcal{B}_0 are 1/3-Hölder continuous. This regularity for \mathcal{B}_0 is inherited by \mathcal{P}_0 and \mathcal{P}_0^* . This proves that B and P are 1/3-Hölder continuous. \square

PROOF OF LEMMA 4.11. Similar to the proof of Corollary 4.8, we find a left eigenvector P_* of \mathcal{B}_* corresponding to β_* , that is, $(\mathcal{B}_*)^*[P_*] = \overline{\beta_*}P_*$, such that

$$(57) \quad P_* = Q^{-1}F_U(Q^*)^{-1} + \mathcal{O}(\varrho + \eta\varrho^{-1})$$

provided that $z \in \mathbb{D}^{\omega_*}$ satisfies $\varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*$. We take the imaginary part of (1) and compute the scalar product with P_* . This yields

$$(58) \quad \beta_* = \frac{\eta \langle P_*, M^*M \rangle}{\varrho \langle P_*, \varrho^{-1}\mathfrak{S}M \rangle}.$$

Using (57) and the balanced polar decomposition, $M = Q^*UQ$, we obtain

$$\begin{aligned} \langle P_*, M^*M \rangle &= \langle F_U Q Q^* \rangle + \mathcal{O}(\varrho + \eta\varrho^{-1}) = \pi + \mathcal{O}(\varrho + \eta\varrho^{-1}), \\ \langle P_*, \varrho^{-1}\mathfrak{S}M \rangle &= \langle F_U^2 \rangle + \mathcal{O}(\varrho + \eta\varrho^{-1}). \end{aligned}$$

Here, we used that U and F_U commute and (45) in order to compute $\langle P_*, M^*M \rangle$. We thus deduce that $|\langle P_*, M^*M \rangle| \sim 1$ and $|\langle P_*, \varrho^{-1}\mathfrak{S}M \rangle| \sim 1$ for all $z \in \mathbb{D}^{\omega_*}$ satisfying $\varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*$ for some sufficiently small $\varrho_* \sim 1$ due to (23d). Therefore, taking the absolute value in (58) and using these scaling relations complete the proof of (51).

For the proof of (ii), we remark that, owing to the continuity of ϱ , we have

$$\lim_{\eta \downarrow 0} \varrho(\tau + i\eta)^{-1} \eta = 0$$

for all $\tau \in \mathbb{R}$ satisfying $\varrho(\tau) > 0$. From (51), we thus conclude that $\beta_*(\tau) = 0$ if $\varrho(\tau) > 0$ for all $\tau \in [\tau_0 - \omega_*/2, \tau_0 + \omega_*/2]$. Owing to the continuity of ϱ , we find a sequence $(\tau_n)_n$ such that $\tau_n \rightarrow \tau_0 \in \partial \text{supp } \varrho$ and $\varrho(\tau_n) > 0$ for all n . Thus the continuity of β_* (see Lemma 4.10) yields $\beta_*(\tau_0) = 0$. Therefore, we have $|\beta_*| + \varrho = 0$ at $z = \tau_0$. Hence, the 1/3-Hölder continuity of $|\beta_*| + \varrho$ implies that there is $\delta_* \sim 1$ such that $\varrho(z) + \eta\varrho(z)^{-1} \leq \varrho_*$ since $\varrho + \eta\varrho^{-1} \sim \varrho + |\beta_*|$ by (51). From $\beta_*(\tau_0) = 0$ and (51), we directly conclude (52). This completes the proof of Lemma 4.11. \square

PROOF OF LEMMA 4.12. First, we use the balanced polar decomposition, $M = Q^*UQ$, (25) and the definition of \mathcal{A} in (50) to obtain

$$(59) \quad \begin{aligned} \mathcal{A}[R, T] &= \frac{1}{2} \mathcal{C}_{\varrho^*, \varrho} [U(\mathcal{F} \mathcal{C}_{\varrho^*, \varrho}^{-1}[R]) \mathcal{C}_{\varrho^*, \varrho}^{-1}[T] \\ &\quad + \mathcal{C}_{\varrho^*, \varrho}^{-1}[T](\mathcal{F} \mathcal{C}_{\varrho^*, \varrho}^{-1}[R])U] \end{aligned}$$

for $R, T \in \mathbb{C}^{N \times N}$.

We choose $\varrho_* \sim 1$ small enough such that Lemma 4.7 is applicable. By using $U = S + \mathcal{O}(\varrho)$ due to (28) as well as (21), (31) and (42a) in (59), we get

$$(60) \quad \mathcal{A}[B_0, B_0] = \mathcal{C}_{\varrho^*, \varrho} [S F_U^2] + \mathcal{O}(\varrho + \eta\varrho^{-1}).$$

In order to show (53a), we use (43a), (43b) and $\sigma = \langle S F_U^3 \rangle$ to obtain

$$\langle P, \mathcal{A}[B, B] \rangle = \langle P_0, \mathcal{A}[B_0, B_0] \rangle + \mathcal{O}(\varrho) = \sigma + \mathcal{O}(\varrho + \eta\varrho^{-1}).$$

This completes the proof of (53a). A similar computation yields (53b). \square

4.3. Derivation of the quadratic equation. In this section, we expand $M(\tau_0 + \omega)$ around $M(\tau_0)$ for a regular edge $\tau_0 \in \partial \text{supp } \varrho$. We show that this approximation is to leading order dominated by a scalar-valued quantity, Θ , which satisfies a quadratic equation. That is the content of the following proposition which is the main result of this section.

PROPOSITION 4.13 (Quadratic equation for shape analysis). *Let (A), (E) as well as (G) be satisfied for some regular edge $\tau_0 \in \partial \text{supp } \varrho$. Then there is $\delta_* \sim 1$ such that the following hold true:*

(a) For all $\omega \in [-\delta_*, \delta_*]$, we have

$$(61) \quad M(\tau_0 + \omega) - M(\tau_0) = \Theta(\omega)B + R(\omega),$$

where $\Theta : [-\delta_*, \delta_*] \rightarrow \mathbb{C}$ and $R : [-\delta_*, \delta_*] \rightarrow \mathbb{C}^{N \times N}$ are defined by

$$(62) \quad \Theta(\omega) := \left\langle \frac{P}{\langle B, P \rangle}, M(\tau_0 + \omega) - M(\tau_0) \right\rangle,$$

$$R(\omega) := \mathcal{Q}[M(\tau_0 + \omega) - M(\tau_0)].$$

Here, $P = P(\tau_0)$, $B = B(\tau_0)$ and $\mathcal{Q} = \mathcal{Q}(\tau_0)$ are the eigenvectors and spectral projection of $\mathcal{B}(\tau_0)$ introduced in Corollary 4.8. We have $B = B^*$ and $P = P^*$ as well as $B \sim 1$ and $P \sim 1$. Moreover, $\Theta(\omega)$ and $R(\omega)$ are bounded by

$$(63) \quad |\Theta(\omega)| \lesssim |\omega|^{1/2}, \quad \Im \Theta(\omega) \geq 0, \quad \|\Im R(\omega)\| \lesssim |\omega|^{1/2} \Im \Theta(\omega)$$

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

(b) The function Θ satisfies the quadratic equation

$$(64) \quad \sigma \Theta^2(\omega) + \omega \Xi(\omega) = 0, \quad \Xi(\omega) = \pi(1 + \nu(\omega)),$$

for all $\omega \in [-\delta_*, \delta_*]$, where $\sigma = \langle P, M\mathcal{S}[B]B \rangle$, $M = M(\tau_0)$, and the error term $\nu(\omega)$ satisfies

$$(65) \quad |\nu(\omega)| \lesssim |\omega|^{1/2}, \quad |\Im \nu(\omega)| \lesssim \Im \Theta(\omega)$$

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

The definition $\sigma = \langle P, M\mathcal{S}[B]B \rangle$ for $\tau_0 \in \partial \text{supp } \varrho$ extends the definition of σ in (21) on \mathbb{H} owing to (53b), (52) as well as the continuity of M , and thus P , B and ϱ .

We warn the reader that, in this section, functions of z like M , B , P , U , Q , etc. without argument are understood to be evaluated at τ_0 instead of the generic spectral parameter z which is the convention in most of the other parts of this work.

PROOF OF PROPOSITION 4.13. The first bound in (63) follows directly from Corollary 4.3.

From (42a), (43a), (43b), $\varrho(\tau_0) = 0$ and (52), we conclude that B and P are the limits of Hermitian, positive-definite matrices which are ~ 1 due to Lemma 4.6. Thus $B = B^* \sim 1$ and $P = P^* \sim 1$. This also implies that $\Im \Theta(\omega) \geq 0$ in (63) as $\Im M(\tau_0 + \omega)$ is always positive semidefinite and $\Im M(\tau_0) = 0$.

In the following lemma whose proof we postpone until the end of this section, we establish a quadratic equation for Θ .

LEMMA 4.14 (Derivation of the quadratic equation). *Let $\Theta(\omega)$ and $R(\omega)$ be defined as in (62) and \mathcal{A} be defined as in (50). Then there is $\delta_* \sim 1$ such that, for all $\omega \in [-\delta_*, \delta_*]$, $\Theta = \Theta(\omega)$ satisfies the quadratic equation*

$$\mu_2 \Theta^2 + \mu_1 \Theta + \mu_0 = e(\omega)$$

with some error term $e(\omega) = \mathcal{O}(|\omega|^{3/2})$ and with coefficients

$$(66) \quad \mu_2 = \langle P, \mathcal{A}[B, B] \rangle, \quad \mu_1 = -\beta \langle P, B \rangle, \quad \mu_0 = \omega \langle P, M^2 \rangle.$$

Moreover, for all $\omega \in [-\delta_*, \delta_*]$, we have

$$(67) \quad |\Im e(\omega)| \lesssim |\omega| \Im \Theta(\omega), \quad \|\Im R(\omega)\| \lesssim |\omega|^{1/2} \Im \Theta(\omega).$$

We now compute the coefficients defined in (66) precisely. This will yield the quadratic equation in (64).

Owing to (53a), (53b), (52), $\varrho(\tau_0) = 0$ and the continuity of M , and thus P , B and ϱ , we have $\mu_2 = \sigma$ as defined in Proposition 4.13(b).

The expansion in (43c) implies $\mu_1 = 0$ at τ_0 by (52). We now compute μ_0 . At $z \in \mathbb{H}$ satisfying $\varrho(z) + \varrho(z)^{-1} \Im z \leq \varrho_*$, we conclude from (43b), (42a) and the balanced polar decomposition, $M = Q^*UQ$, from (19) that

$$\begin{aligned} \langle P, M^2 \rangle &= \langle Q^{-1}F_U(Q^*)^{-1}, Q^*UQQ^*UQ \rangle + \mathcal{O}(\varrho + \eta\varrho^{-1}) \\ &= \langle F_UQQ^* \rangle + \mathcal{O}(\varrho + \eta\varrho^{-1}) = \pi + \mathcal{O}(\varrho + \eta\varrho^{-1}). \end{aligned}$$

Here, we also employed that $U = S + \mathcal{O}(\varrho)$ by (28) and F_U and S commute in the second step and (45) in the last step. Thus we have $\mu_0 = \omega\pi$ at τ_0 by (52).

We set $v(\omega) := -(\pi\omega)^{-1}e(\omega)$ with $e(\omega)$ as introduced in Lemma 4.14. This immediately implies the first bound in (65). From (67), we conclude the second estimate in (65) and the third estimate in (63). This completes the proof of Proposition 4.13. \square

PROOF OF LEMMA 4.14. Owing to the Hölder continuity of M , we conclude that $M(z)$ is invertible and satisfies (1) for all $z \in \overline{\mathbb{D}^{\omega_*/2}}$. Hence, evaluating (1) at $z = \tau_0 + \omega$ and $z = \tau_0$, computing their difference and introducing $M := M(\tau_0)$ as well as $\Delta := M(\tau_0 + \omega) - M$, we obtain

$$(68) \quad \mathcal{B}[\Delta] = \mathcal{A}[\Delta, \Delta] + \omega M^2 + \omega \mathcal{K}[\Delta], \quad \mathcal{K}[\Delta] := \frac{1}{2}(M\Delta + \Delta M).$$

In order to compute $R = \mathcal{Q}[\Delta]$, we apply $\mathcal{B}^{-1}\mathcal{Q}$ to (68), use $\Delta = \Theta B + R$ and, owing to the Hölder continuity of M , $\|\Delta(\omega)\| \lesssim |\omega|^{1/2}$ and $|\Theta(\omega)| \lesssim |\omega|^{1/2}$, find $\delta_* \sim 1$ such that

$$(69) \quad \|R(\omega)\| \lesssim |\omega|, \quad \|\Im R(\omega)\| \lesssim |\omega|^{1/2} \Im \Theta(\omega)$$

for all $\omega \in [-\delta_*, \delta_*]$. Here, in order to estimate $\Im R$, we used that $M = M^*$, and hence, $\mathcal{A}[B, B] = \mathcal{A}[B, B]^*$ as $\tau_0 \in \partial \text{supp } \varrho$. This shows the second estimate in (67).

We apply $\langle P, \cdot \rangle$ to (68) and use the decomposition $\Delta = \Theta B + R$ as well as $\mathcal{B}[B] = \beta B$ which yield

$$\begin{aligned} \Theta \beta \langle P, B \rangle &= \omega \langle P, M^2 \rangle + \Theta^2 \langle P, \mathcal{A}[B, B] \rangle + e, \\ e &:= \langle P, \Theta(\mathcal{A}[B, R] + \mathcal{A}[R, B]) + \mathcal{A}[R, R] \rangle + \omega \langle P, \mathcal{K}[\Delta] \rangle. \end{aligned}$$

From (69), we conclude

$$|e(\omega)| \lesssim |\omega|^{3/2}, \quad |\Im e(\omega)| \lesssim |\omega| \Im \Theta(\omega).$$

This establishes the quadratic equation as well as the missing bounds on e , and thus completes the proof of Lemma 4.14. \square

4.4. *Shape analysis.* In this section, we conclude Theorem 4.1 from Proposition 4.13.

PROOF OF THEOREM 4.1. We recall that $\sigma = \langle P, MS[B]B + BS[B]M \rangle / 2$ as in the proof of Proposition 4.13 and $|\sigma| \sim 1$ by Proposition 4.9(i). We will show that there is $\delta_* \sim 1$ such that

$$(70) \quad \varrho(\tau_0 + \omega) = \begin{cases} \frac{\pi^{1/2}}{|\sigma|^{1/2}} |\omega|^{1/2} + \mathcal{O}(|\omega|) & \text{if } \text{sign } \omega = \text{sign } \sigma, \\ 0 & \text{if } \text{sign } \omega = -\text{sign } \sigma, \end{cases}$$

for all $\omega \in [-\delta_*, \delta_*]$. This directly implies Theorem 4.1 with $c = \sqrt{\pi/|\sigma|}$ as we conclude $\sigma < 0$ from (5) and (70).

We now compute $\Theta(\omega)$ in (61) by identifying the correct solution of (64). The general quadratic equation $\Omega(\zeta)^2 + \zeta = 0$ with $\zeta \in \mathbb{C}$ has two solutions:

$$\Omega_{\pm}(\zeta) = \pm \begin{cases} i\zeta^{1/2} & \text{if } \Re \zeta \geq 0, \\ -(-\zeta)^{1/2} & \text{if } \Re \zeta < 0, \end{cases}$$

where $\zeta^{1/2}$ denotes the standard branch of the square root with the branch cut $(-\infty, 0)$.

Since $\Theta(\omega)$ is a continuous function of ω and $|\nu(\omega)| < 1$ for all $\omega \in [-\delta_*, \delta_*]$ for $\delta_* \sim 1$ sufficiently small due to the first bound in (65), we conclude from (64) that there are $p, q \in \{+, -\}$ such that

$$(71) \quad \begin{aligned} \Theta(\omega) &= \Omega_p(\Lambda(\omega))\mathbf{1}(\omega/\sigma < 0) + \Omega_q(\Lambda(\omega))\mathbf{1}(\omega/\sigma \geq 0), \\ \Lambda(\omega) &:= \frac{\pi\omega}{\sigma}(1 + \nu(\omega)) \end{aligned}$$

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

We now show that $q = +$ by a proof by contradiction. We assume $q = -$. For $\omega/\sigma \geq 0$, we have

$$\Im \Omega_-(\Lambda(\omega)) = -\left(\frac{\pi\omega}{\sigma}\right)^{1/2} + \mathcal{O}(|\nu(\omega)||\omega|^{1/2}).$$

For sufficiently small ω , we thus obtain $\Im \Omega_-(\Lambda(\omega)) < 0$ in contradiction to $\Im \Theta(\omega) \geq 0$ from (63). This implies $q = +$.

Next, we prove that $\Im \Theta(\omega) = 0$ for all $\omega \in I_{\delta_*}$ with $\delta_* \sim 1$ sufficiently small, where $I_{\delta_*} := \{\omega \in \mathbb{R} : \text{sign } \omega = -\text{sign } \sigma, |\omega| \leq \delta_*\}$. We will not determine p in (71) but rather show that $\Im \Theta = 0$ on I_{δ_*} for either choice of p (In fact, $p = +$ can be shown [5], Proposition 7.10(ii)). By possibly shrinking $\delta_* \sim 1$, we get

$$|\Re \Omega_{\pm}(\Lambda(\omega))| \sim |\omega|^{1/2}$$

as $\sigma \in \mathbb{R}$ and $|\sigma| \sim 1$. Therefore, taking the imaginary part of (64) and using the second bound in (65), (71) and $\sigma \in \mathbb{R}$ yield

$$|\omega|^{1/2} \Im \Theta(\omega) \lesssim |\omega| \Im \Theta(\omega)$$

for all $\omega \in I_{\delta_*}$. If $\delta_* \sim 1$ is sufficiently small then we obtain $\Im \Theta(\omega) = 0$ for all $\omega \in I_{\delta_*}$.

We now take the imaginary part of (61) and apply $\langle \cdot \rangle$. Hence, we obtain

$$(72) \quad \begin{aligned} \varrho(\tau_0 + \omega) &= \Im \Theta(\omega)\pi^{-1}\langle B \rangle + \pi^{-1}\langle \Im R(\omega) \rangle \\ &= \Im \Theta(\omega) + \mathcal{O}(|\omega|^{1/2} \Im \Theta(\omega)) \end{aligned}$$

for all $\omega \in [-\delta_*, \delta_*]$. Here, we used $B = B^*$ in the first step and $\langle B \rangle = \pi$ by (43a), (42a), (45) and (52) as well as the third bound in (63) in the second step.

Since $q = +$ in (71), we can bound $\Im \Theta(\omega) = \Im \Omega_+(\Lambda(\omega))$ directly in (72) to obtain the first case in (70). Since $\Im \Theta(\omega) = 0$ for all $\omega \in I_{\delta_*}$, (72) implies the second case in (70). This completes the proof of (70), and thus the one of Theorem 4.1. \square

4.5. *Proof of Proposition 3.1.* We have now established all results which are necessary for the proof of Proposition 3.1.

PROOF OF PROPOSITION 3.1. Claims (i) and (ii) follow directly from [29] and [4] (see also (54)).

Part (iii) is a direct consequence of Theorem 4.1 and the Stieltjes transform representation of $\langle M(z) \rangle$ in (54).

For the proof of (iv), we first remark that (iii) implies $\varrho(z) + \eta\varrho(z)^{-1} \sim \sqrt{|\tau - \tau_0| + \eta}$ for all $z \in \mathbb{H}$ satisfying $|z - \tau_0| \leq \delta_*$. Thus Theorem 4.2 yields the first bound in (iv). Owing to (41), we have $\|\mathcal{B}^{-1}\mathcal{Q}\|_{\text{sp}} \lesssim 1$. Moreover, we choose P and B as in Corollary 4.8. This completes the proof of the second bound in (iv) due to (44).

Moreover, $|\sigma| \sim 1$ by Proposition 4.9. Hence, we conclude $|\langle P, M\mathcal{S}[B]B \rangle| \sim 1$ from Lemma 4.12. Furthermore, owing to (44), we have $|\langle P, B \rangle| \sim 1$. Thus, since $\sigma \in \mathbb{R}$ and $|\sigma| \sim 1$, we get from (43c) that

$$|\beta| \sim |\beta\langle P, B \rangle| \sim \varrho + \eta\varrho^{-1} \sim \sqrt{|\tau - \tau_0| + \eta}.$$

This completes the proof of Proposition 3.1. \square

5. Band rigidity. Within this section, we establish band rigidity for correlated random matrices H . This topological rigidity phenomenon asserts that the number of eigenvalues of H within a spectral band, that is, a connected component of $\text{supp } \varrho$, does not fluctuate and is accurately predicted by the self-consistent density of states with high probability. On the level of the MDE, this phenomenon is reflected by the *band mass formula* (73) below, guaranteeing that $N\varrho$ assigns only integer values to each band. In particular, small continuous deformations of the data (A, \mathcal{S}) of the MDE cannot change these values.

PROPOSITION 5.1 (Band mass formula). *For $\tau \in \mathbb{R} \setminus \text{supp } \varrho$, the integrated self-consistent density of states satisfies*

$$(73) \quad \int_{-\infty}^{\tau} \varrho(x) \, dx = \frac{1}{N} |\text{Spec}(M(\tau)) \cap (-\infty, 0)|.$$

In particular, $N \int_{-\infty}^{\tau} \varrho(x) \, dx$ is an integer.

Before we prove Proposition 5.1, we show how it is used to establish band rigidity for H .

PROOF OF COROLLARY 2.9. We begin with the proof of (7a) and consider a flow that interpolates between $H = H_0$ and a deterministic matrix H_1 . We fix $\tau \notin \text{supp } \varrho$ with $\epsilon := \text{dist}(\tau, \text{supp } \varrho) > 0$ and set

$$(74) \quad \begin{aligned} H_t &:= \sqrt{1-t}W + A_t, & A_t &:= A - t\mathcal{S}[M(\tau)], \\ \mathcal{S}_t &:= (1-t)\mathcal{S}, & t &\in [0, 1]. \end{aligned}$$

The MDE corresponding to H_t is

$$(75) \quad I + (z - A_t + \mathcal{S}_t[M_t(z)])M_t(z) = 0$$

with data (A_t, \mathcal{S}_t) , solution $M_t(z)$ and self-consistent density of states ϱ_t . We refer to this t -dependent MDE as MDE_t . It is designed in such a way that $M_t(\tau)$ at the fixed spectral parameter $z = \tau$ is kept constant at t varies. Moreover, by the following lemma, whose proof we postpone, τ stays away from the self-consistent spectrum along the flow.

LEMMA 5.2. *Let $\epsilon := \text{dist}(\tau, \text{supp } \varrho) > 0$ and M_t be the solution to MDE_t (75). Then $\text{dist}(\tau, \text{supp } \varrho_t) \geq \epsilon$ and $\lim_{\eta \downarrow 0} M_t(\tau + i\eta) = M(\tau)$ for all $t \in [0, 1]$.*

We will now show that along the flow, with overwhelming probability, no eigenvalue crosses the spectral parameter τ . More precisely, we claim that

$$(76) \quad \mathbf{P}(\tau \in \text{Spec } H_t \text{ for some } t \in [0, 1]) \leq \epsilon N^{-D}$$

for any $D > 0$. Since $H_0 = H$ and $H_1 = A - \mathcal{S}[M(\tau)]$, (76) implies that with overwhelming probability

$$\begin{aligned} |\text{Spec } H \cap (-\infty, \tau)| &= |\text{Spec}(A - \mathcal{S}[M(\tau)] - \tau) \cap (-\infty, 0)| \\ &= N(\mathbf{1}_{(-\infty, 0)}(M(\tau))), \end{aligned}$$

where the last identity used the the MDE (1) at $z = \tau$. Now (7a) follows from the band mass formula (73), that is, from $(\mathbf{1}_{(-\infty, 0)}(M(\tau))) = \int_{-\infty}^{\tau} \varrho(\lambda) \, d\lambda$.

It remains to show (76). We first consider the regime of values t close to 1. Since τ is separated away from $\text{supp } \varrho$, and $M(\tau)$ is bounded we conclude from (1) at $z = \tau$ that the spectrum of $A - \mathcal{S}[M(\tau)]$ is also separated away from τ . Moreover, applying [21], Corollary 2.3, to $H = W$ yields $\|W\| \leq C$ with overwhelming probability as the corresponding self-consistent density of states has compact support by Proposition 3.1(ii). Since therefore H_t is a small perturbation of $A - \mathcal{S}[M(\tau)]$ as long as t is close to 1, we conclude that the spectrum of H_t is bounded away from τ as well for every fixed $t \geq 1 - c$ for some small enough constant $c > 0$. We are thus left with the regime $t \leq 1 - c$, where the flatness condition from Assumption (E) for H_t is satisfied. In this regime, we use [21], Corollary 2.3, again. Since $\text{dist}(\tau, \text{supp } \varrho_t) \geq \epsilon$, this corollary implies that the spectrum of H_t is bounded away from τ with overwhelming probability for every fixed $t \leq 1 - c$. Applying a discrete union bound in t together with the Lipschitz continuity of the eigenvalues in t for the flow (74) on the set $\|W\| \leq C$ yields (76).

Finally, (7b) follows from the optimal local law as in the proof of Theorem 2.6 and Corollary 2.7 above. This time, however, (7a) ensures that there is no mismatch between location and label of eigenvalues close to internal edges. In the spectral bulk, this potential discrepancy between label and location does not matter as (7b) allows for a N^ϵ -uncertainty. At the spectral edge, however, neighboring eigenvalues can lie on opposite sides of a spectral gap and we need (7a) to make sure that each eigenvalue has, with high probability, a definite location with respect to the spectral gap. \square

PROOF OF LEMMA 5.2. Note that $M(z)$ is analytic and bounded away from the self-consistent spectrum because it admits a Stieltjes transform representation (cf. Proposition 2.1 of [4]). We consider MDE $_t$ (75) at a spectral parameter $\tau + \zeta$ with some $\zeta \in \mathbb{H}$ such that $|\zeta| \ll 1$ and subtract it from MDE $_t$ at spectral parameter τ . Properly symmetrized the resulting quadratic equation for $\Delta = \Delta(\zeta) = M_t(\tau + \zeta) - M(\tau)$ takes the form

$$(77) \quad \mathcal{B}_t[\Delta] = \zeta M^2 + \frac{\zeta}{2}(M\Delta + \Delta M) + (1 - t)\mathcal{A}[\Delta, \Delta],$$

where $M = M(\tau)$, \mathcal{A} is as in (50) and $\mathcal{B}_t = \text{Id} - (1 - t)\mathcal{C}_M\mathcal{S}$ is the stability operator. We will see that equation (77) is linearly stable in the sense that $\|\mathcal{B}_t^{-1}\| \leq \epsilon$ uniformly in t . Note that the terms containing Δ on the right-hand side are lower order. Thus we may apply the implicit function theorem to show that $\Delta(\zeta)$ is an analytic function for sufficiently small ζ with $\Delta(\zeta) = \zeta \mathcal{B}_t^{-1}[M^2] + \mathcal{O}(|\zeta|^2)$. In particular, it extends to small $\zeta \in \mathbb{C}$. Since $M = M(\tau)$ is self-adjoint and \mathcal{B}_t^{-1} preserves the cone of positive definite matrices, $M + \Delta(\zeta)$ coincides for any small $\zeta \in \mathbb{H}$ with the unique solution to MDE $_t$ with positive definite imaginary part. But since $\Delta(\zeta)$ is analytic in ζ for any small enough ζ , even with negative imaginary part, $M_t(z)$ can be analytically extended to a t -independent neighborhood of τ in \mathbb{C} . Furthermore, since \mathcal{B}_t and $R \mapsto \mathcal{A}[R, R]$ preserve the space of self-adjoint matrices, this extension takes self-adjoint values on the real line. Thus for every t the density $\varrho_t = \frac{1}{\pi} \langle \Im M_t \rangle$ vanishes in a neighborhood of τ , that is, $\text{dist}(\tau, \text{supp } \varrho_t) \geq \epsilon$.

To show the bound on \mathcal{B}_t^{-1} , we use the symmetrization (26) with the self-energy operator $\mathcal{S}_t = (1 - t)\mathcal{S}$ to see that

$$(78) \quad \|\mathcal{B}_t^{-1}\|_{\text{sp}} \leq \epsilon \ \|(\mathcal{C}_U^* - \mathcal{F}_t)^{-1}\|_{\text{sp}} \lesssim \frac{1}{1 - (1 - t)\|\mathcal{F}\|_{\text{sp}}},$$

where U is unitary and $\mathcal{F}_t = (1 - t)\mathcal{F}$ with the self-adjoint operator \mathcal{F} from (25). Exactly as in the proof of Lemma 4.7, the boundedness of \mathcal{B}_t^{-1} in the $\|\cdot\|_{\text{sp}}$ -norm also implies $\|\mathcal{B}_t^{-1}\| \leq \epsilon$. Thus it remains to show that the right-hand side of (78) is bounded. For this purpose, we apply the lower bound on $1 - \|\mathcal{F}\|_{\text{sp}} \geq \epsilon$ from [6], Lemma 3.6, completing the proof of the lemma. \square

PROOF OF PROPOSITION 5.1. Let $\epsilon := \text{dist}(\tau, \text{supp } \varrho) > 0$. Again we make use of MDE_t (75). Recall that $M(\tau)$ solves MDE_t at spectral parameter τ , which stays away from the self-consistent spectrum by Lemma 5.2.

Since $M_t(z)$ is the Stieltjes transform of a matrix valued measure on $\text{supp } \varrho_t$, it can be analytically extended to $\mathbb{C} \setminus \text{supp } \varrho_t$, a set that contains the spectral parameter τ for which $M_t(\tau) = M(\tau)$ by the lemma. When ϱ and $M(\tau)$ are replaced by ϱ_t and $M_t(\tau)$, respectively, in (73) then clearly this identity holds at time $t = 1$ since $M_1(z) = (A - \mathcal{S}[M(\tau)] - z)^{-1} = (\tau + M(\tau)^{-1} - z)^{-1}$ is the resolvent of the self-adjoint matrix $\tau + M(\tau)^{-1}$. As $M_t(\tau) = M(\tau)$, it suffices to establish that the left-hand side of (73) with ϱ replaced by ϱ_t does not change along the flow.

To show that the left-hand side is independent of t , we differentiate the contour integral representation

$$\int_{-\infty}^{\tau} \varrho_t(x) dx = - \oint \frac{dz}{2\pi i} \langle M_t(z) \rangle,$$

where the contour encircles $[\min \text{supp } \varrho_t, \tau)$ counterclockwise, passing through the real line only at τ and to the left of $\inf_t \min \text{supp } \varrho_t$. With $M_t = M_t(z)$, we find

$$\begin{aligned} \frac{d}{dt} \oint \langle M_t \rangle dz &= \oint \langle (\mathcal{C}_{M_t^*}^{-1} - \mathcal{S}_t)^{-1} [I], \mathcal{S}[M(\tau) - M_t] \rangle \\ &= \oint \partial_z \left(\langle M_t \mathcal{S}[M(\tau)] \rangle - \frac{1}{2} \langle M_t \mathcal{S}[M_t] \rangle \right) dz = 0, \end{aligned}$$

where the formula $(\mathcal{C}_{M_t}^{-1} - \mathcal{S}_t)[\partial_t M_t] = \mathcal{S}[M(\tau) - M_t]$, used in the first identity, is obtained by differentiating MDE_t with data (74) with respect to t and the formula $(\mathcal{C}_{M_t}^{-1} - \mathcal{S}_t)[\partial_z M_t] = I$, used in the second identity, follows from differentiating (75) with respect to z . \square

6. Proof of universality. In order to prove Theorem 2.11, we define the Ornstein-Uhlenbeck (OU) process starting from $H = H_0$ by

$$(79) \quad dH_t = -\frac{1}{2}(H_t - A) dt + \Sigma^{1/2}[dB_t], \quad \Sigma[R] := \mathbf{E}W \text{Tr}(WR),$$

where B_t is a matrix of, up to symmetry, independent (real or complex, depending on the symmetry class of H) Brownian motions and $\Sigma^{1/2}$ is the square root of the positive definite operator $\Sigma: \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$. We note that the same process has already been used in [4, 15, 21] to prove bulk universality. The proof now has two steps: First, we will prove edge universality for H_t if $t \gg N^{-1/3}$ and then we will prove that for $t \ll N^{-1/6}$, the eigenvalues of H_t have the same k -point correlation functions as those of $H = H_0$.

6.1. *Dyson Brownian motion.* The process (79) can be integrated, and we have

$$H_t - A = e^{-t/2}(H_0 - A) + \int_0^t e^{(s-t)/2} \Sigma^{1/2} [dB_s],$$

$$\int_0^t e^{(s-t)/2} \Sigma^{1/2} [dB_s] \sim \mathcal{N}(0, (1 - e^{-t})\Sigma).$$

The process is designed in such a way that it preserves expectation $\mathbf{E}H_t = A$ and covariances $\mathbf{Cov}(h_{ab}^t, h_{cd}^t) = \mathbf{Cov}(h_{ab}, h_{cd})$ along the flow. Due to the fullness Assumption (F) there exists a constant $c > 0$ such that $(1 - e^{-t})\Sigma - ct\Sigma^{\text{GOE/GUE}} \geq 0$ for $t \leq 1$, where $\Sigma^{\text{GOE/GUE}}$ denotes the covariance operator of the GOE/GUE ensembles. It follows that we can write

$$H_t = \tilde{H}_t + \sqrt{ct}U, \quad \kappa_t = \kappa - ct\kappa^{\text{GOE/GUE}},$$

$$\mathbf{E}\tilde{H}_t = A, \quad U \sim \text{GOE/GUE},$$

where κ_t here denotes the cumulants of \tilde{H}_t , and U is chosen to be independent of \tilde{H}_t . Due to the fact that Gaussian cumulants of degree more than 2 vanish, it is easy to check that H_t, \tilde{H}_t satisfy the assumptions of Theorem 2.6 uniformly in, say, $t \leq N^{-1/10}$. From now on, we fix $t = N^{-1/3+\epsilon}$ with some small $\epsilon > 0$.

Since the MDE is purely determined by the first two moments of the corresponding random matrix, it follows that $G_t := (H_t - z)^{-1}$ is close to the same M in the sense of a local law for all t . For $\tilde{G}_t := (\tilde{H}_t - z)^{-1}$, we have the MDE

$$(80) \quad I + (z - A + \mathcal{S}_t[M_t])M_t = 0, \quad \mathcal{S}_t := \mathcal{S} - ct\mathcal{S}^{\text{GOE/GUE}}$$

that can be viewed as a perturbation of the original MDE with $t = 0$. The corresponding self-consistent density of states is $\varrho_t(\tau) := \lim_{\eta \searrow 0} \Im \langle M_t(\tau + i\eta) \rangle / \pi$. The fact that M_t remains bounded uniformly in $t \leq N^{-1/10}$ follows from a similar (but much simpler) argument as those leading to the local law in Section 3. The analogue of (10) with G replaced by $M_t(z)$ is obtained by subtracting (1) from (80) and the analogue of the error term D is trivially controlled by t . The details are presented in the MDE perturbation result in [5], Proposition 10.1, with $S = \mathcal{S}$, $\mathcal{S}_t = \mathcal{S}_t$ and $a_t = A$ as the condition on \mathcal{S}_t in [5], equation (10.1), is obviously satisfied for this choice of \mathcal{S}_t due to $\|\mathcal{S}^{\text{GOE/GUE}}[R]\| \lesssim \langle R \rangle$ for all positive semidefinite matrices R . In particular, the shape analysis from Section 4 also applies to M_t .

The Stieltjes transforms of the free convolutions of the empirical spectral density of \tilde{H}_t and ϱ_t with the semicircular distribution generated by $\sqrt{ct}U$ are given implicitly as the unique solutions to the equations

$$\tilde{m}_{\text{fc}}^t(z) = \langle \tilde{G}_t(z + ct\tilde{m}_{\text{fc}}^t(z)) \rangle, \quad m_{\text{fc}}^t(z) = \langle M_t(z + ct m_{\text{fc}}^t(z)) \rangle.$$

We denote the corresponding right-edges close to τ_0 by $\tilde{\tau}_t$ and τ_t . By differentiating the defining equations for m_{fc}^t and \tilde{m}_{fc}^t , we find

$$(81a) \quad \frac{(m_{\text{fc}}^t)'(z)}{1 + ct(m_{\text{fc}}^t)'(z)} = \langle M_t'(\xi_t(z)) \rangle,$$

$$\frac{(\tilde{m}_{\text{fc}}^t)'(z)}{1 + ct(\tilde{m}_{\text{fc}}^t)'(z)} = \langle \tilde{G}_t'(\tilde{\xi}_t(z)) \rangle,$$

$$\frac{(m_{\text{fc}}^t)''(z)}{(1 + ct(m_{\text{fc}}^t)'(z))^3} = \langle M_t''(\xi_t(z)) \rangle,$$

where $\xi_t(z) := z + ct m_{\text{fc}}^t(z)$ and $\tilde{\xi}_t(z) := z + ct \tilde{m}_{\text{fc}}^t(z)$. From the first two equalities in (81a), we conclude

$$(81b) \quad 1 = ct \langle M_t'(\xi_t(\tau_t)) \rangle, \quad 1 = ct \langle \tilde{G}_t'(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle,$$

by considering the $z \rightarrow \tau_t$ and $z \rightarrow \tilde{\tau}_t$ limits and that $(m_{fc}^t)'$, $(\tilde{m}_{fc}^t)'$ blow up at the edge due to the well-known square root behavior of the density along the semicircular flow. We now compare the edge location and edge slope of the densities ϱ_{fc}^t and $\tilde{\varrho}_{fc}^t$ corresponding to m_{fc}^t and \tilde{m}_{fc}^t with that of M . Very similar estimates for deformed Wigner ensembles have been used in [30]. We split the analysis into four claims.

Claim 1. $|\tau_t - \tau_0| \lesssim t/N$. Using that $\mathcal{S}^{\text{GUE}}[R] = \langle R \rangle$, $\mathcal{S}^{\text{GOE}}[R] = \langle R \rangle + R^t/N$ and (80) evaluated at $\xi_t(z)$, we find using the boundedness of M_t ,

$$\begin{aligned} I + (z - A + \mathcal{S}[M_t(\xi_t(z))])M_t(\xi_t(z)) \\ = ct(\mathcal{S}^{\text{GOE/GUE}}[M_t(\xi_t(z))] - \langle M_t(\xi_t(z)) \rangle)M_t(\xi_t(z)) = \mathcal{O}\left(\frac{t}{N}\right). \end{aligned}$$

It thus follows that $M_t(\xi_t(z))$ approximately satisfies the MDE for M at z . By using the first bound in Proposition 3.1(iv) expressing the stability of the MDE against small additive perturbations it follows that

$$\begin{aligned} (82) \quad |m_{fc}^t(z) - \langle M(z) \rangle| &= |\langle M_t(\xi_t(z)) - M(z) \rangle| \\ &\lesssim \frac{t}{N\sqrt{\eta + \text{dist}(\Im z, \partial \text{supp } \varrho)}} \\ &\leq \frac{t}{N\sqrt{\text{dist}(\Im z, \partial \text{supp } \varrho)}}. \end{aligned}$$

Suppose first that $\tau_0 = \tau_t + \delta$ for some positive $\delta > 0$. Then

$$\sqrt{\delta} \lesssim \Im \langle M(\tau_t + \delta/2) \rangle \lesssim \frac{t}{N\sqrt{\delta}},$$

where the first bound follows from the square root behavior of ϱ at the edge τ_0 , while the second bound comes from (82) at $z = \tau_t + \delta/2$ and $\Im m_{fc}^t(\tau_t + \delta/2) = 0$. We thus conclude $\delta \lesssim t/N$. If on the contrary $\tau_0 = \tau_t - \delta$ for some $\delta > 0$, then with a similar argument $\sqrt{\delta} \lesssim \Im m_{fc}^t(\tau_0 + \delta/2) \lesssim t/N$ and we have $\delta \lesssim t/N$ also in this case and the claim follows.

Claim 2. $|\gamma_t - \gamma| \lesssim (t/N)^{1/4}$, where $\gamma = \gamma_{\text{edge}}$ from Definition 2.4. From the third equality in (81a), we can relate the edge-slope of m_{fc}^t to M_t'' . Indeed, if $\gamma_t^{3/2}$ denotes the slope, that is, $\varrho_{fc}^t(x) = \gamma_t^{3/2}\sqrt{(\tau_t - x)_+}/\pi + \mathcal{O}(\tau_t - x)$, then using the elementary integrals

$$\begin{aligned} \lim_{\eta \rightarrow 0} \eta^{1/2} \int_0^\infty \frac{\sqrt{x}/\pi}{(x - i\eta)^2} dx &= \frac{i^{1/2}}{2}, \\ \lim_{\eta \rightarrow 0} \eta^{3/2} \int_0^\infty \frac{\sqrt{x}/\pi}{(x - i\eta)^3} dx &= \frac{i^{3/2}}{8} \end{aligned}$$

we obtain the precise divergence asymptotics of the derivatives $(m_{fc}^t)'(z)$ and $(m_{fc}^t)''(z)$ as $z = \tau_t + i\eta \rightarrow \tau_t$ and conclude

$$\frac{2}{\gamma_t^3} = \lim_{z \rightarrow \tau_t} \frac{(ct)^3(m_{fc}^t)''(z)}{(1 + ct(m_{fc}^t)'(z))^3} = (ct)^3 \langle M_t''(\xi_t(\tau_t)) \rangle,$$

that is,

$$\gamma_t = \frac{(\langle M_t''(\xi_t(\tau_t)) \rangle / 2)^{-1/3}}{ct}.$$

We now use (82) at, say, $z = x := \tau_0 - \sqrt{t/N}$. By Claim 1, we have $\tau_t - x \sim \sqrt{t/N}$, and thus

$$\begin{aligned} \gamma_t^{3/2} &= \frac{\Im m_{fc}^t(x)}{\sqrt{\tau_t - x}} + \mathcal{O}((t/N)^{1/4}) = \frac{\Im \langle M(x) \rangle}{\sqrt{\tau_t - x}} + \mathcal{O}((t/N)^{1/4}) \\ &= \frac{\Im \langle M(x) \rangle}{\sqrt{\tau_0 - x}} + \mathcal{O}((t/N)^{1/4}) = \gamma^{3/2} + \mathcal{O}((t/N)^{1/4}), \end{aligned}$$

where we used Claim 1 again in the third equality. This completes the proof of the claim.

Claim 3. $|\tilde{\tau}_t - \tau_t| < 1/Nt$. Since M_t has a square root edge at some $\hat{\tau}_t$, it follows from the first equality in (81b) that $\xi_t(\tau_t) - \hat{\tau}_t \sim t^2$. Using rigidity in the form of Corollary 2.9 for the matrix \tilde{H}_t to estimate \tilde{G}'_t from below at a spectral parameter outside of the support, we have the bound

$$ct = |(\tilde{G}'_t(\tilde{\xi}_t(\tilde{\tau}_t)))|^{-1} < |\tilde{\xi}_t(\tilde{\tau}_t) - \hat{\tau}_t|^{1/2}.$$

Consequently, using the local law in the form of Lemma A.1 it follows that

$$|(\tilde{M}'_t(\tilde{\xi}_t(\tilde{\tau}_t)))| = 1/ct + \mathcal{O}_{<}(1/Nt^4) \sim 1/t,$$

whence $\tilde{\xi}_t(\tilde{\tau}_t) - \hat{\tau}_t \sim t^2$ where we again used the square root singularity of $\langle M_t \rangle$ at $\hat{\tau}_t$. We can conclude, starting from (81b), that

$$\begin{aligned} 0 &= \langle M'_t(\xi_t(\tau_t)) \rangle - \langle \tilde{G}'_t(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle = \langle M'_t(\xi_t(\tau_t)) \rangle - \langle M'_t(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle \\ &\quad + \langle (M'_t - \tilde{G}'_t)(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle \sim |\xi_t(\tau_t) - \tilde{\xi}_t(\tilde{\tau}_t)|/t^3 + \mathcal{O}_{<}(1/Nt^4), \end{aligned}$$

where we used that $|\langle M'_t(\hat{\tau}_t + rt^2) \rangle| \sim t^{-3}$ for $c < r < C$ and the improved local law $\langle G' - M' \rangle < 1/N\kappa^2$ at a distance $\kappa \sim t^2$ away from the spectrum, as stated in Lemma A.1. We thus find that $|\xi_t(\tau_t) - \tilde{\xi}_t(\tilde{\tau}_t)| < 1/Nt$. It remains to relate this to an estimate on $|\tau_t - \tilde{\tau}_t|$. We have

$$|\tau_t - \tilde{\tau}_t| \lesssim |\xi_t(\tau_t) - \tilde{\xi}_t(\tilde{\tau}_t)| + t|m_{fc}^t(\tau_t) - m_{fc}^t(\tilde{\tau}_t)| + t|(m_{fc}^t - \tilde{m}_{fc}^t)(\tilde{\tau}_t)|,$$

where we bounded the second term by $t|\langle M_t(\xi_t(\tau_t)) - M_t(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle| < 1/Nt$ using $|\langle M'_t(\hat{\tau}_t + rt^2) \rangle| \sim 1/t$ and the third term by $t|\langle (M_t - \tilde{G}_t)(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle| < 1/Nt$ using the local law t^2 away from $\text{supp } \varrho_t$. Thus we can conclude that $|\tau_t - \tilde{\tau}_t| < 1/Nt$.

Claim 4. $|\gamma_t - \tilde{\gamma}_t| < 1/Nt^3$. We first note that $\gamma_t \sim 1$ follows from $|\langle M''_t(\xi_t(\tau_t)) \rangle| \sim t^{-3}$. Therefore, it suffices to estimate

$$\begin{aligned} t^3|\langle M''_t(\xi_t(\tau_t)) - \tilde{G}''_t(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle| &\leq t^3|\langle M''_t(\xi_t(\tau_t)) - M''_t(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle| \\ &\quad + t^3|\langle M''_t(\tilde{\xi}_t(\tilde{\tau}_t)) - \tilde{G}''_t(\tilde{\xi}_t(\tilde{\tau}_t)) \rangle| < \frac{1}{Nt^3}, \end{aligned}$$

as follows from $\langle M''_t(\hat{\tau}_t + rt^2) \rangle \sim t^{-5}$ for $c < r < C$ and the local law from Lemma A.1 at a distance of $\kappa \sim t^2$ away from the spectrum. Thus we have $|\gamma_t - \tilde{\gamma}_t| < 1/Nt^3$.

We now check that \tilde{H}_t is η_* -regular in the sense of [35], Definition 2.1, for $\eta_* := N^{-2/3+\epsilon}$. It follows from the local law that $c\varrho_t(z) < \Im \langle \tilde{G}_t(z) \rangle < C\varrho_t(z)$ for some constants c, C , whenever $\Im z \geq \eta_*$. Now (2.4)–(2.5) in [35] follow in high probability from the assumption that ϱ_t has a regular edge at τ_t . Furthermore, the absence of eigenvalues in the interval $[\tau_t + \eta_*, \tau_t + c/2]$ with high probability follows directly from Corollary 2.7. Finally, $\|\tilde{H}_t\| \leq N$ with high probability follows directly from $\|\tilde{H}_t\| \leq (\text{Tr}|\tilde{H}_t|^2)^{1/2}$. We can thus conclude that with high probability, \tilde{H}_t is $\eta_* = N^{-2/3+\epsilon}$ regular for any positive $\epsilon > 0$.

We denote the eigenvalues of $H_t = \tilde{H}_t + c\sqrt{t}U$ by $\lambda_1^t \leq \dots \leq \lambda_N^t$. Then it follows from [35], Theorem 2.2, that for $N^{-\epsilon} \geq t \geq N^{-2/3+\epsilon}$ with high probability for test functions $F: \mathbb{R}^{k+1} \rightarrow \mathbb{R}$ with $\|F\|_\infty + \|\nabla F\|_\infty \lesssim 1$ there exists some $c > 0$ such that

$$(83) \quad \begin{aligned} & \left| \mathbf{E}[F(\tilde{\gamma}_t N^{2/3}(\lambda_{i_0}^t - \tilde{\tau}_t), \dots, \tilde{\gamma}_t N^{2/3}(\lambda_{i_0-k}^t - \tilde{\tau}_t)) | \tilde{H}_t] \right. \\ & \quad \left. - \mathbf{E}F(N^{2/3}(\mu_N - 2), \dots, N^{2/3}(\mu_{N-k} - 2)) \right| \leq N^{-c}. \end{aligned}$$

By combining (83) with $|\tau_0 - \tilde{\tau}_t| < N^{-2/3-\epsilon}$, $|\gamma - \tilde{\gamma}_t| < N^{-\epsilon}$ from Claims 1–4, we obtain

$$(84) \quad \begin{aligned} & \left| \mathbf{E}[F(\gamma N^{2/3}(\lambda_{i_0}^t - \tau_0), \dots, \gamma N^{2/3}(\lambda_{i_0-k}^t - \tau_0))] \right. \\ & \quad \left. - \mathbf{E}F(N^{2/3}(\mu_N - 2), \dots, N^{2/3}(\mu_{N-k} - 2)) \right| \lesssim N^{-c} + N^{-\epsilon} \end{aligned}$$

for our choice of $t = N^{-1/3+\epsilon}$.

6.2. *Green’s function comparison.* It remains to prove that the local correlation functions of H_t agree with those of H . We want to prove that for any fixed $x_i \in \mathbb{R}$,

$$\lim_{N \rightarrow \infty} \mathbf{P}(N^{2/3}(\lambda_{i_0-i}^t - \tau_0) \geq x_i, i = 0, \dots, k)$$

is independent of t as long as, say, $t \leq N^{-1/3+\epsilon}$. We first note that the local law holds uniformly in t also for H_t . This follows easily from the fact that the assumptions stay uniformly satisfied along the flow because expectation and covariance are preserved while higher order cumulants also remain unchanged up to a multiplication with a t -dependent constant. For $l = N^{-2/3-\epsilon/3}$, $\eta = N^{-2/3-\epsilon}$ and smooth monotonous cut-off functions K_i with $K_i(x) = 0$ for $x \leq i - 1$ and $K_i(x) = 1$ for $x \geq i$ we have

$$(85) \quad \begin{aligned} & \mathbf{E} \prod_{i=0}^k K_{i_0-i} \left(\frac{\Im}{\pi} \int_{x_i N^{-2/3+l}}^{N^{-2/3+\epsilon}} \text{Tr } G_t(x + \tau + i\eta) dx \right) - \mathcal{O}(N^{-\epsilon/9}) \\ & \leq \mathbf{P}(N^{2/3}(\lambda_{i_0-i}^t - \tau) \geq x_i, i = 0, \dots, k) \\ & \leq \mathbf{E} \prod_{i=0}^k K_{i_0-i} \left(\frac{\Im}{\pi} \int_{x_i N^{-2/3-l}}^{N^{-2/3+\epsilon}} \text{Tr } G_t(x + \tau + i\eta) dx \right) + \mathcal{O}(N^{-\epsilon/9}). \end{aligned}$$

We note that the strategy of expressing k -point correlation functions of edge-eigenvalues through a regularized expression involving the resolvent was already used in [28, 30, 32, 37] for proving edge universality. The precise formula (85) has been already used, for example, in [30], equation (4.8).

In order to compare the expectations in (85) at times $t = 0$ and $t = N^{-1/3+\epsilon}$, we claim that we have the bound

$$(86) \quad \begin{aligned} & X_y := \Im \int_{y N^{-2/3 \pm l}}^{N^{-2/3+\epsilon}} \text{Tr } G_t(\tau + x + i\eta) dx, \\ & \left| \mathbf{E}g(X_{x_0}, \dots, X_{x_k}) \frac{dX_{x_j}}{d} \right| \lesssim N^{1/6+3\epsilon} \end{aligned}$$

for any $0 \leq j \leq k$ and smooth function g . Assuming (86), it follows for the smooth functions K_j and by Taylor’s expansion that that for $t \lesssim N^{-1/3+\epsilon}$,

$$\begin{aligned} & \left| \mathbf{E} \prod_{i=0}^k K_{i_0-i} \left(\frac{\Im}{\pi} \int_{x_i N^{-2/3 \pm l}}^{N^{-2/3+\epsilon}} \text{Tr } G_t(x + \tau + i\eta) dx \right) \right. \\ & \quad \left. - \mathbf{E} \prod_{i=0}^k K_{i_0-i} \left(\frac{\Im}{\pi} \int_{x_i N^{-2/3 \pm l}}^{N^{-2/3+\epsilon}} \text{Tr } G_0(x + \tau + i\eta) dx \right) \right| \lesssim \frac{1}{N^{1/6-4\epsilon}}. \end{aligned}$$

Together with (85), we obtain for any k, x_i

$$(87) \quad \begin{aligned} & \mathbf{P}(N^{2/3}(\lambda_{i_0-i}^t - \tau) \geq x_i, i = 0, \dots, k) \\ &= \mathbf{P}(N^{2/3}(\lambda_{i_0-i}^0 - \tau) \geq x_i, i = 0, \dots, k) + \mathcal{O}(N^{-\epsilon/9}). \end{aligned}$$

Equation (86) for $g \equiv 1$ follows from Itô’s lemma in the form

$$\mathbf{E} \frac{df(H)}{dt} = \mathbf{E} \left[-\frac{1}{2} \sum_{\alpha} w_{\alpha} \overline{(\partial_{\alpha} f)}(H) + \frac{1}{2} \sum_{\alpha, \beta} \kappa(\alpha, \beta) (\partial_{\alpha} \partial_{\beta} f)(H) \right]$$

and the general neighborhood cumulant expansion involving *precumulants*, as introduced in [21], Proposition 3.5. This expansion formula was a key input to the Green’s function comparison argument in the spectral bulk in [21], Corollary 2.6, for correlated matrix models under Assumptions (CD). Given the local law, Theorem 2.6, the extension of this proof to the edge is a routine power counting argument even for $g \neq 1$ and is left to the reader.

PROOF OF THEOREM 2.11. The theorem follows directly from (84) and (87). \square

APPENDIX: AUXILIARY RESULTS

PROOF OF LEMMA 3.4. From (70a)–(70b) in [21], we have⁶

$$(88a) \quad \|MS[R]R\|_* \lesssim N^{1/2K} \|R\|_*^2, \quad \|MR\|_* \lesssim N^{1/2K} \|R\|_*$$

and furthermore by a three term geometric expansion also

$$(88b) \quad \begin{aligned} \|\mathcal{B}^{-1}Q\|_{* \rightarrow *} &\leq (1 + \|Q\|_{* \rightarrow *}) (1 + \|C_M \mathcal{S}\|_{* \rightarrow *}) \\ &\quad + \|C_M \mathcal{S}\|_{* \rightarrow \text{hs}} \|\mathcal{B}^{-1}Q\|_{\text{sp}} \|C_M \mathcal{S}\|_{\text{hs} \rightarrow *}. \end{aligned}$$

Since

$$\begin{aligned} \|\mathcal{P}[R]\|_* &= \frac{|\langle P, R \rangle|}{|\langle P, B \rangle|} \|B\|_* \leq \frac{\|B\|}{|\langle P, B \rangle| N} \sum_a |R_{P_a^* a}| \\ &\leq \frac{\|B\| \|R\|_*}{|\langle P, B \rangle| N} \sum_a \|P_a^*\| \leq \frac{\|P\| \|B\|}{|\langle P, B \rangle|} \|R\|_* \end{aligned}$$

it follows that $\|\mathcal{P}\|_{* \rightarrow *} \lesssim 1$ and, therefore, also $\|Q\|_{* \rightarrow *} \lesssim 1$. Now, since $\|R\|_{\max} \leq \|R\|_* \leq \|R\|$ and according to (73) in [21] also $\max\{\|S\|_{\max \rightarrow \|\cdot\|}, \|S\|_{\text{hs} \rightarrow \|\cdot\|}\} \lesssim 1$, the lemma follows together with $\|\mathcal{B}^{-1}Q\|_{\text{sp}} \lesssim 1$ from Proposition 3.1(iv). \square

LEMMA A.1. Fix any $\epsilon, \delta > 0$ and an integer $k \geq 0$. Under the assumptions of Theorem 2.6, for the k th derivatives of M and G we have the bound

$$(89) \quad \left| \langle G^{(k)}(z) - M^{(k)}(z) \rangle \right| < \frac{1}{N \kappa^{k+1}},$$

uniformly in $z \in \mathbb{D}^{\delta}$ with $\kappa = \text{dist}(z, \text{supp } \varrho) \geq N^{-2/3+\epsilon}$.

PROOF. We will fix $z = x + i\eta$ throughout the proof. Let $\chi : \mathbb{R} \rightarrow \mathbb{R}$ be a smooth cut-off function such that $\chi(x') = 1$ for $\kappa' = \text{dist}(x', \text{supp } \varrho) \leq \kappa/3$ and $\chi(x') = 0$ for $\kappa' \geq 2\kappa/3$ and let $\tilde{\chi}$ be a cut-off function such that $\tilde{\chi}(\eta') = 1$ for $\eta' \leq 1$ and $\tilde{\chi}(\eta') = 0$ for $\eta' \geq 2$. We

⁶Cf. Remark 3.2 for the applicability of these bounds in the present setup.

also assume that the cut-off functions have bounded derivatives in the sense $\|\chi'\|_\infty \lesssim 1/\kappa$, $\|\chi''\|_\infty \lesssim 1/\kappa^2$ and $\|\tilde{\chi}'\|_\infty \lesssim 1$. We now define $f(x') := (x' - z)^{-k} \chi(x')$ and the almost analytic extension

$$f^{\mathbb{C}}(z') = f^{\mathbb{C}}(x' + i\eta') := \tilde{\chi}(\eta') [f(x') + i\eta' f'(x')],$$

$$\partial_{\bar{z}} f^{\mathbb{C}}(z') = \frac{i\eta'}{2} \tilde{\chi}(\eta') f''(x') + \frac{i}{2} \tilde{\chi}'(\eta') [f(x') + i\eta' f'(x')].$$

It follows from the Cauchy theorem and the absence of eigenvalues outside $\{\chi = 1\}$ in the sense of Corollary 2.7 that with high probability

$$\langle G^{(k)}(z) - M^{(k)}(z) \rangle = \frac{2}{\pi} \Re \int_{\mathbb{R}} \int_{\mathbb{R}_+} \partial_{\bar{z}} f^{\mathbb{C}}(z') \langle G(z') - M(z') \rangle d\eta' dx'.$$

Due to the fact that $\tilde{\chi}' = 0$ for $\eta' \leq 1$ the second term in $\partial_{\bar{z}} f^{\mathbb{C}}$ only gives a contribution of $1/N\kappa^{k+1}$ even by the local law and the $\|\cdot\|_\infty$ bound for $\partial_{\bar{z}} f^{\mathbb{C}}$ and we now concentrate on the first term. First, we exclude the integration regime $\eta' \lesssim N^{-1+\gamma}$ in which we cannot use the local law but only the trivial bound $\langle G - M \rangle \lesssim 1/\eta'$. For the contribution of this regime to (89), we thus have to estimate

$$N^{-1+\gamma} \int_{\mathbb{R}} |f''(x')| dx'$$

$$\lesssim \frac{1}{N} \int_{|x-x'| \geq 2\kappa/3} \left[\frac{1}{\kappa^2 |x-x'|^k} + \frac{1}{\kappa |x-x'|^{k+1}} + \frac{1}{|x-x'|^{k+2}} \right] dx'$$

$$\lesssim \frac{N^\gamma}{N\kappa^{k+1}}$$

and we have shown that

$$|\langle G^{(k)}(z) - M^{(k)}(z) \rangle|$$

$$< \frac{N^\gamma}{N\kappa^{k+1}}$$

$$+ \int_{\mathbb{R}} \int_{N^{-1+\gamma}}^2 \eta' \left[\frac{\chi(x')}{|x' - z|^{k+2}} + \frac{\chi'(x')}{|x' - z|^{k+1}} + \frac{\chi''(x')}{|x' - z|^k} \right]$$

$$\times |\langle G(z') - M(z') \rangle| d\eta' dx'.$$

We now use the local law of the form $|\langle G - M \rangle| < 1/N(\kappa + \eta')$ and that in the second and third term the integration regime is only of order κ to obtain the final bound of $N^\gamma/N\kappa^{k+1}$ for any $\gamma > 0$. \square

Acknowledgments. The first author was supported in part by ERC Advanced Grant No. 338804 and ERC Starting Grant No. 715539 and the SNSF SwissMap grant.

The second and fourth authors were supported in part by ERC Advanced Grant No. 338804.

The third author was supported in part by the Hausdorff Center for Mathematics.

REFERENCES

[1] ADHIKARI, A. and CHE, Z. (2019). Edge universality of correlated Gaussians. *Electron. J. Probab.* **24** Paper No. 44, 25. MR3949269 <https://doi.org/10.1214/19-EJP273>

[2] AJANKI, O. H., ERDŐS, L. and KRÜGER, T. (2015). Quadratic vector equations on complex upper half-plane. *Mem. Amer. Math. Soc.* To appear. Available at [arXiv:1506.05095](https://arxiv.org/abs/1506.05095).

- [3] AJANKI, O. H., ERDŐS, L. and KRÜGER, T. (2017). Universality for general Wigner-type matrices. *Probab. Theory Related Fields* **169** 667–727. MR3719056 <https://doi.org/10.1007/s00440-016-0740-2>
- [4] AJANKI, O. H., ERDŐS, L. and KRÜGER, T. (2019). Stability of the matrix Dyson equation and random matrices with correlations. *Probab. Theory Related Fields* **173** 293–373. MR3916109 <https://doi.org/10.1007/s00440-018-0835-z>
- [5] ALT, J., ERDŐS, L. and KRÜGER, T. (2018). The Dyson equation with linear self-energy: Spectral bands, edges and cusps. Preprint. Available at [arXiv:1804.07752v2](https://arxiv.org/abs/1804.07752v2).
- [6] ALT, J., ERDŐS, L., KRÜGER, T. and NEMISH, Y. (2019). Location of the spectrum of Kronecker random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 661–696. MR3949949 <https://doi.org/10.1214/18-aihps894>
- [7] AUFFINGER, A., BEN AROUS, G. and PÉCHÉ, S. (2009). Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 589–610. MR2548495 <https://doi.org/10.1214/08-AIHP188>
- [8] BAI, Z. D. and SILVERSTEIN, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Ann. Probab.* **27** 1536–1555. MR1733159 <https://doi.org/10.1214/aop/1022677458>
- [9] BEKERMANN, F. (2018). Transport maps for β -matrix models in the multi-cut regime. *Random Matrices Theory Appl.* **7** 1750013, 36. MR3756421 <https://doi.org/10.1142/S2010326317500137>
- [10] BHATIA, R. (1997). *Matrix Analysis. Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 <https://doi.org/10.1007/978-1-4612-0653-8>
- [11] BONNET, G., DAVID, F. and EYNARD, B. (2000). Breakdown of universality in multi-cut matrix models. *J. Phys. A* **33** 6739–6768. MR1790279 <https://doi.org/10.1088/0305-4470/33/38/307>
- [12] BOROT, G. and GUIONNET, A. (2013). Asymptotic expansion of beta matrix models in the multi-cut regime. Preprint. Available at [arXiv:1303.1045](https://arxiv.org/abs/1303.1045).
- [13] BOURGADE, P., ERDŐS, L. and YAU, H.-T. (2014). Edge universality of beta ensembles. *Comm. Math. Phys.* **332** 261–353. MR3253704 <https://doi.org/10.1007/s00220-014-2120-z>
- [14] CAPITAINE, M. and PÉCHÉ, S. (2016). Fluctuations at the edges of the spectrum of the full rank deformed GUE. *Probab. Theory Related Fields* **165** 117–161. MR3500269 <https://doi.org/10.1007/s00440-015-0628-6>
- [15] CHE, Z. (2017). Universality of random matrices with correlated entries. *Electron. J. Probab.* **22** Paper No. 30, 38. MR3629874 <https://doi.org/10.1214/17-EJP46>
- [16] CIPOLLONI, G., ERDŐS, L., KRÜGER, T. and SCHRÖDER, D. (2018). Cusp universality for random matrices II: The real symmetric case. Preprint. Available at [arXiv:1811.04055](https://arxiv.org/abs/1811.04055).
- [17] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2012). Spectral statistics of Erdős–Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** 587–640. MR2964770 <https://doi.org/10.1007/s00220-012-1527-7>
- [18] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41** 2279–2375. MR3098073 <https://doi.org/10.1214/11-AOP734>
- [19] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** no. 59, 58. MR3068390 <https://doi.org/10.1214/EJP.v18-2473>
- [20] ERDŐS, L., KRÜGER, T. and SCHRÖDER, D. (2018). Cusp universality for random matrices I: Local law and the complex Hermitian case. Preprint. Available at [arXiv:1809.03971](https://arxiv.org/abs/1809.03971).
- [21] ERDŐS, L., KRÜGER, T. and SCHRÖDER, D. (2019). Random matrices with slow correlation decay. *Forum Math. Sigma* **7** e8, 89. MR3941370 <https://doi.org/10.1017/fms.2019.2>
- [22] ERDŐS, L., PÉCHÉ, S., RAMÍREZ, J. A., SCHLEIN, B. and YAU, H.-T. (2010). Bulk universality for Wigner matrices. *Comm. Pure Appl. Math.* **63** 895–925. MR2662426 <https://doi.org/10.1002/cpa.20317>
- [23] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2011). Universality of random matrices and local relaxation flow. *Invent. Math.* **185** 75–119. MR2810797 <https://doi.org/10.1007/s00222-010-0302-7>
- [24] ERDŐS, L., SCHLEIN, B., YAU, H.-T. and YIN, J. (2012). The local relaxation flow approach to universality of the local statistics for random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1–46. MR2919197 <https://doi.org/10.1214/10-AIHP388>
- [25] ERDŐS, L. and SCHNELLI, K. (2017). Universality for random matrix flows with time-dependent density. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1606–1656. MR3729630 <https://doi.org/10.1214/16-AIHP765>
- [26] ERDŐS, L. and YAU, H.-T. (2015). Gap universality of generalized Wigner and β -ensembles. *J. Eur. Math. Soc. (JEMS)* **17** 1927–2036. MR3372074 <https://doi.org/10.4171/JEMS/548>

- [27] ERDŐS, L. and YAU, H.-T. (2017). *A Dynamical Approach to Random Matrix Theory. Courant Lecture Notes in Mathematics* **28**. Courant Institute of Mathematical Sciences, New York; Amer. Math. Soc., Providence, RI. MR3699468
- [28] ERDŐS, L., YAU, H.-T. and YIN, J. (2012). Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.* **229** 1435–1515. MR2871147 <https://doi.org/10.1016/j.aim.2011.12.010>
- [29] HELTON, J. W., RASHIDI FAR, R. and SPEICHER, R. (2007). Operator-valued semicircular elements: Solving a quadratic matrix equation with positivity constraints. *Int. Math. Res. Not. IMRN* **22** Art. ID rnm086, 15. MR2376207 <https://doi.org/10.1093/imrn/rnm086>
- [30] HUANG, J., LANDON, B. and YAU, H.-T. (2017). Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs. Preprint. Available at [arXiv:1712.03936](https://arxiv.org/abs/1712.03936).
- [31] JOHANSSON, K. (2007). From Gumbel to Tracy–Widom. *Probab. Theory Related Fields* **138** 75–112. MR2288065 <https://doi.org/10.1007/s00440-006-0012-7>
- [32] KNOWLES, A. and YIN, J. (2013). Eigenvector distribution of Wigner matrices. *Probab. Theory Related Fields* **155** 543–582. MR3034787 <https://doi.org/10.1007/s00440-011-0407-y>
- [33] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 <https://doi.org/10.1007/s00440-016-0730-4>
- [34] LANDON, B., SOSOE, P. and YAU, H.-T. (2019). Fixed energy universality of Dyson Brownian motion. *Adv. Math.* **346** 1137–1332. MR3914908 <https://doi.org/10.1016/j.aim.2019.02.010>
- [35] LANDON, B. and YAU, H.-T. (2017). Edge statistics of Dyson Brownian motion. Preprint. Available at [arXiv:1712.03881](https://arxiv.org/abs/1712.03881).
- [36] LANDON, B. and YAU, H.-T. (2017). Convergence of local statistics of Dyson Brownian motion. *Comm. Math. Phys.* **355** 949–1000. MR3687212 <https://doi.org/10.1007/s00220-017-2955-1>
- [37] LEE, J. O. and SCHNELLI, K. (2015). Edge universality for deformed Wigner matrices. *Rev. Math. Phys.* **27** 1550018, 94. MR3405746 <https://doi.org/10.1142/S0129055X1550018X>
- [38] LEE, J. O. and SCHNELLI, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** 3786–3839. MR3582818 <https://doi.org/10.1214/16-AAP1193>
- [39] LEE, J. O. and SCHNELLI, K. (2018). Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** 543–616. MR3800840 <https://doi.org/10.1007/s00440-017-0787-8>
- [40] LEE, J. O. and YIN, J. (2014). A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.* **163** 117–173. MR3161313 <https://doi.org/10.1215/00127094-2414767>
- [41] MEHTA, M. L. (1967). *Random Matrices and the Statistical Theory of Energy Levels*. Academic Press, New York. MR0220494
- [42] PASTUR, L. (2006). Limiting laws of linear eigenvalue statistics for Hermitian matrix models. *J. Math. Phys.* **47** 103303, 22. MR2268864 <https://doi.org/10.1063/1.2356796>
- [43] PASTUR, L. and SHCHERBINA, M. (2003). On the edge universality of the local eigenvalue statistics of matrix models. *Zh. Mat. Fiz. Anal. Geom.* **10** 335–365. MR2012268
- [44] SHCHERBINA, M. (2013). Fluctuations of linear eigenvalue statistics of β matrix models in the multi-cut regime. *J. Stat. Phys.* **151** 1004–1034. MR3063494 <https://doi.org/10.1007/s10955-013-0740-x>
- [45] SHCHERBINA, T. (2011). On universality of local edge regime for the deformed Gaussian unitary ensemble. *J. Stat. Phys.* **143** 455–481. MR2799948 <https://doi.org/10.1007/s10955-011-0196-9>
- [46] SOSHNIKOV, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** 697–733. MR1727234 <https://doi.org/10.1007/s002200050743>
- [47] TAO, T. and VU, V. (2010). Random matrices: Universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* **298** 549–572. MR2669449 <https://doi.org/10.1007/s00220-010-1044-5>
- [48] TAO, T. and VU, V. (2011). Random matrices: Universality of local eigenvalue statistics. *Acta Math.* **206** 127–204. MR2784665 <https://doi.org/10.1007/s11511-011-0061-3>
- [49] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246
- [50] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083