Commun. Math. Phys. (2025) 406:253
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-025-05417-z

Communications in
Mathematical Physics

Cusp Universality for Correlated Random Matrices

László Erdős, Joscha Henheik, Volodymyr Riabov

Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria.
E-mail: lerdos@ist.ac.at

Received: 3 November 2024 / Accepted: 14 July 2025
© The Author(s) 2025

Abstract: For correlated real symmetric or complex Hermitian random matrices, we prove that the local eigenvalue statistics at any cusp singularity are universal. Since the density of states typically exhibits only square root edge or cubic root cusp singularities, our result completes the proof of the Wigner–Dyson–Mehta universality conjecture in all spectral regimes for a very general class of random matrices. Previously only the bulk and the edge universality were established in this generality (Alt et al. in Ann Probab 48(2):963–1001, 2020), while cusp universality was proven only for Wigner-type matrices with independent entries (Cipolloni et al. in Pure Appl Anal 1:615–707, 2019; Erdős et al. in Commun. Math. Phys. 378:1203–1278, 2018). As our main technical input, we prove an optimal local law at the cusp using the *Zigzag strategy*, a recursive tandem of the characteristic flow method and a Green function comparison argument. Moreover, our proof of the optimal local law holds uniformly in the spectrum, thus we also provide a significantly simplified alternative proof of the local eigenvalue universality in the previously studied bulk (Erdős et al. in Forum Math. Sigma 7:E8, 2019) and edge (Alt et al. in Ann Probab 48(2):963–1001, 2020) regimes.

1. Introduction

The celebrated Wigner–Dyson–Mehta (WDM) conjecture asserts that the local eigenvalue statistics of large random matrices become *universal*: they depend only on the symmetry class of the matrix and not on the precise details of its distribution. This remarkable effect is extremely robust and manifests in all spectral regimes. The correlation functions of the eigenvalues are governed by one of three universal determinantal processes, whose kernel functions depend on the local shape of the eigenvalue density. As proven by Dyson, Gaudin and Mehta [54] for the Gaussian GOE/GUE ensembles, the local statistics of the eigenvalues in the *bulk* of the spectrum are driven by the *sine*

Supported by the ERC Advanced Grant "RMTBeyond" No. 101020331.

Published online: 01 September 2025

kernel. At the spectral edges, where the density of states vanishes like a square root, Tracy and Widom [65,66] computed that the correlation functions for GOE/GUE are given by the *Airy kernel*. As was first observed by Wigner [69], and formalized as a conjecture for standard Wigner matrices by Dyson and Mehta in the 1960s, these statistics hold well beyond the Gaussian ensembles. After the first proofs for standard Wigner matrices [19,38,40,61,63,64], these universality results in the bulk and at the edge saw rapid development and were gradually extended[1] to ensembles of ever greater generality: for Wigner matrices with diagonal [51,53] and non-diagonal deformations [47], Wigner-type ensembles with not necessarily identically distributed but still independent entries [7], and even to random matrices allowing for substantial correlations among the entries [9,11,35].

The third and final class of universal local statistics emerges at the *cusp-like* singularities of the density with cubic-root behavior. There, the eigenvalues form a *Pearcey process*, which was first identified by Brézin and Hikami for a Gaussian unitary (GUE) matrix with a special deterministic deformation [22,23]. Compared to the bulk and edge, the cusp regime is less understood and universality in this most delicate spectral regime was established only recently in [32,36], however only for a special class of random matrices. More precisely, these proofs were restricted to Wigner-type ensembles with independent entries and diagonal deformations, and did not cover the broadest class of correlated ensembles, for which bulk and edge universality had already been proven.

Our main result completes the picture by proving the universality of the local eigenvalues statistics at the cusp for random matrices with correlated entries and an arbitrary deformation, as stated in our main result, Theorem 2.13. The proof follows the *three-step strategy*, a general method for proving universality of local spectral statistics, summarized in [41]. The first step in this strategy is the *local law*, which asserts that the resolvent $G(z) = (H - z)^{-1}$ at $z = E + \mathrm{i}\eta \in \mathbb{H}$ of the random matrix H concentrates around a deterministic matrix $M(z)$ as the dimension of the matrix tends to infinity. This concentration estimate holds for η just above the local eigenvalue spacing at E, resolving the empirical distribution of eigenvalues at this scale. The second step is to establish universality for ensembles with a tiny Gaussian component, and the third step is a perturbative argument that removes the Gaussian component. Crucially, the optimal local law is used as a key input for both the second and third steps. These latter two steps have proven to be extremely robust and essentially model-independent tools [11,32,35,36]. Nevertheless, the critical first step, the proof of the *local law*, remains highly model-dependent.

As our main technical result, Theorem 2.8, we prove the *optimal average and isotropic local laws* for correlated random matrices. These local laws assert that for any fixed $\xi > 0$, any deterministic matrix B and test vectors $\boldsymbol{x}$, $\boldsymbol{y}$, the bounds

$$\left|\left\langle\big(G(z) - M(z)\big)B\right\rangle\right| \lesssim N^{\xi}\,\frac{\|B\|_{\mathrm{hs}}}{N\eta} \quad \text{and} \quad \left|\big(G(z) - M(z)\big)_{\boldsymbol{x}\boldsymbol{y}}\right| \lesssim N^{\xi}\sqrt{\frac{\rho(z)}{N\eta}}\,\|\boldsymbol{x}\|\,\|\boldsymbol{y}\| \tag{1.1}$$

hold with very high probability. Here N is the dimension of the random matrix H, $\langle\cdot\rangle := N^{-1}\mathrm{Tr}[\cdot]$ denotes the normalized trace, and $\rho(z) := \pi^{-1}\langle\Im M(z)\rangle > 0$ is the *self consistent density of states*. Moreover, Theorem 2.8 provides further optimal improvements to the right-hand sides of (1.1) for spectral parameters $z = E + \mathrm{i}\eta$ with energy E

[1] In another direction of generalization, sparse matrices [4,37,45,52], adjacency matrices of regular graphs [14], band matrices [20,21,60], and dynamically defined matrices [3] have also been considered. In parallel to that, universal statistics in the bulk and at the edge have been established for invariant β-ensembles (see, e.g., [12,15,18,19,33,34,48,55–58,68]) and their discrete analogs [13,16,42,46], although often using very different methods.

outside of the self-consistent spectrum. We point out that the local laws in (1.1) are optimal in terms of their dependence on $\rho(z)$ and the (normalized) Hilbert–Schmidt norm $\|B\|_{\mathrm{hs}} := \langle BB^*\rangle^{1/2}$ of the observable matrix B. In many cases, such as for low-rank observables, $\|B\|_{\mathrm{hs}}$ is much smaller than the operator norm $\|B\|$, which has traditionally been used in previous single-resolvent local laws [11,35,36]. Thus, our local law (1.1) unifies and improves upon the previous local laws, even in the Wigner-type case.

Traditional proofs of the local laws relied on solving an approximate self-consistent equation for the difference $G - M$. They consisted of two parts: a stability analysis of the underlying deterministic *Dyson equation* and a probabilistic estimate on the fluctuations. Both steps become quite cumbersome beyond the simple Wigner matrices. In particular, for general Wigner-type [7,36] and correlated random matrices [11,35], the stability analysis became intricate [8,10], and the probabilistic part relied on sophisticated Feynman graph expansions. Recently, a completely new approach, the *Zigzag strategy* [24,27–29,31,39], has been developed. This approach consists of an iterated application of two steps in tandem (cf. Figure 3 below): the *characteristic flow method* [1,2,6,17,44,49,50], coined the *zig-step*, and a Green function comparison (GFT) argument driven by an Ornstein-Uhlenbeck flow, called the *zag-step*. Remarkably, the Zigzag strategy circumvents many of the difficulties that arise along the more traditional local law proofs. It even removes the key obstacles that previously hindered the proof of the optimal local law at the cusp for the most general correlated matrices. We now explain this crucial aspect in more detail.

For traditional proofs of the local laws, the bulk regime is the easiest since the underlying Dyson equation is stable when $\rho(z)$ is separated away from zero. In the regime where the density $\rho(z)$ vanishes, this stability deteriorates – specifically, the corresponding stability factor behaves like $\rho(z)^{-1}$ at a square-root edge and as $\rho(z)^{-2}$ at a cubic-root cusp. This blow-up had to be compensated by a fine control on the error term in the approximate Dyson equation. On the probabilistic side, obtaining the optimal very-high-probability estimate on the fluctuation error required a high moment calculation that exploited various *fluctuation averaging* mechanisms, even in the simplest bulk regime. In the edge regime, an additional factor $\rho(z)$ needed to be extracted, which essentially relied on the emergence of the imaginary part of the resolvent via the *Ward identity*, $GG^* = \Im G/\eta$. However, for cusp singularities, an additional *second order* cancellation effect was necessary. This delicate effect, coined the *cusp fluctuation averaging* [36], arises from a finite set of critical Feynman subdiagrams, called the σ*-cells*. Roughly speaking, a σ-cell consists of four resolvents interconnected through the deterministic approximation M and the correlation four-tensor of the matrix elements. In the case of Wigner-type matrices with diagonal deformations, M becomes a *diagonal* matrix, leading to a simplification of the original *matrix* Dyson equation into a *vector* equation. Moreover, since the entries of a Wigner-type matrix are independent, the correlation tensor is reduced to a matrix acting on the diagonal. These substantial simplifications facilitated the intricate extraction of σ-cells, effectively capturing the second order cancellation effect. Identifying the analog of the σ-cells for correlated matrices, when M is no longer diagonal and the correlation is a full-fledged four-tensor remains out of reach.

In this paper, we leverage the *Zigzag strategy* to conveniently avoid the complicated graphical expansions and, more importantly, circumvent the extraction of σ-cells. The only stability input required is a trivial bound of the form $\rho(z)/\eta$, that is precisely tracked by the Ward identity. The characteristic flow at the heart of the Zigzag strategy has previously proved itself to be effective in dealing with a *first order* blow-up of the stability factor, such as at the edge of Wigner matrices [27], and in capturing the $z_1 - z_2$

decorrelation effect for the Hermitizations of non-Hermitian i.i.d. matrices [30,31]. The current work demonstrates that the Zigzag strategy is even capable of circumnavigating general *second order* instabilities arising at the cusp. Evidence of this feature of the characteristic flow has already been observed for unitary Brownian motion [3] and in a special non-Hermitian setting [24], where an additional symmetry was available.

Besides unraveling this remarkable power of the Zigzag approach in full generality, our paper is the first to implement the method in a correlated setting, which requires adjustments to the Zigzag dynamics. The GFT argument at the core of the zag step requires an a-priori bound on the resolvent as an input, which typically stems from a single resolvent local law. This, however, would render our argument circular. Hence, to remedy the situation, we augment the zag step with an internal induction[2] (*bootstrap*) in η. Furthermore, our result has two additional features: (i) for the averaged law in (1.1), we obtain the optimal estimate on the observable B in terms of its Hilbert–Schmidt norm, and (ii) we extend the Zigzag approach beyond the typical *above the scale* regime of $N\eta\rho(z) \geq N^{\varepsilon}$ (see Sect. 6). We emphasize that, in addition to covering the missing cusp regime, our proof also provides a unified approach to optimal local laws for the most general class of random matrices with correlated entries, completely eliminating any dependence of the proof on the specific spectral regime. The price we pay for our simple and self-contained Zigzag proof of the local law is assuming *fullness* of the correlated random matrix (cf. Assumption 2.4), rather than the slightly weaker *flatness* condition (cf. [35, Assumption (E)]). However, this stronger assumption is justified because fullness is necessary for deducing universality using the three-step strategy, regardless of how the local law is proven.

Notations and conventions. We use the notation $[N]$ to represent the index set $\{1, \ldots, N\}$. The letters a, b, j, and k are used to denote integer indices, while α (with various subscripts) denotes elements of $[N]^2$. All unrestricted summations of the form $\sum_a$ and $\sum_\alpha$ are understood to run over $a \in [N]$ and $\alpha \in [N]^2$, respectively.

We denote vectors in $\mathbb{C}^{N\times N}$ using boldface letters, e.g., $\boldsymbol{x}$. The scalar product on $\mathbb{C}^N$ is defined by $\langle \boldsymbol{x}, \boldsymbol{y}\rangle := \sum_{j=1}^N \overline{x_j} y_j$, and the corresponding Euclidean norm is denoted by $\|\boldsymbol{x}\| := \langle \boldsymbol{x}, \boldsymbol{x}\rangle^{1/2}$.

Matrices are denoted by capital letters. Unless explicitly stated otherwise, all matrices we consider are $N \times N$. For a matrix $A \in \mathbb{C}^{N\times N}$, the angle brackets $\langle A\rangle := N^{-1}\mathrm{Tr}[A]$ denote its normalized trace. We use the following notations for the matrix norms:

$$\|A\|_{\max} := \max_{a,b} |A_{ab}|, \quad \|A\| := \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|, \quad \|A\|_{\mathrm{hs}} := \left\langle |A|^2 \right\rangle^{1/2},$$

where $|A|^2 := AA^*$. Furthermore, for any $a \in [N]$ and vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we use the following notation:

$$A_{\boldsymbol{x}\boldsymbol{y}} := \langle \boldsymbol{x}, A\boldsymbol{y}\rangle, \quad A_{\boldsymbol{x}a} := \langle \boldsymbol{x}, A\boldsymbol{e}_a\rangle, \quad A_{a\boldsymbol{y}} := \langle \boldsymbol{e}_a, A\boldsymbol{y}\rangle,$$

where $\boldsymbol{e}_a$ is the standard a-th basis vector of $\mathbb{C}^N$.

We denote the complex upper half-plane by $\mathbb{H}$, that is, $\mathbb{H} := \{z \in \mathbb{C} : \Im z > 0\}$, and its closure by $\overline{\mathbb{H}} := \mathbb{H} \cup \mathbb{R}$. For a complex number $z \in \mathbb{C}$, we use the notation $\langle z\rangle := 1 + |z|$.

[2] This argument is reminiscent of [47] and we also refer to [62] for an alternative approach.

We use c and C to denote unspecified, positive constants-small and large, respectively-that are independent of N and may change from line to line. Various tolerance exponents are denoted by Greek letters such as $\varepsilon, \xi, \delta, \zeta, \mu, \nu$. The notation $\xi \ll \varepsilon$ means that there exists a small absolute constant $c > 0$ such that $\xi \le c\varepsilon$. We use $\nu > 0$ to denote arbitrary small tolerance exponents.

For two positive quantities $\mathcal{X}$ and $\mathcal{Y}$, we write $\mathcal{X} \lesssim \mathcal{Y}$ if there exists a constant $C > 0$ that depends only on the *model parameters* in Assumptions 2.1–2.5 (unless explicitly stated otherwise), such that $\mathcal{X} \le C\mathcal{Y}$. We use the notation $\mathcal{X} \sim \mathcal{Y}$ if both $\mathcal{X} \lesssim \mathcal{Y}$ and $\mathcal{Y} \lesssim \mathcal{X}$ hold. For an arbitrary quantity $\mathcal{X}$ and a positive quantity $\mathcal{Y}$, we use the notation $\mathcal{X} = \mathcal{O}(\mathcal{Y})$ to indicate that $|\mathcal{X}| \lesssim \mathcal{Y}$.

Let $\Omega := \{\Omega^{(N)}(u) \mid N \in \mathbb{N},\ u \in \mathcal{U}^{(N)}\}$ be a family of events depending on N and possibly on a parameter u that varies over some parameter set $\mathcal{U}^{(N)}$. We say that Ω holds *with very high probability* (w.v.h.p.) uniformly in $u \in \mathcal{U}^{(N)}$ if, for any $D > 0$,

$$\sup_{u \in \mathcal{U}^{(N)}} \mathbb{P}\big[\Omega^{(N)}(u)\big] \ge 1 - N^{-D},$$

for any $N \ge N_0(D)$. We often discard the explicit dependence of $\Omega^{(N)}$ and $\mathcal{U}^{(N)}$ on N, and simply refer to Ω as a very-high-probability event. A bound is said to hold w.v.h.p. if it holds on a very-high-probability event.

2. Main Results

We consider real symmetric or complex Hermitian random matrices H of the form

$$H = A + W\,, \qquad \mathbb{E}W = 0\,, \tag{2.1}$$

where $A \in \mathbb{C}^{N\times N}$ is a bounded deterministic matrix (cf. Assumption 2.1 below) and W has sufficiently fast decaying correlations between its matrix elements (cf. Assumption 2.3 below).

For any random matrix H, we define the *self-energy operator* $\mathcal{S}_H$ corresponding to H by its action on any deterministic matrix $X \in \mathbb{C}^{N\times N}$,

$$\mathcal{S}_H[X] := \mathbb{E}\big[(H - \mathbb{E}H)X(H - \mathbb{E}H)\big]. \tag{2.2}$$

The Matrix Dyson Equation (MDE) with a *data pair* $(A, \mathcal{S})$ is given by

$$-M(z)^{-1} = z - A + \mathcal{S}\big[M(z)\big] \tag{2.3}$$

for the unknown matrix valued function $M(z)$, $z \in \mathbb{C}\backslash\mathbb{R}$. It is well known (Theorem 2.1 [8]) that the MDE has a unique solution under the constraint that $(\Im z)\Im M(z) > 0$, where $\Im M = \frac{1}{2\mathrm{i}}(M - M^*)$. The corresponding *self-consistent density of states* (scDOS) ρ is a probability density function on the real line defined via the Stieltjes inversion formula,

$$\rho(x) := \lim_{\eta\to+0} \frac{1}{\pi}\big\langle \Im M(x + \mathrm{i}\eta)\big\rangle. \tag{2.4}$$

We define $\rho(z) := \pi^{-1}\langle \Im M(z)\rangle$ to be the harmonic extension of the scDOS to the complex upper-half plane. With a slight abuse of notation, we also refer to $\rho(z)$ as scDOS. As shown in [10], under suitable assumptions (which are formulated precisely in Sect. 2.1 below) on the data pair $(A, \mathcal{S})$ and the solution M of the MDE (2.3), the scDOS ρ is $1/3$-Hölder continuous. Furthermore, the set where the scDOS is positive,

$\{x \in \mathbb{R} : \rho(x) > 0\}$, splits into finitely many connected components, that are called *bands*. Inside the bands, the density is real-analytic with a square root growth behavior at the *edges*. If two bands touch, however, a cubic root *cusp* emerges. These are the only two possible types of singularities. Precise universal asymptotic formulas in the *almost cusp regime* are given, e.g., in [36, Eqs. (2.4a)–(2.4e)].

As the main result of this paper, Theorem 2.13, we show the universality of the local eigenvalue statistics of correlated real symmetric and complex Hermitian random matrices at cusp-like singularities. As mentioned in the introduction, the proof of cusp universality follows the *three-step strategy* [41], the first step of which is a *local law* (see Theorem 2.8) identifying the empirical eigenvalue distribution on a scale slightly above the typical eigenvalue spacing, with very high probability. After precisely formulating the assumptions that we impose on the random matrix (2.1) in Sect. 2.1, we present our novel local law in Sect. 2.2. Afterwards, in Sect. 2.3, we formulate our main result on cusp universality and other consequences of the local law, such as eigenvector delocalization and eigenvalue rigidity.

2.1. Assumptions. In this section, we precisely formulate the assumptions, under which our main result, Theorem 2.8, holds, and comment on them.

Assumption 2.1 *(Bounded expectation). There exists a constant $C_A > 0$ such that $\|A\| \le C_A$, uniformly in N.*

Assumption 2.2 *(Finite moments). For every $p \in \mathbf{N}$, there exists a constant μ_p such that $\mathbb{E}|\sqrt{N} w_\alpha|^p \le \mu_p$ for all $\alpha \in [N]^2$.*

Before formulating our assumption on the correlation structure of the random matrix W, we introduce some custom notation to keep the definition of the norms of the (normalized) *cumulants*[3],

$$\kappa(\alpha_1, ..., \alpha_k) \equiv \kappa(\sqrt{N} w_{\alpha_1}, ..., \sqrt{N} w_{\alpha_k}) , \tag{2.5}$$

relatively compact. First, a double index $\alpha_i \in [N]^2$ is represented by two single indices $a_i, b_i \in [N]$, identifying $\alpha_i \equiv (a_i, b_i)$. For brevity, we often use the notation $a_i b_i = (a_i, b_i)$. Next, if, instead of an index $a \in [N]$, we write a dot $(\cdot)$ in a scalar quantity, then we consider it as an N-vector indexed by the coordinate in place of the dot. As an example, $\kappa(a_1\cdot, a_2 b_2)$ is an N-vector, whose i-entry is $\kappa(a_1 i, a_2 b_2)$ and $\|\kappa(a_1\cdot, a_2 b_2)\|$ is its Euclidean (vector) norm. Similarly, $\|X(*, *)\|$ refers to the operator norm of the $N^2 \times N^2$ matrix with entries $X(\alpha_1, \alpha_2)$. We also introduce a combination of these conventions. In particular, $\big\| \|\kappa(\boldsymbol{x}*, \cdot*)\| \big\|$ denotes the operator norm $\|Y\|$ of the matrix Y with entries $Y(i, j) = \|\kappa(\boldsymbol{x}i, \cdot j)\| = \| \sum_a x_a \kappa(ai, \cdot j)\|$. Since the operator norm is invariant under transposition of the matrix, this does not lead to ambiguity regarding the

[3] Let $\boldsymbol{w} = (w_1, ..., w_k)$ be a random vector. Recall that its joint cumulants, $\kappa_{\boldsymbol{m}}$ with $\boldsymbol{m} \in \mathbf{N}_0^k$, are traditionally given as the coefficients of the log-characteristic function

$$\log \mathbb{E} e^{\mathrm{i}\boldsymbol{w}\cdot\boldsymbol{t}} = \sum_{\boldsymbol{m}} \kappa_{\boldsymbol{m}} \frac{(\mathrm{i}\boldsymbol{t})^{\boldsymbol{m}}}{\boldsymbol{m}!} .$$

For $\boldsymbol{w} = (\sqrt{N} w_{\alpha_1}, ..., \sqrt{N} w_{\alpha_k})$ we use the notation $\kappa(\alpha_1, ..., \alpha_k) \equiv \kappa(\sqrt{N} w_{\alpha_1}, ..., \sqrt{N} w_{\alpha_k}) := \kappa_{(1,...,1)}$ and note that, by construction, $\kappa(\alpha_1, ..., \alpha_k)$ is invariant under permutations of its arguments. For example, for $k = 2$, $\kappa(\alpha_1, \alpha_2) = N\mathbb{E}[w_{\alpha_1} w_{\alpha_2}]$.

order of i and j. Note that we use dot $(\cdot)$ as a placeholder for the variable related to the inner norm, and star $(*)$ for the outer norm.

The following assumption on the correlation structure of W is formulated in the real symmetric case. For complex Hermitian matrices, we require the cumulant norms introduced below to be bounded for all choices of real and imaginary in each of the arguments of a cumulant, i.e. for $\kappa(\alpha_1^{\mathfrak{X}_1}, ..., \alpha_k^{\mathfrak{X}_k}) = \kappa(\sqrt{N}\mathfrak{X}_1 w_{\alpha_1}, ..., \sqrt{N}\mathfrak{X}_k w_{\alpha_k})$ and all choices of $\mathfrak{X}_i \in \{\Re, \Im\}$ (see [35, Appendix C] for a more detailed discussion).

Assumption 2.3 *(Correlation structure). The correlations among the matrix entries $(w_\alpha)_\alpha$ of W satisfy the following.*

(i) The cumulants $\kappa(\alpha_1, ..., \alpha_k)$ have bounded matrix norms (viewed as an $N^2 \times N^2$ matrix), i.e. for all $k \geq 2$ there exists a constant $C_k > 0$ such that[4]

$$|||\kappa|||_k := \left\| \sum_{\alpha_1,...,\alpha_{k-2}} |\kappa(\alpha_1, ..., \alpha_{k-2}, *, *)| \right\| \leq C_k . \tag{2.6}$$

Moreover, we suppose that

$$|||\kappa|||_2^{\mathrm{iso}} := \inf_{\kappa=\kappa_{\mathrm{c}}+\kappa_{\mathrm{d}}} \big(|||\kappa_{\mathrm{c}}|||_c + |||\kappa_{\mathrm{d}}|||_d\big) \leq C_2 , \tag{2.7}$$

where the infimum is taken over all decompositions of κ in two functions $\kappa_{\mathrm{c}}, \kappa_{\mathrm{d}}$, where the subscripts stand for "direct" and "cross" (see [35, Remark 2.8] for an explanation of this terminology) and the corresponding norms are defined as

$$|||\kappa|||_d := \sup_{\|\boldsymbol{x}\|\leq 1} \Big\| \, \|\kappa(\boldsymbol{x}*, \cdot*)\| \, \Big\|, \quad \textit{and} \quad |||\kappa|||_c := \sup_{\|\boldsymbol{x}\|\leq 1} \Big\| \, \|\kappa(\boldsymbol{x}*, *\cdot)\| \, \Big\| .$$

Finally, we assume that

$$\begin{aligned} |||\kappa|||_3^{av} &:= N^{-3/2} \\ &\times \sup_{\substack{X, Y, Z \in \mathbb{C}^{N\times N} : \\ \|X\|, \|Y\| \leq 1, \ \|Z\|_{hs} \leq 1}} \sum_{ab, a_1b_1, a_2b_2} \big|\kappa(ab, a_1b_1, a_2b_2)\big| |X_{b_1a_2}| \, |Y_{b_2a_3}| \, |Z_{b_3a_1}| \leq C_3. \end{aligned} \tag{2.8}$$

(ii) There exists a positive $\mu > 0$, such that for every α there exists an index set $\mathcal{N}(\alpha)$ of cardinality $|\mathcal{N}(\alpha)| \leq N^{1/2-\mu}$ with the property that[5] *$w_\alpha \perp w_\beta$ for all $\beta \notin \mathcal{N}(\alpha)$. That is, every element is correlated with at most $N^{1/2-\mu}$ other matrix elements and is independent of the rest.*

The first part of Assumption 2.3 is needed to control every finite order term in a cumulant expansion in Proposition 5.2, analogously to Assumption (C) in [35]. The condition in (2.8) is needed only since we are dealing with Hilbert–Schmidt norm error terms and thus did not appear in [35], where the observables were bounded in terms of their operator norm. In Example 2.6 below, we present a prototypical class of models with a polynomially decaying metric correlation structure satisfying Assumption 2.3 (i). Complementary to Assumption 2.3 (i), the only purpose of the second part of Assumption 2.3

[4] We remark that the constants C_k in the bounds (2.6)–(2.8) could also be replaced by $C_{k,\nu}N^{\nu}$ for any $\nu > 0$, where $C_{k,\nu}$ is a positive constant. All our proofs hold under this more general condition, but we omit it for simplicity.

[5] In this context, the symbol $\perp$ means that the random variables are independent.

is to ensure that the cumulant expansion can be truncated. In [35], this was guaranteed by a more complicated and slightly more general condition on the correlation decay (cf. [35, Assumption (D)]).

Assumption 2.4 *(Fullness). We say that a random matrix H satisfies the fullness condition with a constant $c > 0$ if*

$$N\,\mathbb{E}\big[|\mathrm{Tr}[(H-\mathbb{E}H)X]|^2\big] \geq c\,\mathrm{Tr}[X^2], \tag{2.9}$$

for any deterministic matrix X of the same symmetry class as H (real symmetric or complex Hermitian).

We assume that there exists a constant $c_{\text{full}} > 0$ such that the random matrix H satisfies the fullness condition as in (2.9) with the constant $c := c_{\text{full}}$.

Assumption 2.5 *(Bounded self-consistent Green function). Fix $C_M, c_M > 0$ and define the set of* admissible energies *as*

$$\begin{aligned}\mathcal{I} \equiv \mathcal{I}_{C_M,c_M} := \{e \in \mathbb{R} : \|M(z)\| \leq C_M\langle z\rangle^{-1} \ \ \textit{for all} \ \ z \in \mathbb{C}\\ \textit{with} \ \ \Re z \in [e-c_M, e+c_M]\}\,.\end{aligned} \tag{2.10}$$

We assume that $\mathcal{I} \neq \emptyset$.

Recall that we refer to the constants in Assumptions 2.1–2.5 as *model parameters*.

Example 2.6 (Polynomially Decaying Metric Correlation Structure). A prime example of correlated random matrix satisfying the Assumption 2.3 (i) is the polynomially decaying model. For second order cumulants, we assume that

$$\big|\kappa(a_1b_1, a_2b_2)\big| \leq \frac{C_2}{1+d(a_1b_1,a_2b_2)^s}, \tag{2.11a}$$

for some $s > 2$, where we define the distance d on the set of labels $[N]^2$ as

$$d(a_1b_1, a_2b_2) := \min\big\{|a_1-a_2|+|b_1-b_2|, |a_1-b_2|+|b_1-a_2|\big\}. \tag{2.11b}$$

For cumulants of order $k \geq 3$, we assume the following decay condition

$$\big|\kappa(\alpha_1,\ldots,\alpha_k)\big| \leq C_k \prod_{e\in\mathfrak{T}_{\min}} \frac{1}{1+d(e)^s}, \tag{2.11c}$$

where $\mathfrak{T}_{\min}$ is a minimal spanning tree, i.e., a spanning tree for which the sum of the edge weights is minimal, in a complete graph with vertices $\alpha_1, \alpha_2, \ldots, \alpha_k$ and edge weights induced by the distance d, defined in (2.11b). The validity of (2.6)–(2.7) was asserted in Example 2.10 of [35], and we verify the new condition (2.8) in Appendix B.

2.2. Local law. In this section, we formulate our main technical result, the optimal local laws in Theorem 2.8. These show that $G(z) = (H-z)^{-1}$ is very well approximated by $M(z)$ in the $N \to \infty$ limit, with optimal convergence rate even at all singular points of the scDOS down to the typical eigenvalue spacing. We now define the scale on which the eigenvalues are predicted to fluctuate around a given energy e_0.

Definition 2.7 (Local fluctuation scale). Let $e_0 \in \mathcal{I}$ be an admissible energy. We define the self-consistent *fluctuation scale* $\eta_\mathfrak{f} = \eta_\mathfrak{f}(e_0) > 0$ (indicated by subscript $\mathfrak{f}$) at energy e_0 via

$$\int_{-\eta_\mathfrak{f}}^{\eta_\mathfrak{f}} \rho(e_0 + x)\mathrm{d}x = \frac{1}{N}\,, \tag{2.12}$$

if $e_0 \in \operatorname{supp}\rho$. In case that $e_0 \notin \operatorname{supp}\rho$, we define $\eta_\mathfrak{f}$ as the fluctuation scale at a nearby edge. More precisely, let I be the largest interval with $e_0 \in I \subset \mathbb{R}\backslash\operatorname{supp}\rho$ and set $\Delta := \min\{|I|, 1\}$. Then, $\eta_\mathfrak{f}$ satisfies the scaling relation

$$\eta_\mathfrak{f} \sim \begin{cases} N^{-2/3}\Delta^{1/9} & \text{if} \quad \Delta > N^{-3/4} \\ N^{-3/4} & \text{if} \quad \Delta \le N^{-3/4}\,. \end{cases} \tag{2.13}$$

While for e_0 in the *bulk*, where the scDOS satisfies $\rho \sim 1$, we have $\eta_\mathfrak{f} \sim N^{-1}$, it holds that $\eta_\mathfrak{f} \sim N^{-2/3}$ at a regular *edge* and $\eta_\mathfrak{f} \sim N^{-3/4}$ at an exact *cusp*.

Theorem 2.8 *(Optimal Local Laws). Fix small N-independent constants $\varepsilon_0, \xi_0 > 0$. Let $H \in \mathbb{C}^{N\times N}$ be a real symmetric or complex Hermitian correlated random matrix. Suppose that Assumptions 2.1–2.5 are satisfied, and let $\mathcal{I}$ be the set of admissible energies from* (2.10)*. Then, uniformly for all $z \in \mathbb{H}$ with $\Re z \in \mathcal{I}$ and* $\operatorname{dist}(z, \operatorname{supp}\rho) \in [N^{\varepsilon_0}\eta_\mathfrak{f}(\Re z), N^D]$*, the resolvent $G(z) := (H - z)^{-1}$ satisfies the* optimal isotropic local law*,*

$$\left|\big(G(z) - M(z)\big)_{\boldsymbol{x}\boldsymbol{y}}\right| \le N^{\xi_0}\sqrt{\frac{\rho(z)}{\langle z\rangle^2 N\eta}}\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|\,, \tag{2.14a}$$

for any deterministic vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$, and the optimal average local law*,*

$$\left|\big\langle\big(G(z) - M(z)\big)B\big\rangle\right| \le \frac{N^{\xi_0}}{\langle z\rangle N \operatorname{dist}\big(z, \operatorname{supp}\rho\big)}\|B\|_{\mathrm{hs}}\,, \tag{2.14b}$$

for any deterministic matrix $B \in \mathbb{C}^{N\times N}$, both with very high probability.

2.3. Delocalization, rigidity, and universality. The local law in Theorem 2.8 is the main input for eigenvector delocalization, eigenvalue rigidity, and universality, as stated below. While Corollaries 2.10–2.11 and Theorem 2.13 are proven as corollaries to Theorem 2.8 in Sect. 3.3, the exclusion of eigenvalues outside the support of the scDOS in Theorem 2.9 is obtained alongside the proof of Theorem 2.8 and presented in Sect. 6.

Theorem 2.9 *(No eigenvalues outside the support of the scDOS). Under the assumptions of Theorem 2.8 we have the following: Let $e_0 \in \mathcal{I}\setminus\operatorname{supp}\rho$. There exists a constant $c > 0$ such that for any fixed small N-independent constant $\theta_0 > 0$*

$$\operatorname{dist}\big(\operatorname{spec} H \cap [e_0 - c, e_0 + c], \operatorname{supp}\rho\big) \le N^{\theta_0}\eta_\mathfrak{f}(e_0), \tag{2.15}$$

with very high probability. Here we use the convention that $\operatorname{dist}(\emptyset, \ldots) = 0$.

Corollary 2.10 *(Eigenvector delocalization). Let $\boldsymbol{u}_i \in \mathbb{C}^N$ with $\|\boldsymbol{u}_i\| = 1$ be a normalized eigenvector of H corresponding to the eigenvalue λ_i. Then, under the assumptions of Theorem 2.8, for any small N-independent constant $\omega_0 > 0$, the estimate*

$$\max_{\substack{i \in [N]: \\ \lambda_i \in \mathcal{I}}} \left| \langle \boldsymbol{x}, \boldsymbol{u}_i \rangle \right| \leq \frac{N^{\omega_0}}{\sqrt{N}} \tag{2.16}$$

holds with very high probability, uniformly in deterministic vectors $\boldsymbol{x} \in \mathbb{C}^N$ with $\|\boldsymbol{x}\| = 1$.

Corollary 2.11 *(Band rigidity and eigenvalue rigidity). Assume the conditions of Theorem 2.8 with $\mathcal{I} = \mathbb{R}$ in Assumption 2.5. Then, the following holds.*

(a) For any $\theta > 0$, whenever $e_0 \in \mathbb{R} \setminus \operatorname{supp}\rho$ with $\operatorname{dist}(e_0, \operatorname{supp}\rho) \geq N^\theta \eta_{\mathfrak{f}}(e_0)$, the number of eigenvalues less than e_0 is deterministic with high probability. More precisely,

$$\left| \operatorname{spec} H \cap (-\infty, e_0) \right| = N \int_{-\infty}^{e_0} \rho(x) \mathrm{d}x, \quad \textit{w.v.h.p.} \tag{2.17}$$

(b) Let $\lambda_1 \leq \ldots \leq \lambda_N$ denote the ordered eigenvalues of H and assume that $e_0 \in \operatorname{int}(\operatorname{supp}\rho)$. Then, for any small N-independent constant $\chi_0 > 0$, it holds that

$$\left| \lambda_{k(e_0)} - e_0 \right| \leq N^{\chi_0} \eta_{\mathfrak{f}}(e_0), \tag{2.18}$$

with very high probability, where we defined the (self-consistent) eigenvalue index *as $k(e_0) := \lceil N \int_{-\infty}^{e_0} \rho(x) \mathrm{d}x \rceil$.*

Remark 2.12 (Integer mass). We point out that (2.17) entails the nontrivial fact that, whenever $e_0 \notin \operatorname{supp}\rho$ satisfies $\operatorname{dist}(e_0, \operatorname{supp}\rho) \geq N^\theta \eta_{\mathfrak{f}}(e_0)$ for some $\theta > 0$, the integral $N \int_{-\infty}^{e_0} \rho(x) \mathrm{d}x$ is always an integer. An immediate consequence is that, for each connected component $[a, b]$ of $\operatorname{supp}\rho$, it holds that $N \int_a^b \rho(x) \mathrm{d}x$ is an integer. That is, each *spectral band* contains that number of eigenvalues with very high probability. For spectral bands which are separated by a distance of order one, this was previously shown in [11, Corollary 2.9]. Our Corollary 2.11 improves this to the optimal minimal distance $N^\epsilon \eta_{\mathfrak{f}}(e_0)$.

As our last consequence to the optimal local laws in Theorem 2.8, we prove cusp universality in Theorem 2.13 below. Since universality is already known in the bulk [35] as well as the edge regime [11], we will henceforth focus on the (approximate) cubic-root cusp. However, the optimal local laws of Theorem 2.8 can be used as an input for the three-step strategy to yield bulk and edge universality as well. From the in-depth analysis of the MDE (2.3) and its solution in [10], we know that the scDOS ρ is described by explicit universal shape functions in the vicinity of local minima with a small value of ρ and near small gaps in the support of ρ; see, e.g., [36, Eqs. (2.4a)–(2.4e)] for precise formulas.

Whenever the local length scale of such an almost cusp shape around a point $\mathfrak{b}$ matches (or is smaller than) the local eigenvalue spacing, i.e. if $\mathfrak{b}$ is a small local minimum, satisfying $\rho(\mathfrak{b}) \lesssim N^{-1/4}$, or a midpoint of a gap with width $\Delta \lesssim N^{-3/4}$, then we call the local shape around $\mathfrak{b}$ a *physical cusp* – reflecting the fact that it becomes indistinguishable from an exact cusp when resolved with a precision (slightly) above the local eigenvalue spacing $\sim N^{-3/4}$. In this case, $\mathfrak{b}$ is called a *physical cusp point*. Besides the local length scale of a physical cusp point $\mathfrak{b}$, the specific shape of the scDOS around $\mathfrak{b}$ is characterized by a single additional parameter $\gamma > 0$, called the *slope parameter*.

In order to formulate our result on cusp universality in Theorem 2.13, it is natural to consider the rescaled k-point function $p_k^{(N)}$, which is implicitly defined as

$$\mathbb{E}\binom{N}{k}^{-1} \sum_{\{j_1,\dots,j_k\}\subset[N]} f(\lambda_{j_1},\dots,\lambda_{j_k}) =: \int_{\mathbb{R}^k} f(\boldsymbol{x}) p_k^{(N)}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,, \tag{2.19}$$

for any test function f. Here, the summation is over all distinct subsets of k integers from $[N]$.

Theorem 2.13 *(Cusp universality for correlated random matrices). Let $H \in \mathbb{C}^{N\times N}$ be a real symmetric or complex Hermitian correlated random matrix as in (2.1). Suppose that Assumptions 2.1–2.5 are satisfied, assume that a* physical cusp point $\mathfrak{b} \in \mathcal{I}$ *lies in the set of admissible energies (2.10), and let $\gamma > 0$ be the appropriate slope parameter at $\mathfrak{b}$. Then, the local k-point correlation function at $\mathfrak{b}$ is universal. That is, for every $k \in \mathbf{N}$ there exists a k-point correlation function $p_{k,\alpha}^{\mathrm{GOE/GUE}}$ such that for any test function $F \in C_c^1(\overline{\Omega})$ on a bounded open set $\Omega \subset \mathbb{R}^k$, it holds that,*[6]

$$\int_{\mathbb{R}^k} F(\boldsymbol{x}) \left[\frac{N^{k/4}}{\gamma^k} p_k^{(N)}\left(\mathfrak{b} + \frac{\boldsymbol{x}}{\gamma N^{3/4}}\right) - p_{k,\alpha}^{\mathrm{GOE/GUE}}(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x} = \mathcal{O}_{k,\Omega}(N^{-c(k)}\|F\|_{C^1})\,, \tag{2.20}$$

where the parameter α depends on γ, the local length scale and the specific shape of the scDOS around $\mathfrak{b}$, i.e., whether it is an exact cusp, a small gap, or a small minimum (see [36, Eq. (2.6)] or [32, Eq. (2.5)]). The constant $c(k) > 0$ in (2.20) depends only on k, and the implicit constant in the error term depends on k and the diameter of the set Ω.

Remark 2.14 (On $p_{k,\alpha}^{\mathrm{GUE/GOE}}$). For the universal k-point correlation function $p_{k,\alpha}^{\mathrm{GOE/GUE}}$, we have the following.

(i) In the *complex Hermitian* symmetry class, the k-point function takes the determinantal form

$$p_{k,\alpha}^{\mathrm{GUE}}(\boldsymbol{x}) = \det\left(K_\alpha(x_i, x_j)\right)_{i,j=1}^k\,, \tag{2.21}$$

where the *extended Pearcey kernel* with parameter $\alpha \in \mathbb{R}$ is given by

$$K_\alpha(x,y) = \frac{1}{(2\pi\mathrm{i})^2}\int_\Xi \mathrm{d}z \int_\Phi \mathrm{d}w\, \frac{\exp\left(-w^4/4 + \alpha w^2/2 - yw + z^4/4 - \alpha z^2/2 + xz\right)}{w-z}\,. \tag{2.22}$$

Here, Ξ is a contour consisting of rays from $\pm e^{\mathrm{i}\pi/4}$ to 0 and rays from 0 to $\pm e^{-\mathrm{i}\pi/4}$, and Φ is the ray from $-\mathrm{i}\infty$ to $\mathrm{i}\infty$. See [5,23,67] and the references in [36] for more details.

(ii) In the *real symmetric* case, the k-point correlation function $p_{k,\alpha}^{\mathrm{GOE}}$ (possibly only a distribution) is not known explicitly, not even if it is Pfaffian. However, $p_{k,\alpha}^{\mathrm{GOE}}$ exists in the dual of C^1 as the limit of correlation functions of a suitable one-parameter family of Gaussian comparison models (see Sec. 3 and in particular Eq. (3.5) of [32]).

[6] Here, $\mathfrak{b}$ is identified with the vector $(\mathfrak{b}\dots,\mathfrak{b}) \in \mathbb{R}^k$.

3. Zigzag Strategy: Proof of the Main Results

To streamline the presentation, we assume that the set of admissible energies $\mathcal{I}$, defined in (2.10) of Assumption 2.5, is the entire real line, that is, $\mathcal{I} = \mathbb{R}$. We discuss the straightforward modifications for general $\mathcal{I}$ in Remark 3.8.

Definition 3.1 (Local Laws). Let H_u be a random matrix depending on a parameter[7] $u \in \mathcal{U}$, and let M_u be the solution to the MDE (2.3) with the data pair $(\mathbb{E}H_u, \mathcal{S}_{H_u})$, where $\mathcal{S}_{H_u}$ is defined in (2.2). For all $u \in \mathcal{U}$, let $\mathcal{D}_u \subset \mathbb{H}$ and let $\xi > 0$. We say that the resolvent $G_u(z) := (H_u - z)^{-1}$ satisfies the averaged local law and the isotropic local law, respectively, with data $(\mathcal{D}_u, \xi)$ uniformly in $u \in \mathcal{U}$, if and only if the bounds

$$\left|\left\langle \big(G_u(z) - M_u(z)\big)B\right\rangle\right| \le \frac{N^{3\xi}}{N\eta}, \quad \text{and} \quad \left|\big(G_u(z) - M_u(z)\big)_{\boldsymbol{x}\boldsymbol{y}}\right| \le N^{\xi}\left(\sqrt{\frac{\rho_u(z)}{N\eta}} + \frac{1}{N\eta}\right), \tag{3.1}$$

hold uniformly in $z := E + \mathrm{i}\eta \in \mathcal{D}_u$ and in $u \in \mathcal{U}$, with very high probability, for any deterministic vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$ with $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$, and any deterministic matrices B with $\|B\|_{\mathrm{hs}} = 1$. Here $\rho_u(z) := \frac{1}{\pi}\langle \Im M_u(z)\rangle$.

The goal of the present section is to prove the local laws in the *above the scale* regime, where $\rho(z)N|\Im z|$ is large. Fix a (small) N-independent constant $\varepsilon > 0$, a large constant $C_L > 0$, and define the spectral domain $\mathcal{D}^{\mathrm{abv}}$ as

$$\mathcal{D}^{\mathrm{abv}} \equiv \mathcal{D}^{\mathrm{abv}}(\varepsilon, C_L) := \left\{z := E + \mathrm{i}\eta \in \mathbb{H} \,:\, \rho(z)N\eta \ge N^{\varepsilon},\ |E| \le C_L,\ \eta \le C_L\right\}. \tag{3.2}$$

The regime $\rho(z)N\eta \ge N^{\varepsilon}$ is natural for studying the local laws, since $\rho(E + \mathrm{i}\eta)N\eta$ is the typical number of eigenvalues in the interval of size η around the energy E.

Theorem 3.2 *(Local Laws above the Scale). Fix a (small) N-independent constant $\varepsilon > 0$, a large constant $C_L > 0$. Let H be a random matrix satisfying the Assumptions 2.1–2.5, then the resolvent $G(z) := (H - z)^{-1}$ satisfies the local laws (3.1) with data $(\mathcal{D}^{\mathrm{abv}}, 2\xi)$, for any fixed tolerance exponent $0 < \xi \le \frac{1}{100}\varepsilon$, where $\mathcal{D}^{\mathrm{abv}} = \mathcal{D}^{\mathrm{abv}}(\varepsilon, C_L)$.*

To prove Theorem 2.8 in the *below the scale* regime, that is, to handle the case when $\rho(z)N|\Im z|$ is small, we proceed in two steps. In the key first step we use the local laws above the scale of Theorem 3.2 to prove Theorem 2.9 that asserts the absence of spectrum outside of the support of the scDOS ρ. Then the second step is a routine derivation of (2.14b) and (2.14a) from (2.15) and (3.1). Both steps are presented in Sect. 6. In the main part of the proof, we only consider spectral parameters z satisfying $\mathrm{dist}(z, \mathrm{supp}\rho) \lesssim 1$. The easy extension to the regime $\mathrm{dist}(z, \mathrm{supp}\rho) \gtrsim 1$ and the resulting $\langle z\rangle^{-2}$-decay are briefly addressed in the discussion above (6.28).

In the sequel, we treat the constants ε, C_L in (3.2) as additional model parameters and omit them from the arguments of $\mathcal{D}^{\mathrm{abv}}$.

Throughout the paper, we consistently use the notation $\varepsilon, \xi, \zeta, \delta$ to represent positive N-independent tolerance exponents, each playing a particular role in the proof. Specifically, ε denotes the tolerance exponent from the definition of the domain $\mathcal{D}^{\mathrm{abv}}$ (see (3.2) and (3.21) below); ξ and its multiples represent the target tolerance exponents for the local laws above the scale in (3.1). The exponent ζ appears in the below-the-scale part of the proof (Sect. 6). Multiples of $-\zeta$ are used in the exclusion estimate (6.9) and in

[7] In applications, the parameter u will typically be time and the set $\mathcal{U}$ will be a bounded subinterval of $\mathbb{R}$.

the lower bound on $\rho N \eta$ in (6.8). The exponent δ refers to the step size used in various inductive arguments. In the sequel, we adhere to the following conventions:

$$\delta \ll \xi \ll \varepsilon, \quad \zeta \ll \xi, \quad \delta < \mu, \tag{3.3}$$

where $\mu > 0$ is the constant from Assumption 2.3 (ii). We also assume that the arbitrary exponent $\nu > 0$ is much smaller than the other tolerance exponents, that is, $\nu \ll \delta$ and $\nu \ll \zeta$.

3.1. Input: global laws. Let $\rho(z)$ be the harmonic extension to $\mathbb{H}$ of the scDOS corresponding to a solution of (2.3). Given small positive constants $\varepsilon, \xi > 0$, and a large constant $D > 0$, we define the global domain as

$$\begin{aligned} &\mathcal{D}^{\text{glob}} \equiv \mathcal{D}^{\text{glob}}(D, \varepsilon, \xi, \rho) := \\ &\quad \big\{z := E + \mathrm{i}\eta \in \mathbb{H} : \, |E| \le N^D, \ N^{-1+\varepsilon} \le \eta \le N^D, \ \rho(z)^{-1}\eta \ge N^{-\xi/4}\big\}. \end{aligned} \tag{3.4}$$

Effectively, the function $\rho(z)^{-1}\eta$ in (3.4) controls the proximity of the spectral parameter z to the support ρ.

Proposition 3.3. *Let H be a random matrix satisfying the Assumptions 2.1–2.5, and let $\rho(z)$ be the scDOS arising from the solution to the MDE (2.3) corresponding to H. Let $\mathcal{D}^{\text{bdd}} := \mathcal{I} + [-c_M, c_M] + \mathrm{i}\mathbb{R} \subset \mathbb{C}$, where $\mathcal{I}$ is defined in (2.10). Fix a large constant $D > 0$ and a tolerance exponent $0 < \xi < \frac{1}{10}\varepsilon$. Then the resolvent $G(z) := (H - z)^{-1}$ satisfies*

$$\Big|\big(G(z) - M(z)\big)_{\boldsymbol{x}\boldsymbol{y}}\Big| \le N^{\xi}\Psi(z)\|\boldsymbol{x}\|\|\boldsymbol{y}\|, \tag{3.5a}$$

$$\Big|\big\langle\big(G(z) - M(z)\big)B\big\rangle\Big| \le N^{3\xi}\Psi(z)\sqrt{\frac{\langle z\rangle}{N\eta}}\|B\|_{\text{hs}}, \tag{3.5b}$$

with very high probability, uniformly in $z := E + \mathrm{i}\eta \in \mathcal{D}^{\text{glob}}(D, \varepsilon, \xi, \rho) \cap \mathcal{D}^{\text{bdd}}$, for any deterministic vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and matrices B. Here the control parameter $\Psi(z)$ is defined as

$$\Psi(z) := \sqrt{\frac{\rho(z)}{\langle z\rangle^2 N\eta}} + \frac{1}{\langle z\rangle^2 N\eta}, \quad \eta := \Im z. \tag{3.6}$$

We prove Proposition 3.3 in Sect. 7.

3.2. Local law via zigzag strategy: Proof of Theorem 3.2.

3.2.1. Preliminaries: Two Random Matrix Flows For any random matrix H, we define the covariance tensor Σ_H corresponding to H by its action on any deterministic matrix $X \in \mathbb{C}^{N\times N}$,

$$\Sigma_H[X] := \mathbb{E}\big[\mathrm{Tr}\big[(H - \mathbb{E}H)X\big](H - \mathbb{E}H)\big]. \tag{3.7}$$

Note that Σ_H is different from the self-energy operator (2.2), but they both carry equivalent information. Moreover, it is positive definite on the space of matrices equipped with the usual scalar product $(X, Y) = \langle X^*Y\rangle$ and we will denote by $\Sigma^{1/2}$ its square root.

Along the proof, we use two distinct flows in the space of $N \times N$ random matrices: the *zig-flow* (standard Ornstein–Uhlenbeck process), defined as

$$\mathrm{d}H_t = -\frac{1}{2}H_t\mathrm{d}t + \frac{\mathrm{d}\mathfrak{B}_t}{\sqrt{N}}, \qquad t \geq 0; \tag{3.8}$$

and the *zag-flow* (modified Ornstein–Uhlenbeck process), distinguished by the superscript t,

$$\mathrm{d}H^t = -\frac{1}{2}\big(H^t - \mathbb{E}H^t\big)\mathrm{d}t + \Sigma_{H^0}^{1/2}\big[\mathrm{d}\mathfrak{B}_t\big], \qquad t \geq 0, \tag{3.9}$$

where Σ_{H^0} is the covariance tensor of H^0, defined according to (3.7). In both (3.8) and (3.9), $\mathfrak{B}_t$ denotes the real symmetric or complex Hermitian Brownian motion, depending on the symmetry class of H.

Note that along the zig-flow (3.8), the covariance tensor $\Sigma_t := \Sigma_{H_t}$, corresponding to H_t via (3.7), satisfies the ordinary differential equation

$$\mathrm{d}\Sigma_t = \big(-\Sigma_t + \Sigma_{\mathrm{G}}\big)\mathrm{d}t, \tag{3.10}$$

where Σ_{G} is the covariance tensor of a GOE/GUE matrix in the same symmetry class as H. That is $\Sigma_{\mathrm{G}}[X] = N^{-1}X$ in the complex Hermitian case, and $\Sigma_{\mathrm{G}}[X] = N^{-1}(X + X^{\mathfrak{t}})$ in the real-symmetric case, where $X^{\mathfrak{t}}$ denotes the transpose of X. On the other hand, along the zag-flow (3.9), the expectation and the covariance tensor of H_t (and hence the self-energy $\mathcal{S}_{H_t}$) are preserved. Therefore, the deterministic approximation M remains unchanged along the zag-flow.

For any $t \geq 0$, we define the flow maps $\mathfrak{F}^t_{\mathrm{zig}}$ and $\mathfrak{F}^t_{\mathrm{zag}}$ on the space of probability distribution $\mathcal{P}(\mathbb{C}^{N\times N})$ by

$$\mathfrak{F}^t_{\mathrm{zig}}\big[H\big] := H_t, \quad \text{where } H_t \text{ solves (3.8) with the initial condition } H_0 = H. \tag{3.11}$$

$$\mathfrak{F}^t_{\mathrm{zag}}\big[H\big] := H^t, \quad \text{where } H^t \text{ solves (3.9) with the initial condition } H^0 = H. \tag{3.12}$$

The key relation between the flow maps $\mathfrak{F}^t_{\mathrm{zig}}$ and $\mathfrak{F}^t_{\mathrm{zag}}$ is captured by the following lemma.

Lemma 3.4 *(Flow Distribution Surjectivity). Let H be a random matrix satisfying the fullness condition* (2.9) *with a constant* $0 < c < 1$, *then there exists a random matrix* $\mathfrak{H}_{c,t}(H)$ *such that*

$$\mathfrak{F}^t_{\mathrm{zig}}\big[\mathfrak{H}_{c,t}(H)\big] \overset{d}{=} \mathfrak{F}^{s(t)}_{\mathrm{zag}}\big[H\big], \quad 0 \leq t \leq -\log(1-c), \tag{3.13}$$

where the function $s(t) \equiv s_c(t)$ is defined as

$$s(t) \equiv s_c(t) := \log c - \log\big(c - 1 + \mathrm{e}^{-t}\big), \tag{3.14}$$

and satisfies

$$s(t) \leq 2c^{-1}t, \quad 0 \leq t \leq c/2. \tag{3.15}$$

We defer the proof of Lemma 3.4 to the Appendix A.

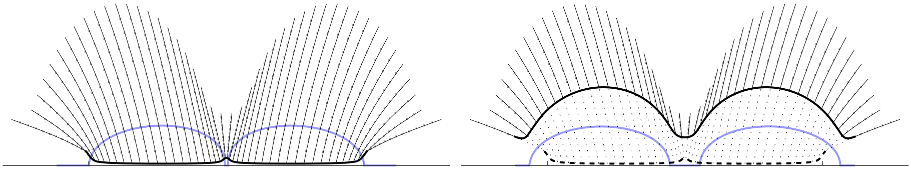

Fig. 1. The left panel depicts several trajectories of the flow (3.18) that terminate at the *scale curve* $\rho_T(z)N\Im z = c$ (solid black line), while the the graph of scDOS ρ_T is superimposed in light blue. The right panel depicts trajectories up to an intermediate time $t \in (0, T)$ with their continuations beyond t shown as thin dotted lines. The pre-image of the scale curve at the time t is depicted as a solid black line, and the scale curve itself is depicted as a dashed black line. The graph of scDOS ρ_t is superimposed in light blue. In both panels, the black markers along the trajectories of (3.18) are evenly spaced in time

3.2.2. Zigzag approach: Iterative application of the characteristic flow and GFT We consider the time-dependent matrix Dyson equation (MDE),

$$-M_t(z)^{-1} = z - A_t + \mathcal{S}_t\big[M_t(z)\big], \quad z \in \mathbb{C}\backslash\mathbb{R}, \quad (\Im z)\Im M_t(z) > 0, \tag{3.16}$$

where the data pair $(A_t, \mathcal{S}_t)$ is given as the unique solutions to the differential equations

$$\mathrm{d}A_t = -\frac{1}{2}A_t\mathrm{d}t\,, \quad \mathrm{d}\mathcal{S}_t = (-\mathcal{S}_t + \langle\cdot\rangle)\mathrm{d}t\,. \tag{3.17}$$

with the *terminal conditions* $A_T = A = \mathbb{E}H$ and $\mathcal{S}_T = \mathcal{S} = \mathbb{E}[(H-A)(\,\cdot\,)(H-A)]$, respectively.

Given $M_t(z)$, we consider the *characteristic ODE* for the time dependent spectral parameter $z_t \in \mathbb{C}$ (see Figure 1),

$$\mathrm{d}z_t = -\frac{1}{2}z_t\mathrm{d}t - \big\langle M_t(z_t)\big\rangle\mathrm{d}t. \tag{3.18}$$

By trivial ODE arguments, for all $0 \le s \le t$, the corresponding (inverse) flow map $\varphi_{s,t} : \overline{\mathbb{H}} \to \overline{\mathbb{H}}$ is defined uniquely by

$$\varphi_{s,t}(z_t) := z_s, \quad \text{where } z_s \text{ solves (3.18).} \tag{3.19}$$

It can be directly checked that along the trajectories of (3.18), the solution to the time-dependent MDE (3.16) satisfies

$$\mathrm{d}M_t(z_t) = \frac{1}{2}M_t(z_t)\mathrm{d}t. \tag{3.20}$$

Lemma 3.5 *(Time-Dependent Domains). There exist a constant* $C' \sim 1$ *such that for any constant* $0 < c' \le \pi$ *and any terminal time* $0 < T \lesssim 1$*, the time-dependent domains* $\mathcal{D}_t^{\mathrm{abv}}$*,* $t \in [0, T]$*, (see Figure 2), defined as*

$$\begin{aligned}\mathcal{D}_t^{\mathrm{abv}} &\equiv \mathcal{D}_t^{\mathrm{abv}}(\varepsilon, C_L, c', T)\\ &:= \big\{z := E + \mathrm{i}\eta \in \mathbb{H} \,:\, \rho_t(z)N\eta \ge N^{\varepsilon},\ |E| \vee \eta \le C_L + C'\cdot(T-t),\\ &\qquad \rho_t(z)^{-1}\eta \ge c'\cdot\big(N^{-1+\varepsilon} + T - t\big)\big\},\end{aligned} \tag{3.21}$$

satisfy $\varphi_{s,t}(\mathcal{D}_t^{\mathrm{abv}}) \subset \mathcal{D}_s^{\mathrm{abv}}$ *for all* $0 \le s \le t \le T$*, where* $\varphi_{s,t}$ *is the flow map defined in* (3.19).

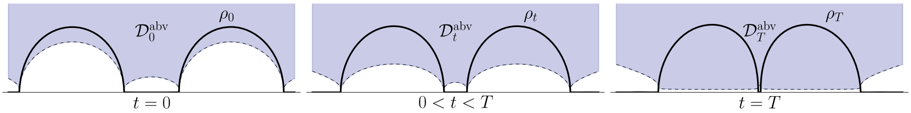
$\mathcal{D}_0^{\mathrm{abv}}$
ρ_0
$t=0$
$\mathcal{D}_t^{\mathrm{abv}}$
ρ_t
$0<t<T$
$\mathcal{D}_T^{\mathrm{abv}}$
ρ_T
$t=T$

Fig. 2. The time-dependent domain $\mathcal{D}_t^{\mathrm{abv}}$, defined in (3.21), is illustrated in blue at three distinct times: the initial time $t=0$ (left), an intermediate time $0<t<T$ (center), and the terminal time $t=T$ (right). The graph of the scDOS ρ_t is superimposed in black on each panel (not to scale)

We defer the proof of Lemma 3.5 to Appendix A.

As in (3.4), the function $\rho_t(z)^{-1}\eta$ in the definition (3.21) effectively controls the distance between z and the support of ρ_t. Therefore the time-dependent family of domains $\mathcal{D}_t^{\mathrm{abv}}$ effectively interpolates between the global regime $\mathcal{D}^{\mathrm{glob}}$ and the final target domain $\mathcal{D}^{\mathrm{abv}}$.

Indeed, since $\rho(z)\lesssim 1$, by choosing the constant $c'\sim 1$ in (3.21) small enough, we can guarantee that $\mathcal{D}^{\mathrm{abv}}\subset\mathcal{D}_T^{\mathrm{abv}}$, where we recall that $\mathcal{D}^{\mathrm{abv}}$ is defined in (3.2). On the other hand, it follows from (3.4) that by choosing

$$T := CN^{-\xi/4}, \tag{3.22}$$

with a sufficiently large constant $C\gtrsim 1$, we can guarantee that $\mathcal{D}_0^{\mathrm{abv}}\subset\mathcal{D}^{\mathrm{glob}}$, where $\mathcal{D}^{\mathrm{glob}}$ is defined in (3.4).

We conduct the proof inductively. Fix a tolerance exponent $0<\xi\ll\varepsilon$, a step size $0<\delta\ll\xi$ (recall (3.3)). For the terminal time T chosen as in (3.22), let K be the smallest integer such that $N^{-K\delta}T\le N^{-1+\varepsilon}$, and define a sequence of times $\{t_k\}_{k=0}^K$ as

$$t_0 := 0, \quad t_k := T - N^{-k\delta}T, \quad k\in\{1,\dots,K-1\}, \quad t_K := T. \tag{3.23}$$

Let $\{\vartriangle t_k\}_{k=1}^K$ denote the difference sequence of $\{t_k\}_{k=0}^K$, that is

$$\vartriangle t_k := t_k - t_{k-1}, \quad k\in\{1,\dots,K\}. \tag{3.24}$$

Let Σ_t solve the equation (3.10) with the terminal condition $\Sigma_T=\Sigma$, where Σ, defined via (3.7), is the covariance tensor of the target matrix H, for which we eventually prove the local laws in Theorem 2.8. Observe that for all $0\le t\le T$, the solution Σ_t satisfies

$$\Sigma_t \ge \widetilde{c}\,\Sigma_{\mathrm{G}}, \quad \widetilde{c} := \frac{c_{\mathrm{flat}}}{2}\wedge 1, \tag{3.25}$$

where c_{flat} is the constant in Assumption 2.4. Given the target random matrix ensemble H, we construct two sequences of random matrices, $\{H_k\}_{k=0}^K$ and $\{H^k\}_{k=1}^K$ recursively by

$$H_K := H, \quad H^k := \mathfrak{F}_{\mathrm{zag}}^{s(\vartriangle t_k)}\big[H_k\big], \quad H_{k-1} := \mathfrak{H}_{\widetilde{c},\vartriangle t_k}\big(H_k\big), \quad k\in\{1,\dots,K\}, \tag{3.26}$$

where $s(t) := s_{\widetilde{\alpha}}(t)$ and $\mathfrak{H}_{\widetilde{c},\vartriangle t_k}$ are given by Lemma 3.4, and $\widetilde{c}$ is the constant in (3.25). It follows by a simple backward inductive argument starting at $k=K$ that the covariance tensor of both H_k and H^k is given by Σ_{t_k}, hence by (3.25), H_{k-1} is well-defined.

Proposition 3.6 *(Zig Step). Fix $k\in\{1,\dots,K\}$, and denote*

$$G_t(z) := \big(\mathfrak{F}_{\mathrm{zig}}^{t-t_{k-1}}[H_{k-1}] - z\big)^{-1}, \quad t_{k-1}\le t\le t_k. \tag{3.27}$$

Assume that for some $\xi,\nu>0$ with $\xi+K\nu\ll\varepsilon$, and $\ell\le 2k$, the resolvent G_t satisfies the local laws (3.1) with data $(\mathcal{D}_t^{\mathrm{abv}},\xi+\ell\nu)$ at time $t=t_{k-1}$, then the resolvent G_t satisfies the local laws (3.1) with data $(\mathcal{D}_t^{\mathrm{abv}},\xi+(\ell+1)\nu)$ uniformly in $t\in[t_{k-1},t_k]$.

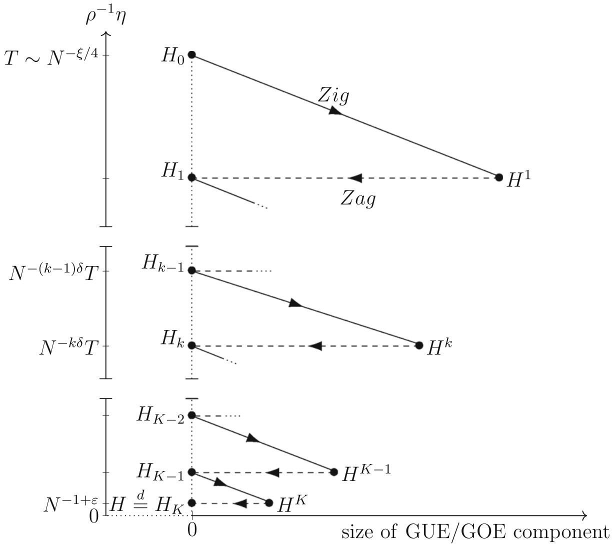

Fig. 3. Schematic representation of the Zigzag induction. The random matrices H_k, H^k, as defined in (3.26), are situated within an abstract coordinate system. The horizontal axis represents the size of the Gaussian component, while the vertical axis indicates the lower bound on $\rho(z)^{-1}\eta$ in the domains, c.f. (3.21), where we prove the local laws (3.1). Solid arrows denote applications of Proposition 3.6 (referred to as *Zig* steps, in which we reduce $\rho^{-1}\eta$ at the cost of introducing a Gaussian component), and dashed arrows indicate applications of Proposition 3.7 (*Zag* steps, in which we keep the spectral parameter fixed and remove the previously introduced Gaussian component)

Proposition 3.7 *(Zag Step). Fix $k \in \{1, \dots, K\}$. Let $s_k := s(\vartriangle t_k)$ be the time defined in* (3.14)*, let H_k be the random matrix defined in* (3.26)*, and denote*

$$G^s(z) := \left(\mathfrak{F}^s_{\text{zag}}[H_k] - z\right)^{-1}, \quad 0 \le s \le s_k. \tag{3.28}$$

Assume that for some $\xi, \nu > 0$ with $\xi + K\nu \ll \varepsilon$, and $\ell \le 2k$, the resolvent G^s satisfies the local laws (3.1) *with data $(\mathcal{D}^{\text{abv}}_{t_k}, \xi + \ell\nu)$ at time $s = s_k$, then G^s satisfies the local laws* (3.1) *with data $(\mathcal{D}^{\text{abv}}_{t_k}, \xi + (\ell + 1)\nu)$ uniformly in $s \in [0, s_k]$.*

Having formulated the cardinal steps of the Zigzag strategy, we now put them together to prove our key theorem on the local laws above the scale. Note that in the above the scale regime $\rho(z)N\eta \ge N^{\varepsilon}$, the term $1/(N\eta)$ in the isotropic bound is (3.1) is dominated by $\sqrt{\rho/(N\eta)}$, and hence will be ignored in Sects. 4 and 5.

Proof of Theorem 3.2. Recall our choice of the constant $c' \sim 1$ in (3.21) and the terminal time $T \sim N^{-\xi/4}$ in (3.22) that guarantees the inclusions $\mathcal{D}^{\text{abv}}_0 \subset \mathcal{D}^{\text{glob}}$ and $\mathcal{D}^{\text{abv}} \subset \mathcal{D}^{\text{abv}}_T$. Therefore, Proposition 3.3 implies that the resolvent $G_0(z) := (H_0 - z)^{-1}$ of a random matrix H_0, defined in (3.26), satisfies the local laws (3.1) with data $(\mathcal{D}^{\text{abv}}_0, \xi)$. Using Propositions 3.6 and 3.7 in tandem K times, we prove by forward induction on k that for any $\nu > 0$, the resolvent $G_k(z) := (H_k - z)^{-1}$ satisfies the local laws (3.1) with data $(\mathcal{D}^{\text{abv}}_{t_k}, \xi + 2k\nu)$, for all $k \in \{1, \dots, K\}$. Since $H_K = H$ and $\mathcal{D}^{\text{abv}}_{t_K} = \mathcal{D}^{\text{abv}}_T \supset \mathcal{D}^{\text{abv}}$, this concludes the proof of Theorem 3.2. □

Remark 3.8 (On Locality of Assumption 2.5). In the case of a general set of admissible energies $\mathcal{I}$, defined in (2.10), our proof holds verbatim, except the spectral domains

$\mathcal{D}^{\rm glob}$, $\mathcal{D}^{\rm abv}$, $\mathcal{D}_t^{\rm abv}$ used along the proof have to be restricted. More precisely, we need the following modifications:

(i) we restrict the domain $\mathcal{D}_t^{\rm abv}$, defined in (3.21), by intersecting it with the region

$$\mathcal{D}_t^{\rm bdd} := \big\{z \in \mathbb{C} \,:\, \mathrm{dist}(\Re z, \mathcal{I}) \le c_M/2 + C' \cdot (T - t)\big\}, \quad 0 \le t \le T; \tag{3.29}$$

(ii) we restrict the domain $\mathcal{D}^{\rm abv}$, defined in (3.2), by intersecting it with $\mathcal{D}_T^{\rm bdd}$;

(iii) we restrict the global domain $\mathcal{D}^{\rm glob}$, defined in (3.4) by intersecting it with the set $\{z \in \mathbb{C} \,:\, \mathrm{dist}(\Re z, \mathcal{I}) \le \frac{3}{4} c_M\}$.

3.3. Proofs of Corollaries 2.10–2.11 and Theorem 2.13. In this section, we deduce eigenvector delocalization, band rigidity and eigenvalue rigidity, as well as cusp universality from the local law in Theorem 2.8. These arguments are essentially independent of the correlation structure of the random matrix, so we only refer to analogous proofs, which can easily be adjusted to our case with straightforward modifications.

Proof of Corollary 2.10 on eigenvector delocalization. As usual, eigenvector delocalization is an immediate consequence of the optimal isotropic local law from Theorem 2.8 for $\Im G$; see [35, Proof of Corollary 2.4] or [7, Proof of Corollary 1.14] for this argument. □

Proof of Corollary 2.11 on band rigidity and eigenvalue rigidity. The proof of band rigidity was first done for correlated matrices in [11, Proof of Corollary 2.5 in Section 5] but with $\mathrm{dist}(e_0, \mathrm{supp}\rho) \gtrsim 1$. The adjustments for $\mathrm{dist}(e_0, \mathrm{supp}\rho) \ge N^{\theta} \eta_{\mathfrak{f}}(e_0)$ are carried out in [36, Proof of Corollary 2.6] for the case of Wigner-type matrices (i.e. without correlations). This argument immediately translates to our setting, hence we omit the details for brevity.

Armed with band rigidity as in (2.17), the proof of Corollary 2.11 (b) is conducted in the same way as in [7, Proofs of Corollaries 1.10 and 1.11] or [36, Proof of Corollary 2.6]. □

Proof of Theorem 2.13 on cusp universality. Given the optimal local law in Theorem 2.8, universality at the cusp follows by the *three-step strategy*: The first step is the (model dependent) local law. The second step establishes universality for matrices with a small Gaussian component using the *Dyson Brownian Motion*, while the third step removes the Gaussian component via a comparison argument. The second and third step have already been worked out in the general correlated case in both the complex Hermitian [36] and real symmetric [32] symmetry class. More precisely, as explained in [32, Beginning of Section 3], once an appropriate local law for correlated matrices is available, the arguments in [32,36] directly yields the desired universality. Now, our Theorem 2.8 provides the necessary local law and thus cusp universality follows by application of [32,36]. □

4. Characteristic Flow: Proof of Proposition 3.6

First, we collect the necessary properties of the solution M_t to the time-dependent MDE (3.16).

Lemma 4.1 *(Preliminary bounds on M_t). Let $(A, \mathcal{S})$ be a data-pair satisfying the Assumptions 2.1, 2.4, and 2.5. Then there exists a threshold $T_* \sim 1$ such that for any terminal time $0 < T < T_*$, the solution M_t to the time-dependent MDE (3.16), with the terminal condition on the data pair $(A_T, \mathcal{S}_T) = (A, \mathcal{S})$, satisfies*

$$\|M_t(z)\| \lesssim 1, \quad c\rho_t(z) \le \Im M_t(z) \le C\rho_t(z), \tag{4.1}$$

uniformly in z with $\Re z \in \mathcal{I}$, where $\mathcal{I}$ is the set of admissible energies from (2.10). Here the second inequality holds in the sense of quadratic forms, with $1 \lesssim c \le C \lesssim 1$.

Essentially, at the terminal time $t = T$, the bounds (4.1) follow from the assumptions of the lemma, while at all other times $0 \le t < T$, the equations (3.17) guarantee that the data pair $(A_t, \mathcal{S}_t)$ constitutes only a small perturbation around $(A_T, \mathcal{S}_T)$. We give a more detailed proof of Lemma 4.1 in Appendix A.

Equipped with Lemma 4.1, we are ready to prove Proposition 3.6.

Proof of Proposition 3.6. We conduct the proof in the complex Hermitian case, the obvious modifications in the real symmetric case[8] are left to the reader. Throughout the proof we consider the step index k to be fixed, and hence omit it from the subscripts.

It suffices to prove that the resolvent G_t satisfies the local laws (3.1) with data $(\mathcal{D}_t^{\mathrm{abv}}, \xi + (\ell+1)\nu)$ for any fixed $t := t_{\mathrm{final}} \in [t_{k-1}, t_k]$ and $z \in \mathcal{D}_{t_{\mathrm{final}}}^{\mathrm{abv}}$, since uniformity in t and z can be obtained by a simple grid argument.[9] Let $t_{\mathrm{init}} := t_{k-1}$, and for all $t \in [t_{\mathrm{init}}, t_{\mathrm{final}}]$, let $z_t := \varphi_{t,t_{\mathrm{final}}}(z)$, where the map φ is defined in (3.19). It follows from Lemma 3.5 that $z_t \in \mathcal{D}_t^{\mathrm{abv}}$ for all $t \in [t_{\mathrm{init}}, t_{\mathrm{final}}]$. We denote $G_t := (H_t - z_t)^{-1}$, and $M_t := M_t(z_t)$, where M_t is the solution to (3.16).

Using Itô's formula, we deduce that for any deterministic $N \times N$ matrix B,

$$\mathrm{d}\big\langle (G_t - M_t)B\big\rangle = \bigg(\frac{1}{2}\big\langle (G_t - M_t)B\big\rangle + \big\langle G_t - M_t\big\rangle\big\langle G_t^2 B\big\rangle\bigg)\mathrm{d}t + \frac{1}{\sqrt{N}}\sum_{ab}\partial_{ab}\big\langle G_t B\big\rangle \mathrm{d}\big(\mathfrak{B}_t\big)_{ab}, \tag{4.2}$$

where $\partial_{ab} := \partial_{H_{ab,t}}$ denotes the partial derivative with respect to the matrix entry $H_{ab,t}$. In particular, for a fixed pair of deterministic vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$ with $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$, setting $B := N\boldsymbol{y}\boldsymbol{x}^*$ we obtain

$$\mathrm{d}\big(G_t - M_t\big)_{\boldsymbol{x}\boldsymbol{y}} = \bigg(\frac{1}{2}\big(G_t - M_t\big)_{\boldsymbol{x}\boldsymbol{y}} + \big\langle G_t - M_t\big\rangle\big(G_t^2\big)_{\boldsymbol{x}\boldsymbol{y}}\bigg)\mathrm{d}t + \frac{1}{\sqrt{N}}\sum_{ab}\partial_{ab}\big(G_t\big)_{\boldsymbol{x}\boldsymbol{y}}\mathrm{d}\big(\mathfrak{B}_t\big)_{ab}. \tag{4.3}$$

First, we prove that the resolvent $G_{t_{\mathrm{final}}}$ satisfies the isotropic local law and averaged local law in (3.1) for $B := I$ with data $(\mathcal{D}_{t_{\mathrm{final}}}^{\mathrm{abv}}, \xi + (\ell + \frac{1}{2})\nu)$. Define a set of deterministic

[8] For a detailed treatment of the real-symmetric case in the setting of standard Wigner matrices, we refer the reader to Section 4 of [27]. The only difference is the presence of an additional term in equation (4.5) due to $\sigma = 1$. However, its treatment follows identically to that of the other terms already present in the $\sigma = 0$ case-see, for instance, equations (4.28) and (4.27) in [27]. The necessary modifications for the more general ensembles considered here are entirely analogous.

[9] The grid argument relies on two straightforward observations: First, the resolvent $G_t(z)$ with $|\Im z| \ge N^{-1}$ – and, therefore, all quantities we consider – are Lipschitz continuous with a Lipschitz constant $\lesssim N^C$ for some $C > 0$ both in z and in t. Second, for any $C > 0$, the intersection of N^C-many very-high-probability events also occurs with very high probability. Therefore, a uniform very-high probability bounds are first established over a sufficiently fine N^{-C} grid in the domain of z or t, and then extended to the entire domain by Lipschitz continuity.

vectors $\mathcal{V} := \{\boldsymbol{x}, \boldsymbol{y}\}$. Define the stopping time τ

$$\begin{aligned}\tau := \inf&\left\{ t_{\mathrm{init}} \le t \le t_{\mathrm{final}} \;:\; \max_{\boldsymbol{u},\boldsymbol{v}\in\mathcal{V}} \left| \sqrt{\rho_t(z_t)^{-1} N\eta_t} \big(G_t - M_t\big)_{\boldsymbol{u}\boldsymbol{v}} \right| \ge N^{\xi+(\ell+\frac{1}{2})\nu} \right\} \\ &\wedge \inf\left\{ t_{\mathrm{init}} \le t \le t_{\mathrm{final}} \;:\; \left| N\eta_t \langle G_t - M_t\rangle \right| \ge N^{3\xi+3(\ell+\frac{1}{2})\nu} \right\},\end{aligned} \tag{4.4}$$

where we denote $\eta_t := \Im z_t > 0$.

Computing the quadratic variation of the martingale term in (4.2), we obtain

$$\begin{aligned}\left[\int_{t_{\mathrm{init}}}^{\cdot} \frac{1}{\sqrt{N}} \sum_{ab} \partial_{ab} \big\langle G_s \big\rangle \mathrm{d}\big(\mathfrak{B}_s\big)_{ab} \right]_{t\wedge\tau} &\le \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle (\Im G_s)^2 \big\rangle}{N^2 \eta_s^2} \mathrm{d}s \le \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle \Im G_s \big\rangle}{N^2\eta_s^3} \mathrm{d}s \\ &\le \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle \Im M_s \big\rangle + \frac{1}{2}\eta_s}{N^2\eta_s^3} \mathrm{d}s + \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle \Im G_s - \Im M_s \big\rangle}{N^2\eta_s^3} \mathrm{d}s\,.\end{aligned} \tag{4.5}$$

In the first step, we used that $\partial_{ab}\langle G_s\rangle = N^{-1}(G_s^2)_{ba}$ and employed a *Ward identity* $G_s G_s^* = \Im G_s/\eta_s$ twice. Moreover, in the penultimate step we used the norm bound $\|\Im G_s\| \le \eta_s^{-1}$, and in the ultimate step we used the fact that $\eta_s > 0$ in $\mathcal{D}_s^{\mathrm{abv}}$. We now estimate the two integrals in the last line of (4.5) separately. For the first integral, we use the imaginary part of (3.18) to obtain

$$\int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle \Im M_s \big\rangle + \frac{1}{2}\eta_s}{N^2\eta_s^3} \mathrm{d}s = \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{-\mathrm{d}\eta_s}{N^2\eta_s^3} \le \frac{1}{N^2\eta_{t\wedge\tau}^2}. \tag{4.6}$$

For the second integral, we use the definition (4.4) of the stopping time τ, and the imaginary part of (3.18) to deduce that

$$\begin{aligned}\left| \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle \Im G_s - \Im M_s \big\rangle}{N^2\eta_s^3} \mathrm{d}s \right| &\lesssim \left| \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{N^{3\xi+3(\ell+\frac{1}{2}\nu)}}{N^3\eta_s^4} \mathrm{d}s \right| \lesssim \left| \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{N^{3\xi+3(\ell+\frac{1}{2}\nu)}}{N^3\eta_s^4 \langle \Im M_s\rangle} \mathrm{d}\eta_s \right| \\ &\lesssim \frac{N^{-\varepsilon+3\xi+3(\ell+\frac{1}{2}\nu)}}{N^2\eta_{t\wedge\tau}^2},\end{aligned} \tag{4.7}$$

where in the last inequality we used that, for all $t_{\mathrm{init}} \le s \le t_{\mathrm{final}}$ it holds that $\langle \Im M_s\rangle N\eta_s \sim \rho_s(z_s) N\eta_s \gtrsim N^{\varepsilon}$ by (3.21).

Therefore, using the path-wise Burkholder-Davis-Gundy inequality (see Lemma 5.6 in [31] and Appendix B.6, Eq. (18) in [59]) and the fact that $\xi + K\nu \ll \varepsilon$, we deduce that, with very high probability

$$\max_{t_{\mathrm{init}}\le s\le t} \left| \int_{t_{\mathrm{init}}}^{s\wedge\tau} \frac{1}{\sqrt{N}} \sum_{ab} \partial_{ab}\big\langle G_s \big\rangle \mathrm{d}\big(\mathfrak{B}_s\big)_{ab} \right| \le \frac{N^{\nu}}{N\eta_{t\wedge\tau}}. \tag{4.8}$$

Next, using the Ward identity and the definition (4.4) of the stopping time τ, we obtain

$$\left| \frac{1}{2} + \big\langle G_s^2 \big\rangle \right| \le \frac{1}{2} + \frac{\big\langle \Im G_s \big\rangle}{\eta_s} \le -\frac{1}{\eta_s}\frac{\mathrm{d}\eta_s}{\mathrm{d}s} + \frac{N^{3\xi+3(\ell+\frac{1}{2})\nu}}{N\eta_s^2} \le -\frac{1}{\eta_s}\frac{\mathrm{d}\eta_s}{\mathrm{d}s}\left(1 + CN^{-\varepsilon+3\xi+3(\ell+\frac{1}{2})\nu}\right), \tag{4.9}$$

where in the last inequality we used the imaginary part of (3.18) and the bound $\langle \Im M_s\rangle N\eta_s \sim \rho_s(z_s)N\eta_s \gtrsim N^{\varepsilon}$ from (3.21). By integrating (4.2), it follows from the assumption of Proposition 3.6 at $t = t_{k-1} = t_{\mathrm{init}}$ and (4.8) that the bound

$$\big|\big\langle G_{t\wedge\tau} - M_{t\wedge\tau}\big\rangle\big| \le -\Big(1 + CN^{-\varepsilon+3\xi+(\ell+\frac{1}{2})\nu}\Big)\bigg(\int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big|\big\langle G_s - M_s\big\rangle\big|}{\eta_s}\frac{\mathrm{d}\eta_s}{\mathrm{d}s}\mathrm{d}s + \frac{N^{3\xi+3\ell\nu}}{N\eta_{t\wedge\tau}}\bigg), \tag{4.10}$$

holds with very high probability. Here we used that $\xi \ll \varepsilon$ from (3.3), and the assumption that $\ell\nu \le 2K\nu \ll \varepsilon$. Applying the Gronwall inequality yields the very-high-probability bound,

$$\big|\big\langle G_{t\wedge\tau} - M_{t\wedge\tau}\big\rangle\big| \le \frac{N^{3\xi+3(\ell+\frac{1}{4})\nu}}{N\eta_{t\wedge\tau}}, \tag{4.11}$$

uniformly in $t_{\mathrm{init}} \le t \le t_{\mathrm{final}}$.

Similarly, computing the quadratic variation of the martingale term in (4.3), we obtain

$$\begin{aligned}\bigg[\int_{t_{\mathrm{init}}}^{\cdot} \frac{1}{\sqrt{N}}\sum_{ab}\partial_{ab}\big(G_s\big)_{\boldsymbol{uv}}\mathrm{d}\big(\mathfrak{B}_s\big)_{ab}\bigg]_{t\wedge\tau} &\le \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big(\Im G_s\big)_{\boldsymbol{uu}}\big(\Im G_s\big)_{\boldsymbol{vv}}}{N\eta_s^2}\mathrm{d}s \\ &\lesssim \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\rho_s(z_s)^2}{N\eta_s^2}\bigg(1 + \frac{N^{\xi+(\ell+1)\nu}}{\sqrt{\rho_s(z_s)N\eta_s}}\bigg)^2\mathrm{d}s \\ &\lesssim \frac{\rho_{t\wedge\tau}(z_{t\wedge\tau})}{N\eta_{t\wedge\tau}},\end{aligned} \tag{4.12}$$

where we used the imaginary part of (3.20) to obtain $\rho_s(z_s) \sim \rho_{t\wedge\tau}(z_{t\wedge\tau})$. Therefore, using the path-wise Burkholder-Davis-Gundy inequality, we deduce the very-high-probability bound

$$\max_{t_{\mathrm{init}}\le s\le t}\bigg|\int_{t_{\mathrm{init}}}^{s\wedge\tau} \frac{1}{\sqrt{N}}\sum_{ab}\partial_{ab}\big(G_s\big)_{\boldsymbol{uv}}\mathrm{d}\big(\mathfrak{B}_s\big)_{ab}\bigg| \le N^{\nu}\sqrt{\frac{\rho_{t\wedge\tau}(z_{t\wedge\tau})}{N\eta_{t\wedge\tau}}}. \tag{4.13}$$

Moreover, within in the Schwarz estimate

$$|\big(G_s^2\big)_{\boldsymbol{uv}}| \le \sqrt{(|G_s|^2)_{\boldsymbol{uu}}(|G_s|^2)_{\boldsymbol{vv}}},$$

using the Ward identity, together with (4.1), (4.4), and the relations $\xi + K\nu \ll \varepsilon$, we deduce that

$$\big|\big(G_s^2\big)_{\boldsymbol{uv}}\big| \lesssim \rho_s(z_s)\bigg(1 + \frac{N^{\xi+(\ell+\frac{1}{2})\nu}}{\sqrt{\rho_s(z_s)N\eta_s}}\bigg) \lesssim \rho_s(z_s). \tag{4.14}$$

Therefore, from (3.21) and the bound (4.11), we conclude that

$$\bigg|\int_{t_{\mathrm{init}}}^{t\wedge\tau}\big\langle G_s - M_s\big\rangle\big(G_s^2\big)_{\boldsymbol{uv}}\mathrm{d}s\bigg| \lesssim \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{N^{3\xi+3(\ell+\frac{1}{2})\nu}}{N\eta_s}\frac{\rho_s(z_s)}{\eta_s}\mathrm{d}s \le \frac{N^{3\xi+3(\ell+1)\nu}}{N^{\varepsilon/2}}\sqrt{\frac{\rho_{t\wedge\tau}(z_{t\wedge\tau})}{N\eta_{t\wedge\tau}}}. \tag{4.15}$$

Integrating (4.3), and combining the assumption of Proposition 3.6 at time $t = t_{k-1} = t_{\mathrm{init}}$, (4.13) and (4.15) yields

$$\big|\big(G_{t\wedge\tau} - M_{t\wedge\tau}\big)_{\boldsymbol{uv}}\big| \le N^{\xi+(\ell+\frac{1}{4})\nu}\sqrt{\frac{\rho_{t\wedge\tau}(z)}{N\eta_{t\wedge\tau}}}, \tag{4.16}$$

uniformly in $t_{\mathrm{init}} \le t \le t_{\mathrm{final}}$, with very high probability, for all $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}$. Note that the term $\frac{1}{2}(G_t - M_t)_{\boldsymbol{uv}}$ on the right-hand side of (4.3) can be removed by differentiating $\mathrm{e}^{-t/2}(G_t - M_t)_{\boldsymbol{uv}}$ with the harmless prefactor $\mathrm{e}^{-t/2} = 1 + \mathcal{O}(T)$.

Hence, using (4.11) and (4.16), we conclude that $\tau = t_{\mathrm{final}}$ with very high probability, therefore establishing the isotropic local law and averaged local law in (3.1) for $B := I$ with data $(\mathcal{D}^{\mathrm{abv}}_{t_{\mathrm{final}}}, \xi + (\ell + \frac{1}{2})\nu)$.

For a general observable $B \in \mathbb{C}^{N\times N}$, we use the bound (4.11) as input to obtain the very-high-probability estimate

$$\Big|\big\langle G_s - M_s\big\rangle\big\langle G_s^2 B\big\rangle\Big| \le \frac{N^{3\xi+3(\ell+\frac{1}{4})\nu}}{N\eta_s}\,\frac{\big\langle \Im G_s\big\rangle^{1/2}\big\langle \Im G_s BB^*\big\rangle^{1/2}}{\eta_s} \lesssim \frac{N^{3\xi+3(\ell+\frac{1}{4})\nu}}{N\eta_s}\,\frac{\rho_s(z_s)}{\eta_s}\,\|B\|_{\mathrm{hs}}\,. \tag{4.17}$$

uniformly in $t_{\mathrm{init}} \le s \le t_{\mathrm{final}}$. Here, in the last step we used the isotropic bound (4.16) for the eigenvectors $\boldsymbol{v}_j$ of BB^*, corresponding to the eigenvalues $|\sigma_j|^2$, to conclude that, with very high probability,

$$\langle \Im GBB^*\rangle = \frac{1}{N}\sum_j |\sigma_j|^2 (\Im G)_{\boldsymbol{v}_j\boldsymbol{v}_j} \lesssim \rho_s(z_s)\|B\|_{\mathrm{hs}}^2. \tag{4.18}$$

Similarly, using (4.18), we estimate the quadratic variation of the corresponding martingale term in (4.2) for general B,

$$\left[\int_{t_{\mathrm{init}}}^{\cdot} \frac{1}{\sqrt{N}}\sum_{ab} \partial_{ab}\big\langle G_s B\big\rangle \mathrm{d}\big(\mathfrak{B}_s\big)_{ab}\right]_{t\wedge\tau} \lesssim \frac{N^{\nu}}{N^2\eta_{t\wedge\tau}^2}\,\|B\|_{\mathrm{hs}}\,. \tag{4.19}$$

Combining (4.2), (4.17), and (4.19), we conclude that the resolvent $G_{t_{\mathrm{final}}}$ satisfies the averaged local law in (3.1) with data $(\mathcal{D}^{\mathrm{abv}}_{t_{\mathrm{final}}}, \xi + (\ell + 1)\nu)$ for any $B \in \mathbb{C}^{N\times N}$. This concludes the proof of Proposition 3.6. □

5. Green Function Comparison: Proof of Proposition 3.7

The goal of this section is to prove Proposition 3.7 and thereby conclude the argument for the *zag* step of our proof. For simplicity and in order to avoid unnecessary complications, we will carry out the proof only in the real symmetric case; the complex-Hermitian case can be dealt with minor modifications.[10] and is thus omitted. Moreover, since throughout the argument the time t_k defined in (3.23) remains fixed, for the remainder of this section, we drop the superscript t_k from $\mathcal{D}^{\mathrm{abv}}_{t_k}$, ρ_{t_k}, and M_{t_k}. To further condense the notation, we abbreviate $\mathcal{D} := \mathcal{D}^{\mathrm{abv}}_{t_k}$ and $s_{\mathrm{final}} := s(\vartriangle t_k)$.

The proof will be conducted iteratively along vertical truncations of the domain $\mathcal{D}$, defined as

$$\mathcal{D}_\gamma \equiv \mathcal{D}^{\mathrm{abv}}_{t_k,\gamma} := \big\{z := E + \mathrm{i}\eta \in \mathcal{D} \equiv \mathcal{D}^{\mathrm{abv}}_{t_k} : \eta \ge N^{-1+\gamma}\}, \quad 0 < \gamma \le 1. \tag{5.1}$$

This is formalized in the following proposition, which we prove in Sect. 5.2.

[10] The notation in the cumulant expansion is slightly more involved in the complex case, as the real and imaginary parts are treated separately; see [35, Appendix C]

Proposition 5.1 *(Zag Bootstrap). Fix a constant $0 < \gamma_0 \leq 1$ and assume that the very-high-probability bounds on the matrix elements of the resolvent* (3.28)

$$\left|\left(G^s(z)\right)_{\boldsymbol{u}\boldsymbol{v}}\right| \lesssim 1, \quad \left|\left(\Im G^s(z)\right)_{\boldsymbol{u}\boldsymbol{u}}\right| \lesssim \rho(z), \tag{5.2}$$

hold uniformly in $z \in \mathcal{D}_{\gamma_0}$ and $s \in [0, s_{\text{final}}]$, for any deterministic $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^N$ with $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$.

Fix $\gamma_1 \geq \gamma_0 - \delta$ with $\delta < \mu$ satisfying $\delta \ll \xi$, and assume that for some $\nu > 0$ and $\ell \in \mathbb{N}$, the resolvent G^s satisfies the local laws (3.1) *with data $(\mathcal{D}_{\gamma_1}, \xi + \ell\nu)$ at time $s = s_{\text{final}}$. Then the resolvent G^s satisfies the local laws* (3.1) *with data $(\mathcal{D}_{\gamma_1}, \xi + (\ell+1)\nu)$ uniformly in $s \in [0, s_{\text{final}}]$.*

Armed with Proposition 5.1, we can easily conclude Proposition 3.7.

Proof of Proposition 3.7. The proof goes via induction in $\gamma(k) := 1 - k\delta$ by iteratively applying Proposition 5.1. As the base case, clearly, the estimates (5.2) hold for $\gamma_0 = \gamma(0) = 1$ as a direct consequence of the bounds $\|G^s(E + \mathrm{i}\eta)\| \leq \eta^{-1}$ and $\rho(E + \mathrm{i}\eta) \sim 1$ for $\eta \sim 1$. Moreover, the resolvent G^s satisfies the local laws (3.1) with data $(\mathcal{D}_{\gamma_1}, \xi + \ell\nu)$ with $\gamma_1 = \gamma(1)$ at time $s = s_{\text{final}}$ by assumption. Hence, the resolvent G^s satisfies the local laws (3.1) with data $(\mathcal{D}_{\gamma(1)}, \xi + (\ell+1)\nu)$ uniformly in $s \in [0, s_{\text{final}}]$ by Proposition 5.1. As a consequence, since $\xi + (\ell + 1)\nu \ll \varepsilon$, we have that the resolvent G^s satisfies the bounds (5.2) uniformly in $z \in \mathcal{D}_{\gamma(1)}$ and $s \in [0, s_{\text{final}}]$

As the induction step, assume now that for an integer $k \geq 1$ the resolvent G^s satisfies the bounds (5.2) uniformly in $z \in \mathcal{D}_{\gamma(k)}$ and $s \in [0, s_{\text{final}}]$. (Recall that, as above, G^s satisfies the local laws (3.1) with data $(\mathcal{D}_{\gamma(k)}, \xi + \ell\nu)$ at time $s = s_{\text{final}}$ by assumption.) Therefore, the resolvent G^s satisfies the local laws (3.1) with data $(\mathcal{D}_{\gamma(k+1)}, \xi + (\ell+1)\nu)$ uniformly in $s \in [0, s_{\text{final}}]$ by Proposition 5.1. Note that after $K' := \lceil (1 + \varepsilon)/\delta \rceil \sim 1$ steps, $\mathcal{D}_{\gamma(K')} = \mathcal{D}$ and we have hence proven Proposition 3.7. □

It thus remains to prove Proposition 5.1. We begin by collecting several preliminaries in Sect. 5.1. Afterwards, in Sect. 5.2 we give the proof of Proposition 5.1 based on average and isotropic *Gronwall estimates*. These bounds are proven in Sects. 5.3.1 and 5.3.2, respectively.

5.1. Preliminaries. In order to perform the GFT, i.e., compare initial and final W's, given by $W^t = H^t - A$ with H^t being the solution to (3.9), we employ Itô's formula: For a C^2-function $f(W^t)$, it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E} f(W^t) = -\frac{1}{2}\mathbb{E}\sum_{\alpha} w_\alpha(t)(\partial_\alpha f)(W^t) + \frac{1}{2N}\sum_{\alpha,\beta} \kappa_t(\alpha,\beta)\mathbb{E}(\partial_\alpha\partial_\beta f)(W^t), \tag{5.3}$$

where $\kappa_t(\alpha, \beta)$ denotes the (normalized, recall (2.5)) second order cumulant of $w_\alpha(t)$ and $w_\beta(t)$, the matrix entries of W^t. The first summand on the rhs. of (5.3) can now be further treated by cumulant expansion, which is the first key ingredient for our proof.

Proposition 5.2 *(Multivariate cumulant expansion; cf. Proposition 3.2 in [35] and Lemma 3.1 in [43]). Let $f : \mathbb{R}^{N\times N} \to \mathbb{C}$ be a L times differentiable function with bounded derivatives. Let W be a random matrix, whose normalized cumulants satisfy*

Assumption 2.3. Then, for any index $\alpha_0 \in [N]^2$ it holds that (recall the definition of the neighborhood set $\mathcal{N}$ from Assumption 2.3)

$$\mathbb{E} w_{\alpha_0} f(W) = \sum_{k=0}^{L-1} \sum_{\boldsymbol{\alpha} \in \mathcal{N}(\alpha_0)^k} \frac{\kappa(\alpha_0, \boldsymbol{\alpha})}{N^{(k+1)/2} k!} \mathbb{E}(\partial_{\boldsymbol{\alpha}} f)(W) + \Omega_L(f, \alpha_0), \tag{5.4}$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)$ and $\partial_{\boldsymbol{\alpha}} = \partial_{w_{\alpha_1}} ... \partial_{w_{\alpha_k}}$ for $k \geq 1$, and for $k = 0$ is considered as the function f itself. Moreover, the error term in (5.4) satisfies

$$\left|\Omega_L(f, \alpha_0)\right| \lesssim \frac{C_L}{N^{(L+1)/2}} \sum_{\boldsymbol{\alpha} \in \mathcal{N}(\alpha_0)^L} \sup_{\lambda \in [0,1]} \left(\mathbb{E}\left|(\partial_{\boldsymbol{\alpha}} f)(\lambda W|_{\mathcal{N}(\alpha_0)} + W|_{[N]^2 \setminus \mathcal{N}(\alpha_0)})\right|^2\right)^{1/2}, \tag{5.5}$$

for some constant $C_L > 0$ depending only on L. The notation $W|_{\mathcal{N}}$ for $\mathcal{N} \subset [N]^2$ in (5.5) refers to the matrix which equals W at all entries $\alpha \in \mathcal{N}$ and is zero otherwise.

Note that the $k = 1$ term in the expansion of the first summand on the rhs. of (5.3) exactly cancels the second summand on the rhs. of (5.3). For Proposition 5.2 being practically applicable we need to control (i) every order of the expansion, and (ii) the truncation term Ω. These will be guaranteed by Assumption 2.3 above.

The second key input required for the GFT argument is the following monotonicitiy estimate on resolvents.

Lemma 5.3 *(Monotonicity estimate). Fix a constant $0 < \gamma_0 \leq 1$ and assume that the very-high-probability bounds (5.2) hold uniformly in $z \in \mathcal{D}_{\gamma_0}$ and $s \in [0, s_{\text{final}}]$, for any deterministic $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^N$ with $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$.*

Fix $\gamma_1 \geq \gamma_0 - \delta$. Then, we have

$$|G^s(E + \mathrm{i}\eta_1)_{\boldsymbol{u}\boldsymbol{v}}| \lesssim \frac{\eta_0}{\eta_1}, \qquad |\Im G^s(E + \mathrm{i}\eta_1)_{\boldsymbol{u}\boldsymbol{u}}| \lesssim \rho(E + \mathrm{i}\eta_0)\frac{\eta_0}{\eta_1}, \tag{5.6}$$

with very high probability, uniformly in $z := E + \mathrm{i}\eta_1 \in \mathcal{D}_{\gamma_1}$ for any $\eta_0 \geq N^{-1+\gamma_0} \vee \eta_1$, time $s \in [0, s_{\text{final}}]$, and for any deterministic vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^N$ with $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$.

We defer the proof of Lemma 5.3 to Appendix A.

5.2. Gronwall estimates: Proof of Proposition 5.1. In this section, we provide the proof of Proposition 5.1 based on two Gronwall estimates, formulated in Propositions 5.4–5.5 below that will be proven in the next subsection. The isotropic part of Proposition 5.1 will be concluded in a self-contained way, based entirely on the *isotropic Gronwall estimate* in Proposition 5.4. Its conclusion in (5.11) then serves as an input for the *average Gronwall estimate* in Proposition 5.5.

Proof of Proposition 5.1. We remind the reader that, as pointed out below (3.10), the deterministic approximation M is time-independent in the *zag* step.

Proposition 5.4 *(Isotropic Gronwall estimate). Assume the conditions of Proposition 5.1. Fix $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$ of bounded norm, $z := E + \mathrm{i}\eta_1 \in \mathcal{D}_{\gamma_1}$ and $\eta_0 \geq N^{-1+\gamma_0} \vee \eta_1$ such that $\eta_0/\eta_1 \leq N^{\delta}$. For $s \in [0, s_{\text{final}}]$, define*

$$S_s := \left(G^s(E + \mathrm{i}\eta_1) - M(E + \mathrm{i}\eta_1)\right)_{\boldsymbol{x}\boldsymbol{y}}. \tag{5.7}$$

Then, for any (large) even $p \in \mathbf{N}$, it holds that

$$\left|\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}|S_s|^p\right| \lesssim \left(1 + N^{10\delta}\sqrt{\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}}\right)\left[\mathbb{E}|S_s|^p + (\Psi(\eta_1))^p\right], \tag{5.8}$$

uniformly in $s \in [0, s_{\mathrm{final}}]$, bounded $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$, and $z \in \mathcal{D}_{\gamma_1}$. Here, for $\eta \in [\eta_0, \eta_1]$, we denoted

$$\Psi(\eta) := \sqrt{\frac{\rho(E+\mathrm{i}\eta)}{N\eta}}. \tag{5.9}$$

By Gronwall's lemma, uniformly in $s \in [0, s_{\mathrm{final}}]$, from (5.8) we find that

$$\begin{aligned}\mathbb{E}|S_s|^p &\lesssim \exp\left(\left(1 + N^{10\delta}\sqrt{\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}}\right)(s_{\mathrm{final}} - s)\right)\left[\mathbb{E}|S_{s_{\mathrm{final}}}|^p + (\Psi(\eta_1))^p\right] \\ &\lesssim \exp(N^{-\xi/10})\left[\mathbb{E}|S_{s_{\mathrm{final}}}|^p + (\Psi(\eta_1))^p\right] \lesssim \mathbb{E}|S_{s_{\mathrm{final}}}|^p + (\Psi(\eta_1))^p.\end{aligned} \tag{5.10}$$

Here we used that $\rho(E+\mathrm{i}\eta_0)/\eta_0 \lesssim N^{k\delta}/T$ by (3.21), $s_{\mathrm{final}} \lesssim N^{-(k-1)\delta}T$ by (3.15), $T \sim N^{-\xi/4}$ from (3.22), and $\delta \ll \xi$ by (3.3). We point out that in (5.10), we use the final value rather than the initial value, as is more customary in a typical Gronwall argument, since in the zigzag strategy, illustrated in Figure 3, the endpoint of the flow is the known object.

To estimate $\mathbb{E}|S_{s_{\mathrm{final}}}|^p$, recall that the resolvent G^s satisfies the isotropic local law in (3.1) with data $(\mathcal{D}_{\gamma_1}, \xi + \ell\nu)$ at $s = s_{\mathrm{final}}$. Therefore, since p in (5.10) was arbitrary, we find that

$$\left|\left(G^s(z) - M(z)\right)_{\boldsymbol{x}\boldsymbol{y}}\right| \le N^{\xi+(\ell+1)\nu}\sqrt{\frac{\rho(z)}{N\eta_1}}, \tag{5.11}$$

uniformly in $z := E + \mathrm{i}\eta_1 \in \mathcal{D}_{\gamma_1}$, $s \in [0, s_{\mathrm{final}}]$, and bounded $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$, with very high probability.

This proves the isotropic part of Proposition 5.1 and we are left with the average part.

Proposition 5.5 *(Average Gronwall estimate).*
Fix $B \in \mathbb{C}^{N\times N}$ of bounded Hilbert–Schmidt norm, $\|B\|_{\mathrm{hs}} \le 1$, $z := E + \mathrm{i}\eta \in \mathcal{D}_{\gamma_1}$, and $\eta_0 \ge N^{-1+\gamma_0} \vee \eta_1$ such that $\eta_0/\eta_1 \le N^{\delta}$. For $s \in [0, s_{\mathrm{final}}]$, define

$$R_s := \langle (G^s(E+\mathrm{i}\eta_1) - M(E+\mathrm{i}\eta_1))B\rangle. \tag{5.12}$$

Moreover, suppose that (5.11) holds uniformly in $z := E + \mathrm{i}\eta_1 \in \mathcal{D}_{\gamma_1}$, $s \in [0, s_{\mathrm{final}}]$, and bounded $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$. Then, for any (large) even $p \in \mathbf{N}$ it holds that

$$\left|\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}|R_s|^p\right| \lesssim \left(1 + N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\right)\left[\mathbb{E}|R_s|^p + \left(\frac{N^{3\xi}}{N\eta_1}\right)^p\right], \tag{5.13}$$

uniformly in $s \in [0, s_{\mathrm{final}}]$, bounded $B \in \mathbb{C}^{N\times N}$, and $z \in \mathcal{D}_{\gamma_1}$.

Analogously to (5.10), by Gronwall's lemma, uniformly in $s \in [0, s_{\text{final}}]$, we find that

$$\begin{aligned} \mathbb{E}|R_s|^p &\lesssim \exp\left(\left(1 + N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\right)(s_{\text{final}} - s)\right)\left[\mathbb{E}|R_{s_{\text{final}}}|^p + \left(\frac{N^{3\xi}}{N\eta_1}\right)^p\right] \\ &\lesssim \exp(N^{-\delta})\left[\mathbb{E}|R_{s_{\text{final}}}|^p + (\Psi(\eta_1))^p\right] \lesssim \mathbb{E}|R_{s_{\text{final}}}|^p + \left(\frac{N^{3\xi}}{N\eta_1}\right)^p . \end{aligned} \tag{5.14}$$

Here we used that $\rho(E+\mathrm{i}\eta_0)/\eta_0 \lesssim N^{k\delta}/T$ by (3.21), $s_{\text{final}} \lesssim N^{-(k-1)\delta}T$ by (3.15), $T \sim N^{-\xi/4}$ by (3.22), and $\delta \ll \xi$ by (3.3). Note that the small prefactor $N^{-2\delta}$ in (5.13) is absolutely essential, unlike in the isotropic case (5.10), where a large prefactor $N^{10\delta}$ is affordable thanks to the square root. The linear appearance of ρ/η in (5.13) is only due to fact that we estimate B in terms of its Hilbert–Schmidt norm $\|B\|_{\text{hs}}$; cf. the estimate in (5.21). For observables with $\|B\| \sim \|B\|_{\text{hs}}$, such as the identity matrix $B = \mathbf{1}$, the linear dependence on ρ/η can be improved to a $\sqrt{\rho/\eta}$. We exploit this fact in (6.23) below.

Recall that the resolvent G^s satisfies the average local law in (3.1) with data $(\mathcal{D}_{\gamma_1}, \xi + \ell\nu)$ at $s = s_{\text{final}}$. Therefore, since p in (5.14) was arbitrary, we find that

$$\left|\left\langle\left(G^s(z) - M(z)\right)B\right\rangle\right| \leq \frac{N^{3(\xi+(\ell+1)\nu)}}{N\eta_1},$$

uniformly in $z := E + \mathrm{i}\eta_1 \in \mathcal{D}_{\gamma_1}$, $s \in [0, s_{\text{final}}]$, and $B \in \mathbb{C}^{N\times N}$ with $\|B\|_{\text{hs}} \leq 1$, with very high probability.

This concludes the proof of Proposition 5.1. □

5.3. Cumulant expansion: Proofs of Propositions 5.5 and 5.4. The proofs of Propositions 5.4–5.5 are based on the multivariate cumulant expansion from Proposition 5.2 and the monotonicity estimate from Lemma 5.3. We begin by proving the average Gronwall estimate in Proposition 5.5. Moreover, we will henceforth omit the superscript s from the resolvent G^s.

5.3.1. Average case

Proof of Proposition 5.5. Throughout the proof, we will assume that $\|B\|_{\text{hs}} \lesssim 1$. By (5.3) for R_s we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}|R_s|^p = -\frac{1}{2}\mathbb{E}\sum_{\alpha_1} w_{\alpha_1}(s)(\partial_{\alpha_1}|R_s|^p) + \frac{1}{2}\sum_{\alpha_1,\alpha_2}\kappa_s(\alpha_1,\alpha_2)\mathbb{E}\left[\partial_{\alpha_1}\partial_{\alpha_2}|R_s|^p\right], \tag{5.15}$$

where $w_{\alpha_i}(s)$ is the α_i-th entry of W_s, $\kappa_s(\alpha_1, \alpha_2, ...)$ is a joint normalized cumulant of $w_{\alpha_1}(s), w_{\alpha_2}(s), ...$ and $\partial_{\alpha_i} = \partial_{w_{\alpha_i}(s)}$ denotes the partial derivative in the direction of $w_{\alpha_i}(s)$.

The first term on the rhs. of (5.15) can now be expanded by means of Proposition 5.2:

$$\mathbb{E}\left[w_{\alpha_1}(s)(\partial_{\alpha_1}|R_s|^p)\right] = \sum_{k=0}^{L-1}\sum_{\boldsymbol{\alpha}\in\mathcal{N}(\alpha_1)^k}\frac{\kappa_s(\alpha_1,\boldsymbol{\alpha})}{N^{(k+1)/2}\,k!}\mathbb{E}\left[\partial_{\alpha_1}\partial_{\boldsymbol{\alpha}}|R_s|^p\right] + \Omega_L\,. \tag{5.16}$$

Since L derivatives of $|R_s|^p$ create L additional resolvent matrix elements (where each of them is bounded with the aid of Lemma 5.3) and using that $|\mathcal{N}(\alpha_1)| \lesssim N^{1/2-\mu}$ by Assumption 2.3 (ii), the error term Ω_L can be estimated as[11]

$$|\Omega_L| \lesssim N^{-\frac{L+1}{2}} N^{L(1/2-\mu)} N^{(p+L)\delta} \lesssim N^{2p\delta+L(\delta-\mu)}. \tag{5.17}$$

Using the relation $\mu > \delta$ from (3.3) and $L := \lceil((1+\delta)p+2)/(\mu-\delta)\rceil$, we see that $|\Omega_L| \le N^{-2}(N\eta_1)^{-p}$ (the factor N^{-2} is needed to bound the summation over α_1 in (5.15)). With this choice of L, the error term Ω_L will henceforth be ignored.

Plugging (5.16) into (5.15) and using that the $k = 0$ term is zero by $\kappa_s(\alpha_1) = \mathbb{E}w_{\alpha_1}(s) = 0$, and that the $k = 1$ term in (5.16) cancels the second term on the rhs. of (5.15), we obtain

$$\left|\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}|R_s|^p\right| \lesssim \left|\sum_{k=2}^{L-1}\sum_{\alpha_1}\sum_{\boldsymbol{\alpha}\in\mathcal{N}(\alpha_1)^k} \frac{\kappa_s(\alpha_1,\boldsymbol{\alpha})}{N^{(k+1)/2}\,k!}\mathbb{E}(\partial_{\alpha_1}\partial_{\boldsymbol{\alpha}}|R_s|^p)\right| + \left(\frac{1}{N\eta_1}\right)^p. \tag{5.18}$$

We will now first estimate the third order cumulant terms (i.e. those with $k = 2$ in (5.18)), as these are the most delicate, and afterwards turn to the higher order ones that can be handled by simple power counting with a little twist due to the Hilbert–Schmidt norm of the observable B. Moreover, we drop the time dependence of R_s and κ_s whenever it does not lead to confusion. We point out that Assumption 2.3 also holds for W^s from (3.9), uniformly in $s \in [0,\infty)$. Indeed, adding an independent Gaussian random matrix to W_0 has no effect on cumulants of order $k \ge 3$ (by Gaussianity) and leaves the first two joint moments as well as the independence property of Assumption 2.3 (ii) invariant (the covariance tensor Σ is trivial beyond the range $\mathcal{N}(\alpha_1)$) by construction (3.9). In particular, we can freely extend the summation over $\boldsymbol{\alpha} \in \mathcal{N}(\alpha_1)^k$ in (5.18) to $\boldsymbol{\alpha} \in ([N]^2)^k$ and combine the latter two summations in (5.18) into $\sum_{\alpha_1,\boldsymbol{\alpha}}$.

Now, for the third order cumulant terms, we aim to control

$$\left|N^{-3/2}\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)\mathbb{E}(\partial_{\alpha_1}\partial_{\alpha_2}\partial_{\alpha_3}|R|^p)\right|,$$

which, after employing the Leibniz rule, can be broken up into terms of the form $(\partial_\alpha^3 R)|R|^{p-1}$, $(\partial_\alpha R)(\partial_\alpha^2 R)|R|^{p-2}$, and $(\partial_\alpha R)^3|R|^{p-3}$. To further ease the notation, here and in the following, we neglect the difference between R and $\overline{R}$, as these will be estimated in a completely analogous way.

[11] To be precise, note some of the $p+L$ resolvents in the error term Ω_L are actually resolvents of the random matrix $W^{(\lambda)} := \lambda W|_{\mathcal{N}(\alpha_0)} + W|_{[N]^2\setminus\mathcal{N}(\alpha_0)}$ (recall (5.5)) and we need to guarantee their boundedness as well, uniformly in $\lambda \in [0,1]$. We perform a resolvent expansion of $G^{(\lambda)} := (A + W^{(\lambda)} - z)^{-1}$ up to some order $\tilde{m} \in \mathbf{N}$ around $G^{(1)}$ whose boundedness is known. For each $G^{(\lambda)}$, the m^{th} order term in this expansion can be bounded by $N^\delta N^{m(\delta-\mu+\nu)}$ with the aid of Lemma 5.3 (to bound $G^{(1)}$ isotropically) and using the norm estimate $\|W|_{\mathcal{N}(\alpha_0)}\| \le N^{-\mu+\nu}$, w.v.h.p. for any $\nu > 0$, which is a consequence of Assumption 2.3 (ii). By a simple norm bound $\|G^{(\lambda)}\| \le \eta^{-1}$, the last truncation term in the resolvent expansion admits the bound $N^\delta N^{\tilde{m}(\delta-\mu+\nu)}\eta^{-1}$. Therefore, since η depends at most polynomially on N and $\mu > \delta + \nu$ for some $\nu > 0$ small enough, the resolvent expansion can be truncated at finite order, leaving us with the bound N^δ for every matrix element of $G^{(\lambda)}$ employed in (5.17), uniformly in $\lambda \in [0,1]$.

We begin with the terms of the form $(\partial_\alpha^3 R)|R|^{p-1}$, which requires the bound (2.8) in Assumption 2.3 (i). Writing $\langle GB\rangle = N^{-1}\sum_j (GB)_{jj}$ and identifying $\alpha_i \equiv (a_i, b_i) \in [N]^2$, we aim to estimate (ignoring the $|R|^{p-1}$-factor)

$$\begin{aligned} N^{-5/2}&\left|\sum_{j,\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)G_{ja_1}G_{b_1a_2}G_{b_2a_3}(GB)_{b_3j}\right| \\ &= N^{-5/2}\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)G_{b_1a_2}G_{b_2a_3}(GBG)_{b_3a_1}\right|. \end{aligned}$$

For both $G_{b_1a_2}$ and $G_{b_2a_3}$ we write $G_{ba} = M_{ba} + (G-M)_{ba}$ and use $\|M\| \lesssim 1$ for the M-term and the bound (5.11) for the $(G-M)$-term. In particular (recalling the notation (5.9)),

$$\begin{aligned} N^{-5/2}&\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)M_{b_1a_2}(G-M)_{b_2a_3}(GBG)_{b_3a_1}\right| \\ &\lesssim N^{-5/2}\,N^{\xi+(\ell+1)\nu}\,\Psi(\eta_1)\sum_{\alpha_1,\alpha_2,\alpha_3}|\kappa(\alpha_1,\alpha_2,\alpha_3)|\,|(GBG)_{b_3a_1}| \\ &\lesssim N^{-5/2}\,N^{\xi+(\ell+1)\nu}\,\Psi(\eta_1)\sum_{\alpha_1,\alpha_2,\alpha_3}|\kappa(\alpha_1,\alpha_2,\alpha_3)|\,|(GBB^*G^*)_{b_3b_3}|^{1/2}|(GG^*)_{a_1a_1}|^{1/2} \\ &\lesssim N^{\xi+(\ell+1)\nu}\,\Psi(\eta_1)^2\left\|\sum_{\alpha_2}|\kappa(*,\alpha_2,*)|\right\|\,\langle GG^*BB^*\rangle^{1/2} \lesssim N^{-\delta}\sqrt{\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}}\,\frac{N^{3\xi}}{N\eta_1}, \end{aligned} \tag{5.19}$$

with very high probability. In the second step, we used the Schwarz inequality. In the penultimate inequality, we employed (2.6) and used

$$\begin{aligned} N^{-2}\sum_{a_3,b_3}(GBB^*G^*)_{b_3b_3} &= \langle GG^*BB^*\rangle \quad \text{and} \\ N^{-3}\sum_{a_1b_1}(GG^*)_{a_1a_1} &= \frac{\langle \Im G\rangle}{N\eta_1} \lesssim \frac{\rho(E+\mathrm{i}\eta_1)}{N\eta_1} = \Psi(\eta_1), \end{aligned}$$

with very high probability, where in the second relation we additionally used a Ward identity and the already established isotropic law in the form $(\Im G)_{\boldsymbol{e}_i\boldsymbol{e}_i} \lesssim (\Im M)_{\boldsymbol{e}_i\boldsymbol{e}_i} \lesssim \rho(E+\mathrm{i}\eta_1)$. Finally, in the ultimate step, similarly to (4.18), we used (2.6), the Ward identity, the spectral decomposition of BB^*, and (5.11) together with $(\Im M)_{\boldsymbol{v}\boldsymbol{v}} \lesssim \rho(E+\mathrm{i}\eta_1)$ by (4.1), to obtain

$$\langle GG^*BB^*\rangle = \frac{1}{N\eta_1}\sum_j |\sigma_j|^2(\Im G)_{\boldsymbol{v}_j\boldsymbol{v}_j} \lesssim \frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\|B\|_{\mathrm{hs}}^2, \tag{5.20}$$

and used $\delta \ll \xi$ by (3.3), and the fact that $\nu > 0$ is arbitrarily small. Note that the small factor $N^{-\delta}$ in the last line of (5.19) is balanced by an additional N^{ξ}. We stress that here and in the following we estimate $\rho(E+\mathrm{i}\eta_1)/\eta_1 \leq N^{2\delta}\rho(E+\mathrm{i}\eta_0)/\eta_0$ in order to conveniently use that $(\rho(E+\mathrm{i}\eta_0)/\eta_0)s_{\mathrm{final}} \lesssim N^{\delta}$ as discussed below (5.14). The terms with $(G-M)_{b_1a_2}G_{b_2a_3}$ and $(G-M)_{b_1a_2}(G-M)_{b_2a_3}$ are treated analogously and we are thus left with the $M_{b_1a_2}M_{b_2a_3}$-term. Here, using the $|||\kappa|||_3^{\mathrm{av}}$ norm from (2.8), we estimate

$$\begin{aligned} & N^{-5/2}\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)M_{b_1a_2}M_{b_2a_3}(GBG)_{b_3a_1}\right| \\ &\quad \le N^{-1}|||\kappa|||_3^{\mathrm{av}}\|M\|^2\|GBG\|_{\mathrm{hs}} \lesssim \eta_1^{-1/2}\frac{1}{N\eta_1}\langle\Im GBB^*\rangle^{1/2} \\ &\quad \lesssim N^{-\delta}\sqrt{\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}}\,\frac{N^{3\xi}}{N\eta_1}, \end{aligned} \tag{5.21}$$

with very high probability. In the penultimate step we used the definition of $\|\cdot\|_{\mathrm{hs}}$ together with a Ward identity and the trivial bound $\|G\|\le\eta_1^{-1}$; in the last step we employed (5.20) and $\eta_0/\eta_1\le N^\delta$ together with monotonicity of $\eta\mapsto\eta\rho(E+\mathrm{i}\eta)$ and $\delta\ll\xi$. Hence, by two Young inequalities , we thus find

$$\begin{aligned} & \left|N^{-3/2}\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)\mathbb{E}\big[(\partial_{\alpha_1}\partial_{\alpha_2}\partial_{\alpha_3}R)|R|^{p-1}\big]\right| \\ &\quad \lesssim \left(1+N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\right)\left[\mathbb{E}|R|^p+\left(\frac{N^{3\xi}}{N\eta_1}\right)^p\right], \end{aligned} \tag{5.22}$$

where we overestimated $N^{-\delta}\sqrt{\rho/\eta_0}\lesssim 1+N^{-2\delta}\rho/\eta_0$.

Next, we turn to terms of the form $(\partial_\alpha R)(\partial_\alpha^2 R)|R|^{p-2}$. Similarly to (5.19), using (2.6) for $k=3$, we find

$$\begin{aligned} & N^{-7/2}\left|\sum_{j,k,\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)G_{ja_1}G_{b_1a_2}(GB)_{b_2j}G_{ka_3}(GB)_{b_3k}\right| \\ & \lesssim N^{-7/2}\sum_{\alpha_1,\alpha_2,\alpha_3}|\kappa(\alpha_1,\alpha_2,\alpha_3)|\,|(GBG)_{b_3a_3}|\,|(GBG)_{b_2a_1}| \\ & \lesssim N^{-7/2}\sqrt{\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}}\sum_{\alpha_1,\alpha_2,\alpha_3}|\kappa(\alpha_1,\alpha_2,\alpha_3)|\,|(GBG)_{b_3a_3}|\sqrt{(GBB^*G^*)_{b_2b_2}} \\ & \lesssim N^{-7/2}\sqrt{\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}}\,|||\kappa|||_3\sqrt{\sum_{b_3,a_3}|(GBG)_{b_3a_3}|^2}\sqrt{\sum_{b_2a_2}(GBB^*G^*)_{b_2b_2}} \\ & \lesssim \frac{\rho(E+\mathrm{i}\eta_1)^{1/2}}{N^2\eta_1^2}\langle\Im GB\Im GB^*\rangle^{1/2}\langle\Im GBB^*\rangle^{1/2}\lesssim N^{-\delta}\sqrt{\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}}\left(\frac{N^{3\xi}}{N\eta_1}\right)^2, \end{aligned} \tag{5.23}$$

with very high probability. To go to the third line, we used a Schwarz inequality and the estimate $(GG^*)_{a_1a_1}\lesssim\rho/\eta_1$ w.v.h.p. (as follows by a Ward identity and (5.11)). In the penultimate step, we again used several Ward identities. In the last step we used $\langle\Im GB\Im GB^*\rangle\le\langle\Im GBB^*\rangle/\eta_1$ and (5.20) together with $\eta_0/\eta_1\le N^\delta$, monotonicity of $\eta\mapsto\eta\rho(E+\mathrm{i}\eta)$, and $\delta\ll\xi$ by (3.3). Hence, again by Young's inequality and overestimating $N^{-\delta}\sqrt{\rho/\eta_0} \lesssim 1 + N^{-2\delta}\rho/\eta_0$, we find

$$\begin{aligned} &\left| N^{-3/2} \sum_{\alpha_1,\alpha_2,\alpha_3} \kappa(\alpha_1,\alpha_2,\alpha_3)\mathbb{E}\big[(\partial_{\alpha_1}\partial_{\alpha_2}R)(\partial_{\alpha_3}R)|R|^{p-2}\big]\right| \\ &\quad \lesssim \left(1 + N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\right)\left[\mathbb{E}|R|^p + N^{\xi}\left(\frac{N^{\delta}}{N\eta_1}\right)^p\right]. \end{aligned} \tag{5.24}$$

Finally, we estimate terms of the form $(\partial_\alpha R)^3|R|^{p-3}$, which are the most critical ones, since they necessarily contribute the $N^{-2\delta}\rho/\eta$ factor as we estimate B by its Hilbert–Schmidt norm $\|B\|_{\mathrm{hs}}$. For terms of the form $(\partial_\alpha R)^3|R|^{p-3}$, similarly to (5.19) and (5.23), we find

$$\begin{aligned} &N^{-9/2}\left|\sum_{j,k,\ell,\alpha_1,\alpha_2,\alpha_3} \kappa(\alpha_1,\alpha_2,\alpha_3)G_{ja_1}(GB)_{b_1 j}G_{ka_2}(GB)_{b_2 k}G_{\ell a_3}(GB)_{b_3\ell}\right| \\ &\quad \lesssim N^{-9/2}\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\|B\| \sum_{\alpha_1,\alpha_2,\alpha_3} |\kappa(\alpha_1,\alpha_2,\alpha_3)|\left|(GBG)_{b_2a_2}\right|\left|(GBG)_{b_3a_3}\right| \\ &\quad \lesssim N^{-9/2}\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\|B\|\,|||\kappa|||_3 \sum_{a,b}\left|(GBG)_{ab}\right|^2 \lesssim N^{-7/2}\frac{\rho(E+\mathrm{i}\eta_1)^2}{\eta_1^4}\|B\|\|B\|_{\mathrm{hs}}^2 \\ &\quad \lesssim N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\left(\frac{N^{3\xi}}{N\eta_1}\right)^3 . \end{aligned} \tag{5.25}$$

To go to the second line, we used that

$$|(GBG)_{ab}| \le \|B\|\sqrt{(GG^*)_{aa}(GG^*)_{bb}} \lesssim \|B\|\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\,, \tag{5.26}$$

by a Schwarz inequality, a Ward identity and (5.11). In the third line we estimated

$$\sum_{a,b}\left|(GBG)_{ab}\right|^2 = \frac{N}{\eta_1^2}\langle \Im GB\Im GB^*\rangle \lesssim \frac{N\rho(E+\mathrm{i}\eta_1)}{\eta_1^3}\|B\|_{\mathrm{hs}}^2\,, \tag{5.27}$$

with very high probability, by means of Ward identities and (5.20). To go to the fourth line, we used $\|B\| \le \sqrt{N}\|B\|_{\mathrm{hs}}$ and the fact that $\delta \ll \xi$ by (3.3), together with $\eta_0/\eta_1 \le N^{\delta}$ and monotonicity of $\eta \mapsto \eta\rho(E+\mathrm{i}\eta)$.

Hence, (5.25) together with Young's inequality implies that

$$\begin{aligned} &N^{-3/2}\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\mathbb{E}\big[(\partial_{\alpha_1}R)(\partial_{\alpha_2}R)(\partial_{\alpha_3}R)|R|^{p-3}\big]\right| \\ &\quad \lesssim \left(1 + N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\right)\left[\mathbb{E}|R|^p + \left(\frac{N^{3\xi}}{N\eta_1}\right)^p\right]. \end{aligned} \tag{5.28}$$

For the higher order terms in (5.18) with $n = k+1 \ge 4$ we aim to estimate

$$\left| N^{-n/2}\sum_{\alpha_1,\dots,\alpha_n}\kappa(\alpha_1,\dots,\alpha_n)\mathbb{E}\big[\partial_{\alpha_1}\dots\partial_{\alpha_n}|R|^p\big]\right| .$$

In case that the n derivatives are distributed on $k \in [n]$ factors of R, we find that, for $n_\ell \in \mathbf{N}$ with $\sum_{\ell=1}^k n_\ell = n$ and identifying $(\alpha_i)_{i\in[n]} \equiv \big((a_{\ell_i}, b_{\ell_i})\big)_{i\in[n_\ell],\ell\in[k]}$,

$$\begin{aligned}
&\Bigg|N^{-n/2}N^{-k}\sum_{j_1,\dots,j_k}\sum_{\alpha_1,\dots,\alpha_n}\kappa(\alpha_1,\dots,\alpha_n)\prod_{\ell=1}^k\big(G_{j_\ell a_{\ell_1}}G_{b_{\ell_1}a_{\ell_2}}\dots G_{b_{\ell_{n_\ell-1}}a_{\ell_{n_\ell}}}(GB)_{b_{\ell_{n_\ell}}j_\ell}\big)\Bigg|\\
&\lesssim N^{-n/2}N^{-k}\sum_{\alpha_1,\dots,\alpha_n}|\kappa(\alpha_1,\dots,\alpha_n)|\prod_{\ell=1}^k\big|(GBG)_{b_{\ell_{n_\ell}}a_{\ell_1}}\big|\\
&\lesssim N^{-n/2}N^{-k}\left(\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\right)^{k-2}\|B\|^{k-2}\sum_{\alpha_1,\dots,\alpha_n}|\kappa(\alpha_1,\dots,\alpha_n)|\\
&\big|(GBG)_{\tilde b_1\tilde a_1}\big|\,\big|(GBG)_{\tilde b_2\tilde a_2}\big|.
\end{aligned}\tag{5.29}$$

To go to the second line, we performed all the j summations and estimated all the other resolvents without a j index by (5.11); to go to the third line, we used (5.26) for $k-2$ of the k factors and used a simplified notation for the indices $\tilde a, \tilde b$, which agree with some a_{ℓ_i}, b_{ℓ_j}. The two factors of GBG are kept separately, since we aim for an estimate in terms of Hilbert–Schmidt norm $\|B\|_{\mathrm{hs}}$ of the observable B; otherwise the whole argument for the higher order terms would be a simple power counting. However, now we distinguish two cases: (i) $k \le n-2$, and (ii) $k \in \{n-1, n\}$. In the less critical case (i), we use a Schwarz inequality to estimate $\big|(GBG)_{\tilde b\tilde a}\big| \lesssim \sqrt{(GBB^*G^*)_{\tilde b\tilde b}}\sqrt{\rho/\eta_1}$, similarly to (5.26). Then, we continue to estimate (5.29) as

$$\begin{aligned}
&N^{-n/2}N^{-k}\left(\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\right)^{k-1}\|B\|^{k-2}\sum_{\alpha_1,\dots,\alpha_n}|\kappa(\alpha_1,\dots,\alpha_n)|\\
&\quad\sqrt{(GBB^*G^*)_{\tilde b_1\tilde b_1}}\sqrt{(GBB^*G^*)_{\tilde b_2\tilde b_2}}\\
&\quad\le N^{-n/2}N^{-k}\left(\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\right)^{k-1}\|B\|^{k-2}|||\kappa|||_n\sum_{ab}(GBB^*G^*)_{aa}\\
&\quad\lesssim N^{2-n/2}\left(\frac{\rho(E+\mathrm{i}\eta_1)}{N\eta_1}\right)^{k}\|B\|^{k-2}\|B\|_{\mathrm{hs}}^2\lesssim\left(\frac{\rho(E+\mathrm{i}\eta_1)}{N\eta_1}\right)^{k}.
\end{aligned}\tag{5.30}$$

While in the second step, we used (5.20), the final step follows from $\|B\| \le \sqrt{N}\|B\|_{\mathrm{hs}} \lesssim \sqrt{N}$ and $k \le n-2$.

For case (ii), we first note that necessarily $(\tilde a_1, \tilde b_1) = (a_1, b_1) = \alpha_1$, and similarly for index 2, up to permutation of the arguments of κ in (5.29). This simply follows, since $n \ge 4$ derivatives hitting each of $k \in \{n-1, n\}$ factors at least once, means that at least two of them are hit exactly once. Therefore, we can continue estimating (5.29) as

$$\begin{aligned} & N^{-n/2}N^{-k}\left(\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\right)^{k-2}\|B\|^{k-2}\sum_{\alpha_1,\ldots,\alpha_n}|\kappa(\alpha_1,\ldots,\alpha_n)|\left|(GBG)_{b_1a_1}\right|\left|(GBG)_{b_2a_2}\right| \\ & \le N^{-n/2}N^{-k}\left(\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\right)^{k-2}\|B\|^{k-2}|\!|\!|\kappa|\!|\!|_n\sum_{ab}\left|(GBG)_{ba}\right|^2 \\ & \lesssim N^{(k-n)/2}\left(\frac{1}{N\eta_1}\right)^{k}\frac{\rho(E+\mathrm{i}\eta_1)}{\eta_1}\lesssim N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\left(\frac{N^{3\xi}}{N\eta_1}\right)^{k}. \end{aligned} \tag{5.31}$$

Note that for $k=n$ this estimate truly contributes the critical $N^{-2\delta}\rho(E+\mathrm{i}\eta_0)/\eta_0$ factor. Here, in the second step, we used (5.27) together with $\|B\|\le\sqrt{N}\|B\|_{\mathrm{hs}}\lesssim\sqrt{N}$; the final step follows from $\eta_0/\eta_1\le N^{\delta}$ together with monotonicity of $\eta\mapsto\eta\rho(E+\mathrm{i}\eta)$ and $\delta\ll\xi$ by (3.3).

Hence, by Young's inequality, combining (5.30) and (5.31), we deduce

$$\begin{aligned} & \left|N^{-n/2}\sum_{\alpha_1,\ldots,\alpha_n}\kappa(\alpha_1,\ldots,\alpha_n)\mathbb{E}(\partial_{\alpha_1}\ldots\partial_{\alpha_n}|R|^p)\right| \\ & \quad\lesssim\left(1+N^{-2\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{\eta_0}\right)\left[\mathbb{E}|R|^p+\left(\frac{N^{3\xi}}{N\eta_1}\right)^{p}\right]. \end{aligned} \tag{5.32}$$

Therefore, combining (5.18) with (5.22), (5.24), (5.28), and (5.32), we obtain (5.13). This finishes the proof of Proposition 5.5. □

5.3.2. Isotropic case

Proof of Proposition 5.4. Similarly to the proof of Proposition 5.5, after applying Itô's Lemma and a cumulant expansion, we find

$$\left|\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}|S_s|^p\right|\lesssim\left|\sum_{k=2}^{L-1}\sum_{\alpha_1}\sum_{\boldsymbol{\alpha}\in\mathcal{N}(\alpha_1)^k}\frac{\kappa_s(\alpha_1,\boldsymbol{\alpha})}{N^{(k+1)/2}\,k!}\mathbb{E}\big[\partial_{\alpha_1}\partial_{\boldsymbol{\alpha}}|S_s|^p\big]\right|+\Psi(\eta_1)^p. \tag{5.33}$$

for some large enough L.

Employing the same notational simplifications as explained below (5.18), we again first estimate the third order cumulant terms, given by

$$\left|N^{-3/2}\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)\mathbb{E}\big[\partial_{\alpha_1}\partial_{\alpha_2}\partial_{\alpha_3}|S|^p\big]\right|.$$

Distributing the derivatives according to the Leibniz rule, we need to estimate various terms of the forms $(\partial_\alpha^3 S)|S|^{p-1}$, $(\partial_\alpha S)(\partial_\alpha^2 S)|S|^{p-2}$, and $(\partial_\alpha S)^3|S|^{p-3}$. In contrast to the average case treated in the proof of Proposition 5.5, there is no term in the cumulant expansion producing the most critical $N^{-2\delta}\rho/\eta$ factor; instead we get $N^{8\delta}\sqrt{\rho/\eta}=N^{1/2+8\delta}\Psi$.

We start with estimating the first type of terms. In this case, identifying $\alpha_i \equiv (a_i, b_i) \in [N]^2$ and using Lemma 5.3 together with a Ward identity and Assumption 2.3 (i), we find

$$\begin{aligned} &N^{-3/2}\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)G_{\boldsymbol{x}a_1}G_{b_1a_2}G_{b_2a_3}G_{b_3\boldsymbol{y}}\right| \\ &\quad\lesssim N^{-3/2}N^{2\delta}\sum_{\alpha_1,\alpha_2,\alpha_3}|\kappa(\alpha_1,\alpha_2,\alpha_3)|\,|G_{\boldsymbol{x}a_1}|\,|G_{b_3\boldsymbol{y}}| \\ &\quad\lesssim N^{-3/2}N^{2\delta}\,|||\kappa|||_3\left(\sum_{a_1,b_1}|G_{\boldsymbol{x}a_1}|^2\right)^{1/2}\left(\sum_{a_3,b_3}|G_{b_3\boldsymbol{y}}|^2\right)^{1/2}\lesssim N^{1/2+3\delta}\frac{\rho(E+\mathrm{i}\eta_0)}{N\eta_1} \end{aligned} \tag{5.34}$$

with very high probability. Completely analogously we obtain

$$N^{-3/2}\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)G_{\boldsymbol{x}a_1}G_{b_1\boldsymbol{y}}G_{\boldsymbol{x}a_2}G_{b_2a_3}G_{b_3\boldsymbol{y}}\right|\lesssim N^{1/2+4\delta}\left(\frac{\rho(E+\mathrm{i}\eta_0)}{N\eta_1}\right)^{3/2} \tag{5.35}$$

and

$$N^{-3/2}\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)G_{\boldsymbol{x}a_1}G_{b_1\boldsymbol{y}}G_{\boldsymbol{x}a_2}G_{b_2\boldsymbol{y}}G_{\boldsymbol{x}a_3}G_{b_3\boldsymbol{y}}\right|\lesssim N^{1/2+4\delta}\left(\frac{\rho(E+\mathrm{i}\eta_0)}{N\eta_1}\right)^{2}, \tag{5.36}$$

again with very high probability. Hence, combining (5.34), (5.35), and (5.36) with Young's inequality and additionally using that $\eta\mapsto\rho(E+\mathrm{i}\eta)/\eta$ is monotonically decreasing, we infer

$$\begin{aligned} &\left|N^{-3/2}\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa(\alpha_1,\alpha_2,\alpha_3)\mathbb{E}\big[\partial_{\alpha_1}\partial_{\alpha_2}\partial_{\alpha_3}|S|^p\big]\right| \\ &\qquad\lesssim N^{1/2+8\delta}\Psi(\eta_0)\big[\mathbb{E}|S|^p+\Psi(\eta_1)^p\big]\qquad\text{w.v.h.p.} \end{aligned}$$

Next, we turn to the higher order terms, where we aim to estimate

$$\left|N^{-n/2}\sum_{\alpha_1,\ldots,\alpha_n}\kappa(\alpha_1,\ldots,\alpha_n)\mathbb{E}\big[\partial_{\alpha_1}\ldots\partial_{\alpha_n}|S|^p\big]\right|. \tag{5.37}$$

Distributing the n derivatives on $k\in[n]$ factors of S, we find that, for $n_\ell\in\mathbf{N}$ with $\sum_{\ell=1}^k n_\ell=n$ and (w.l.o.g.) $n_1\le n_2\le\ldots\le n_k$, and identifying $(\alpha_i)_{i\in[n]}\equiv\big((a_{\ell_i},b_{\ell_i})\big)_{i\in[n_\ell],\ell\in[k]}$, (5.37) can be rewritten as (ignoring the factor $|S|^{p-k}$)

$$\left|N^{-n/2}\sum_{\alpha_1,\ldots,\alpha_n}\kappa(\alpha_1,\ldots,\alpha_n)\prod_{\ell=1}^k\left(G_{\boldsymbol{x}a_{\ell_1}}G_{b_{\ell_1}a_{\ell_2}}\ldots G_{b_{\ell_{n_\ell-1}}a_{\ell_{n_\ell}}}G_{b_{\ell_{n_\ell}}\boldsymbol{y}}\right)\right|. \tag{5.38}$$

If $n_2=1$, since there are now at least two factors of S hit by a single derivative, we find that (similarly to (5.31) in the proof of Proposition 5.5, cf. also [24, Eqs. (8.82)–(8.85)])

$$\begin{aligned} (5.38) &\lesssim N^{-n/2}N^{(n+k-4)\delta}|||\kappa|||_n\sum_{a,b}|G_{\boldsymbol{x}a}G_{b\boldsymbol{y}}|^2 \\ &\lesssim N^{2-n/2}N^{(n+k-2)\delta}\Psi(\eta_1)^4\le\big[N^{1/2+8\delta}\Psi(\eta_0)\big]\Psi(\eta_1)^k, \end{aligned}$$

with very high probability. If $n_2 > 1$, we find, analogously to (5.30) in the proof of Proposition 5.5 (cf. also [24, Eqs. (8.86)–(8.87)])

$$\begin{aligned}(5.38) &\lesssim N^{-n/2}N^{(n+k-2)\delta}|||\kappa|||_n \sum_{a,b}|G_{\boldsymbol{x}a}|^2 \lesssim N^{2-n/2}N^{(n+k)\delta}\Psi(\eta_1)^2\\ &\lesssim N^{1/2}N^{(n+k)\delta-(n-4)\xi/2}\Psi(\eta_0)\Psi(\eta_1)^{n-2} \lesssim \big[N^{1/2+8\delta}\Psi(\eta_0)\big]\Psi(\eta_1)^k,\end{aligned}$$

with very high probability. Here, to go to the second line, we used that $N^{-1/2+\xi/2} \leq \Psi(\eta_0) \leq \Psi(\eta_1)$. In the ultimate step, we used $\Psi(\eta_1) \leq 1$ and that, since $n_2 > 1$ and $n_1 \leq n_2 \leq ... \leq n_k$, we have $n \geq k+2$. Therefore, using Young's inequality, we infer

$$\left|N^{-n/2}\sum_{\alpha_1,...,\alpha_n}\kappa(\alpha_1,...,\alpha_n)\mathbb{E}\big[\partial_{\alpha_1}...\partial_{\alpha_n}|S|^p\big]\right| \lesssim N^{1/2+8\delta}\Psi(\eta_0)\big[\mathbb{E}|S|^p + \Psi(\eta_1)^p\big], \tag{5.39}$$

with very high probability, and thus, combining (5.34), (5.35), and (5.36) with (5.39), and including the $\Psi(\eta_1)^p$ term from (5.33), we obtain (5.8). This finishes the proof of Proposition 5.4. □

6. Local Law Outside the Support of the scDOS

In this section, we prove Theorem 2.9, that is, the absence of spectrum inside the gaps in the support of ρ_T of size $\Delta_T \geq N^{-3/4+5\varepsilon}$, where $\varepsilon > 0$ is the exponent from (3.2). Recall our choice of the terminal time $T \sim N^{-\xi/4}$ from (3.22).

The characteristic flow was used to exclude outliers near a regular square-root edge for Dyson Brownian motion with general β and potential in [1, Section 4]. In [24, Section 8.1], the approach was used at the edge of non-Hermitian i.i.d. matrices, which corresponds to a cusp-like singularity of the hermitization. We present a modified version of the proof that allows us to avoid moment-matching arguments, used in [24] to remove the order one Gaussian component.

6.1. Time-evolution of the gaps. First, we analyze the dynamics of the gaps in the support the scDOS corresponding to the time-dependent MDE (3.16). For all $t \in [0, T]$, define the density $\rho_t : \mathbb{R} \to \mathbb{R}_+$ via the Stieltjes inversion formula, $\rho_t(x) := \pi^{-1}\lim_{\eta\to+0}\langle \Im M_t(x+\mathrm{i}\eta)\rangle$.

Definition 6.1 (Endpoints of a Gap). For a continuous probability density function ρ on $\mathbb{R}$, we say that $\mathfrak{e}^-, \mathfrak{e}^+$ are left and right *end-points of a gap in the support of* ρ if and only if $\mathfrak{e}^-, \mathfrak{e}^+ \in \partial\{x \in \mathbb{R} : \rho(x) > 0\}$ and $\rho(x) = 0$ for all $x \in [\mathfrak{e}^-, \mathfrak{e}^+]$.

Once Theorem 3.2 is established, the proof of Theorems 2.8 and 2.9 reduces to considering gaps in the support of ρ_T with at least one end point satisfying $\mathrm{dist}(\mathfrak{e}_T, \mathcal{I}) \leq c_M/4$, where $\mathfrak{e}_T \in \{\mathfrak{e}_T^-, \mathfrak{e}_T^+\}$, $\mathcal{I}$ is the set of admissible energies defined in (2.10), and $c_M > 0$ is the constant from Assumption 2.5. We then distinguish between two relevant cases:

(i) The final gap size $\Delta_T := \mathfrak{e}_T^+ - \mathfrak{e}_T^- \leq c_M/4$,
(ii) $\Delta_T > c_M/4$.

We focus on the more challenging case (i), which, in particular, includes all cusp-like singularities in the set of admissible energies. In this case, by Lemma 4.1, the solution $M_t(z)$ remains bounded in and around the gap for all times $0 \leq t \leq T$.

In the simpler case (ii), it is straightforward to verify that the singularity at the endpoint $\mathfrak{e}_t := \varphi_{t,T}(\mathfrak{e}_T)$ is a regular edge-point for all $0 \leq t \leq T$, where $\varphi_{t,T}$ is the flow map defined in (3.19). Consequently, there is no need to track the precise behavior of the opposite endpoint of the gap, and the analysis in Section 6 holds with Δ_t replaced by 1. The definition of the sub-scale domain $\mathcal{D}_t^{\mathrm{sub}}$ (see (6.8) below) must be adjusted by the condition $\varkappa_t(z) := \mathrm{dist}(\mathfrak{e}_t, z) \leq c_M/8 + C'(T-t)$, where $C' \sim 1$ is an appropriate constant (e.g., from Lemma 3.5). The rest of the proof then follows verbatim. Therefore, for the remainder of this section, we assume that $\Delta_T \leq c_M/4$.

For any $t \in [0, T]$ and any $z := E + \mathrm{i}\eta$ with E lying inside the gap $[\mathfrak{e}_t^-, \mathfrak{e}_t^+]$ in the support of ρ_t, the scDOS $\rho_t(z)$ satisfies (see Remark 7.3 in [10])

$$\rho_t(z) \sim \frac{\eta}{(\varkappa_t(z) + \eta)^{1/2}(\Delta_t + \varkappa_t(z) + \eta)^{1/6}}, \quad \varkappa_t(z) := \mathrm{dist}(E, \mathfrak{e}_t^{\pm}). \tag{6.1}$$

In the following lemma, we collect the necessary properties of the quantities $\mathfrak{e}_t^{\pm}$, Δ_t, $\varkappa_t(z_t)$ along the flow (3.18), that we later use in the proof of Proposition 6.6. Recall that the terminal time is small, $T \sim N^{-\xi/4} \ll 1$ by (3.22), and the final gap is also sufficiently small $\Delta_T \leq c_M/4$.

Lemma 6.2 *(Characteristic Flow near Small Gaps). For any time $0 \leq t \leq T$, let $\mathfrak{e}_t^-$, $\mathfrak{e}_t^+$ be the left and right end-points of a gap in the support of ρ_t with size $0 < \Delta_t \lesssim 1$, then for any $0 \leq s \leq t$, there exist a gap in the support of ρ_s with endpoints $\mathfrak{e}_s^-$, $\mathfrak{e}_s^+$ and width $\Delta_s := \mathfrak{e}_s^+ - \mathfrak{e}_s^-$, that satisfy*

$$\Delta_s \sim \Delta_t + (t-s)^{3/2}, \tag{6.2}$$

$$\mathrm{d}\mathfrak{e}_s^{\pm} = -\frac{1}{2}\mathfrak{e}_s^{\pm}\mathrm{d}s - \langle M_s(\mathfrak{e}_s^{\pm})\rangle \mathrm{d}s. \tag{6.3}$$

Pick an $E_t \in (\mathfrak{e}_t^-, \mathfrak{e}_t^+)$ and $\eta_t \lesssim N^{-\nu}\Delta_t$ for some $\nu > 0$. Let $z_s = E_s + \mathrm{i}\eta_s := \varphi_{s,t}(E_t + \mathrm{i}\eta_t)$, as defined in (3.19), then

$$\eta_s \lesssim N^{-\nu/2}\Delta_s, \quad E_s \in (\mathfrak{e}_s^-, \mathfrak{e}_s^+), \quad 0 \leq s \leq t. \tag{6.4}$$

Moreover, for any $0 \leq s \leq t$, recall $\varkappa_s(z) := \mathrm{dist}(\Re z, \mathfrak{e}_s^{\pm})$, and assume that $\varkappa_t(z_t) \gtrsim N^{\nu}\eta_t$, then

$$\eta_s^{-1}\varkappa_s(z_s) \gtrsim \eta_t^{-1}\varkappa_t(z_t), \quad 0 \leq s \leq t. \tag{6.5}$$

Finally, there exists a constant $\mathfrak{c} > 0$, such that for any $0 \leq t \leq T$, if $E_t \in (\mathfrak{e}_t^-, \mathfrak{e}_t^+)$ and $\eta_t \lesssim N^{-\nu}\varkappa_t$, then $z_s := \varphi_{s,t}(E_t + \mathrm{i}\eta_t)$ satisfies

$$\sqrt{\varkappa_s(z_s)} \geq \sqrt{\varkappa_t(z_t)} + \mathfrak{c}(t-s)\Delta_s^{-1/6}. \tag{6.6}$$

We defer the proof of Lemma 6.2 to Appendix A.

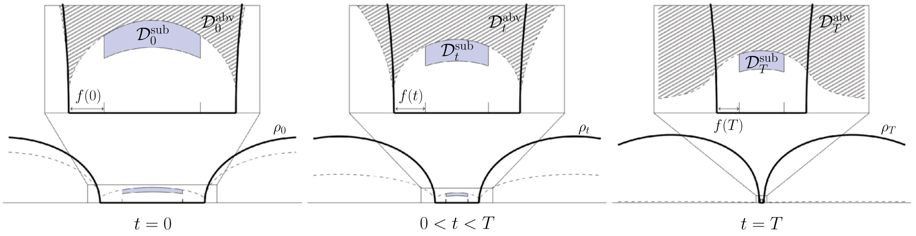
$\mathcal{D}_0^{\mathrm{sub}}$
$\mathcal{D}_0^{\mathrm{abv}}$
$f(0)$
ρ_0
$t=0$
$\mathcal{D}_t^{\mathrm{sub}}$
$\mathcal{D}_t^{\mathrm{abv}}$
$f(t)$
ρ_t
$0<t<T$
$\mathcal{D}_T^{\mathrm{sub}}$
$\mathcal{D}_T^{\mathrm{abv}}$
$f(T)$
ρ_T
$t=T$

Fig. 4. Shaded in blue is the illustration of the time-dependent domain $\mathcal{D}_t^{\mathrm{sub}}$, defined in (6.8), at three distinct times: the initial time $t = 0$ (left), an intermediate time $0 < t < T$ (center), and the terminal time $t = T$ (right). The domain $\mathcal{D}_t^{\mathrm{abv}}$ at the corresponding time t is indicated with crosshatching in the zoomed-in insert, with its boundary indicated by a dashed line in the main plot. The zoomed-in insert also depicts the distance $f(t)$, defined in (6.7), between the edge of the support of ρ_t and the corresponding horizontal cut-off of the domain $\mathcal{D}_t^{\mathrm{sub}}$. The graph of the scDOS ρ_t is superimposed in black on each panel (not to scale)

6.2. Absence of spectrum inside small gaps. Proof of Theorem 2.9. In the sequel, we always assume that the final gap satisfies $\Delta_T \geq N^{-3/4+5\varepsilon}$. Recall the constant ε from (3.2), and define the function $f \equiv f_\varepsilon$ by

$$f(t) \equiv f_\varepsilon(t) := \left[\frac{N^{-1+\varepsilon} + \mathfrak{r}(T-t)}{2\Delta_t^{1/6}} \vee N^{\varepsilon}\sqrt{\eta_{\mathfrak{f},t}}\right]^2, \quad \eta_{\mathfrak{f},t} := N^{-2/3}\Delta_t^{1/9}, \quad t \in [0,T], \tag{6.7}$$

where we chose the constant $\mathfrak{r}$ satisfying $1 \lesssim \mathfrak{r} \leq \mathfrak{c}$ (where $\mathfrak{c}$ is the constant from (6.6)) to be sufficiently small such that $f(t) \leq \frac{1}{4}\Delta_t$. This is indeed possible, since it follows from (6.2) that $\Delta_t^{2/3} \gtrsim \Delta_T^{2/3} + (T-t)$, and $\Delta_T^{2/3} \gg N^{-1/2}$ by assumption on the final gap size.

Fix a tolerance exponent $0 < \zeta < \frac{1}{100}\xi$, where ξ is the exponent from (3.22), and define the time-dependent sub-scale domain $\mathcal{D}_t^{\mathrm{sub}}$ by (see Fig. 4)

$$\mathcal{D}_t^{\mathrm{sub}} \equiv \mathcal{D}_t^{\mathrm{sub}}(\varepsilon,\zeta) := \left\{z := E + \mathrm{i}\eta \in \mathbb{H} : \varkappa_t(z) \geq f(t),\ N^{-\zeta/2} \leq \rho_t(z)N\eta \leq N^{\varepsilon}\right\}, \tag{6.8}$$

where we recall $\varkappa_t(z) = \mathrm{dist}(\Re z, \mathfrak{e}_t^{\pm})$. In the sequel, we omit the arguments ε, ζ of the domain $\mathcal{D}_t^{\mathrm{sub}}$ from the notation.

Definition 6.3 (Exclusion Estimate). Let H_u be a random matrix depending on some parameter[12] $u \in \mathcal{U}$, and let M_u be the solution to the MDE (2.3) with the data pair $(\mathbb{E}H_u, \mathcal{S}_u)$, where $\mathcal{S}_u$ is the self-energy operator corresponding to H_u via (2.2). For all $u \in \mathcal{U}$, let $\mathcal{D}_u$ be a subset of $\mathbb{C}$, and let $\zeta > 0$. We say that the resolvent $G_u(z) := (H_u - z)^{-1}$ satisfies the *exclusion estimate*, with data $(\mathcal{D}_u, \zeta, \Omega)$ uniformly in $u \in \mathcal{U}$, if and only if the bound

$$\left|\left\langle G_u(z) - M_u(z)\right\rangle\right| \leq \frac{N^{-\zeta}}{N|\Im z|}, \tag{6.9}$$

holds uniformly in $z \in \mathcal{D}_u$ and in $u \in \mathcal{U}$, on the event Ω.

The goal of the present subsection is to deduce the following claim.

[12] As in Definition 3.1, the parameter u will typically be time and the set $\mathcal{U}$ will be a bounded subinterval of $\mathbb{R}$.

Claim 6.4. *If a random matrix H satisfies the assumptions of Theorem 2.8, then for any $0 < \zeta < \frac{1}{100}\xi$, the resolvent $G(z) := (H - z)^{-1}$ satisfies the exclusion estimate (6.9) with data $(\mathcal{D}_T^{\mathrm{sub}}, 2\zeta, \Omega)$ for some very-high-probability event Ω.*

Then, using Claim 6.4 as an input, we conclude (2.15) using the following lemma.

Lemma 6.5 *(Eigenvalue Exclusion). Fix a time $t \in [0, T]$, with the terminal time T as in (3.22), and let H be a random matrix satisfying $\mathbb{E}H = A_t$ and $\mathcal{S}_H = \mathcal{S}_t$, where $\mathcal{S}_H$ is the self-energy corresponding to H via (2.2). Assume that for some tolerance exponent $\nu > 0$ and $\ell \in \mathbb{N}$ with $\ell\nu \ll \zeta$, the resolvent $G(z) := (H - z)^{-1}$ satisfies the exclusion estimate (6.9) with data $(\mathcal{D}_t^{\mathrm{sub}}, \zeta - \ell\nu, \Omega)$, then*

$$\operatorname{spec}(H) \cap [\mathfrak{e}_t^- + f(t), \mathfrak{e}_t^+ - f(t)] = \emptyset \quad \text{on } \Omega. \tag{6.10}$$

We defer the proof of Lemma 6.5 to Appendix A.

Proof of Theorem 2.9. Choose $\varepsilon := \frac{1}{5}\theta_0$, $\xi := \frac{1}{10}\varepsilon$ and $\zeta < \frac{1}{100}\xi$. It follows from Claim 6.4 that $G(z) := (H - z)^{-1}$ satisfies the exclusion estimate (6.9) with data $(\mathcal{D}_T^{\mathrm{sub}}, 2\zeta, \Omega)$ for some very-high-probability event Ω, where $\mathcal{D}_T^{\mathrm{sub}} := \mathcal{D}_T^{\mathrm{sub}}(\varepsilon, \xi)$ is defined in (6.8). Hence, (2.15) follows immediately from (6.10) of Lemma 6.5, since $f(T) := f_\varepsilon(T) \geq N^{2\varepsilon}\eta_{\mathfrak{f}}(e_0)$ by definition (6.7). This concludes the proof of Theorem 2.9. □

To prove Claim 6.4, we augment the Zigzag induction of Section 3 with the following propositions. Recall that the relations (3.3) between the fixed tolerance exponents ζ, ξ, ε from (3.2), (3.22) and (6.8), respectively.

Proposition 6.6 *(Zig Step below the Scale). Fix $k \in \{1, \ldots, K\}$, and recall the definition of t_k from (3.23). Let $G_t(z)$ be the time-dependent resolvent defined in (3.27). Assume that for some $\nu > 0$ and $\ell \in \mathbb{N}$ with $\ell\nu \ll \zeta$, the resolvent G_t satisfies the exclusion estimate (6.9) with data $(\mathcal{D}_t^{\mathrm{sub}}, \zeta - \ell\nu, \Omega)$ at time $t = t_{k-1}$, for some very-high-probability event Ω. Then the resolvent G_t satisfies the exclusion estimate (6.9) with data $(\mathcal{D}_t^{\mathrm{sub}}, \zeta - (\ell + 1)\nu, \Omega')$ uniformly in $t \in [t_{k-1}, t_k]$, for some very-high-probability event $\Omega' \subset \Omega$.*

Proposition 6.7 *(Zag Step below the Scale). Fix $k \in \{1, \ldots, K\}$, and let $G^s(z)$ be the time-dependent resolvent defined in (3.28), and let $s_k := s(\vartriangle t_k)$ be as defined in (3.14). Assume that for some $\nu > 0$ and $\ell \in \mathbb{N}$ with $\ell\nu \ll \zeta$, the resolvent $G^s(z)$ satisfies the exclusion estimate (6.9) with data $(\mathcal{D}_{t_k}^{\mathrm{sub}}, \zeta - \ell\nu, \Omega)$ at time $s = s_k$, for some very-high-probability event Ω, and the isotropic local law in (3.1) with data $(\mathcal{D}_{t_k}^{\mathrm{abv}}, \xi + \ell\nu)$ uniformly in time $s \in [0, s_k]$. Then the bound $G^s(z)$ satisfies the exclusion estimate (6.9) with data $(\mathcal{D}_{t_k}^{\mathrm{sub}}, \zeta - (\ell + 1)\nu, \Omega')$ uniformly in time $s \in [0, s_k]$, for some very-high-probability event $\Omega' \subset \Omega$.*

Proof of Claim 6.4. Claim 6.4 follows by induction in k as in Sect. 3 using the tandem of Propositions 6.6 and 6.7, and using the global law of Proposition 3.3 for H_0 as the initial estimate at step $k = 0$. This is indeed sufficient, since for all $z := E + \mathrm{i}\eta \in \mathcal{D}_0^{\mathrm{sub}}$, $N\eta \gtrsim N^{1/4+\varepsilon-\zeta/4}$. Indeed, (6.1), (6.2), (6.7) and (6.8), together with the assumption $\Delta_T \geq N^{-3/4+5\varepsilon}$, imply that

$$N\eta \sim \sqrt{N\,\rho(z) N\eta\, \varkappa(z)^{1/2} \Delta_t^{1/6}} \gtrsim \sqrt{N^{1-\zeta/2}\, N^{-1/3+\varepsilon}\, \Delta_t^{2/9}} \gtrsim N^{1/4+\varepsilon-\zeta/4}. \tag{6.11}$$

Hence, the right-hand side of (3.5b) satisfies

$$N^{3\xi}\Psi(z)\sqrt{\frac{\langle z\rangle}{N\eta}} \lesssim \frac{N^{3\xi}}{N\eta}\sqrt{\frac{1+\rho_0(z)N\eta}{N\eta}} \lesssim \frac{N^{-1/8+3\xi+\zeta/8}}{N\eta} \le \frac{N^{-\zeta}}{N\eta}. \tag{6.12}$$

This concludes the proof of Claim 6.4. □

Proof of Proposition 6.6. The proof is essentially analogous to that in Sect. 4, hence we only outline the key differences.

It follows from (3.18) that $z_s := \varphi_{s,t}(z_t) \in \mathcal{D}_s^{\mathrm{sub}}$ for all $z_t \in \mathcal{D}_t^{\mathrm{sub}}$ and all $0 \le s \le t$. Moreover, using (6.1), we conclude that

$$\frac{\Im z}{\varkappa_s(z)} \lesssim \left(\frac{\eta_{\mathfrak{f},s}}{\varkappa_s(z)}\right)^{3/4}\sqrt{\rho_s(z)N\Im z} \lesssim N^{-\varepsilon}, \quad z \in \mathcal{D}_s^{\mathrm{sub}}. \tag{6.13}$$

Therefore, it follows from (6.6) that for all $z_t \in \mathcal{D}_t^{\mathrm{sub}}$, the trajectory $z_s := \varphi_{s,t}(z_t)$ satisfies

$$\begin{aligned} \varkappa_s(z_s) - f(s) &= \left(\sqrt{\varkappa_s(z_s)} + \sqrt{f(s)}\right)\left(\sqrt{\varkappa_s(z_s)} - \sqrt{f(s)}\right) \\ &\gtrsim \sqrt{\varkappa_s(z_s)}\,\frac{t-s}{\Delta_s^{1/6}}, \quad 0 \le s \le t, \end{aligned} \tag{6.14}$$

where, in the second step, we used (6.6) and (6.7) to estimate $\sqrt{\varkappa_s(z_s)} - \sqrt{f(s)}$.

Let $t_{\mathrm{init}} := t_{k-1}$ and $t_{\mathrm{final}} := t_k$. Define the stopping time τ by

$$\tau := \inf\left\{ t_{\mathrm{init}} < t \le t_{\mathrm{final}} : \sup_{z\in\mathcal{D}_t^{\mathrm{sub}}} \left|N\eta_t\langle G_t(z) - M_t(z)\rangle\right| \ge N^{-\zeta+(\ell+1)\nu} \right\}. \tag{6.15}$$

Statement (6.10) of Lemma 6.5 then implies that on the event $\Omega := \{t \le \tau\}$, the resolvent G_t satisfies the norm bound

$$\|G_t(z)\| \le \frac{\Im z}{\big(\varkappa_t(z) - f(t)\big)^2 + (\Im z)^2}, \quad z \in \mathcal{D}_t^{\mathrm{sub}}. \tag{6.16}$$

Therefore, computing the quadratic variation of the martingale term in (4.2) with $B = \mathbf{1}$ similarly to (4.5) yields

$$\begin{aligned} &\left[\int_{t_{\mathrm{init}}}^{\cdot} \frac{1}{\sqrt{N}} \sum_{ab} \partial_{ab}\big\langle G_s\big\rangle \mathrm{d}\big(\mathfrak{B}_s\big)_{ab}\right]_{t\wedge\tau} \\ &\quad\le \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle (\Im G_s)^2\big\rangle}{N^2\eta_s^2}\mathrm{d}s \le \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{\big\langle \Im G_s\big\rangle}{N^2\eta_s^2}\frac{\eta_s}{\big(\varkappa_s - f(s)\big)^2 + \eta_s^2}\mathrm{d}s \\ &\quad\lesssim \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{1}{N^2\varkappa_s^{3/2}\Delta_s^{-1/6}\big((t\wedge\tau - s)^2 + \varkappa_s^{-1}\Delta_s^{1/3}\eta_s^2\big)}\mathrm{d}s \\ &\quad\lesssim \frac{1}{N^2\varkappa_{t\wedge\tau}^{3/2}\Delta_{t\wedge\tau}^{-1/6}} \int_{t_{\mathrm{init}}}^{t\wedge\tau} \frac{1}{(t\wedge\tau - s)^2 + \varkappa_{t\wedge\tau}^{-1}\Delta_{t\wedge\tau}^{1/3}\eta_{t\wedge\tau}^2}\mathrm{d}s \\ &\quad\lesssim \frac{1}{N^2\varkappa_{t\wedge\tau}^{3/2}\Delta_{t\wedge\tau}^{-1/6}}\frac{1}{\varkappa_{t\wedge\tau}^{-1/2}\Delta_{t\wedge\tau}^{1/6}\eta_{t\wedge\tau}} \lesssim \frac{1}{N^2\eta_{t\wedge\tau}^2}\frac{\eta_{t\wedge\tau}}{\varkappa_{t\wedge\tau}} \lesssim \frac{N^{-\varepsilon}}{N^2\eta_{t\wedge\tau}^2}, \end{aligned} \tag{6.17}$$

abbreviating $G_s := G_s(z_s)$, $\eta_s := \Im z_s$, and $\varkappa_s := \varkappa_s(z_s)$. In (6.17), to go to the third line, we used (6.1), (6.14) and (6.15), while in the last line we used the fact that $\eta_s \geq \eta_t$, $\varkappa_s \gtrsim \varkappa_t$, $\Delta_s \gtrsim \Delta_t$ and $\varkappa_s^{1/2}\Delta_s^{-1/6} \gtrsim \varkappa_t^{1/2}\Delta_t^{-1/6}$ for all $s \leq t$, that follows from (3.18), (6.1), (6.2), (6.5) and (6.6).

The remainder of the proof follows analogously to Sect. 4. □

Proof of Proposition 6.7. Note that by choosing the constant $c' \sim 1$ in (3.21) small enough, we can guarantee that for any $t \in [0, T]$ and any $z := E + \mathrm{i}\eta \in \mathcal{D}_t^{\mathrm{sub}}$, the point $E+\mathrm{i}\eta(E)$ lies in $\mathcal{D}_t^{\mathrm{abv}}$, where $\eta(E)$ is defined implicitly via $\eta(E)\rho_t(E+\mathrm{i}\eta(E)) = N^{-1+\varepsilon}$. Indeed, we only need to check that $\rho_t(E+\mathrm{i}\eta(E))^{-1}\eta(E) \geq c'(N^{-1+\varepsilon}+T-t)$. However, it follows from (6.1) and the definition of $f(t)$ in (6.7) that $\rho_t(E+\mathrm{i}\eta(E))^{-1}\eta(E) \gtrsim N^{\varepsilon}\eta_{\mathfrak{f},t}^{1/2}\Delta_t^{1/6} + T - t$. Together with $\Delta_t \gtrsim \Delta_T \gtrsim N^{-3/4+5\varepsilon}$, this immediately implies that the inclusion $E + \mathrm{i}\eta(E) \in \mathcal{D}^{\mathrm{abv}}$ for sufficiently small $c' \sim 1$.

Since throughout the proof the time t_k remains fixed, for the remainder of this section, we drop the superscript t_k from $\mathcal{D}_{t_k}^{\mathrm{abv}}$, $\mathcal{D}_{t_k}^{\mathrm{sub}}$, ρ_{t_k}, $\varkappa_{t_k}$, Δ_{t_k}, and M_{t_k}.

First, using a monotonicity estimate analogous to Lemma 5.3 (see (A.10) and (A.11) in Remark A.1), we conclude from the isotropic local law in (3.1) for $G^s(z)$ that, uniformly in $z \in \mathcal{D}^{\mathrm{sub}}$, in $a, b \in [N]$ and in $s \in [0, s_k]$,

$$\left|(\Im G^s)_{aa}\right| \lesssim \frac{N^\varepsilon}{N\eta}, \quad \left|(G^s - M)_{ab}\right| \lesssim \frac{N^\varepsilon}{N\eta}, \quad \left|(G^s)_{ab}\right| \lesssim 1, \quad \text{w.v.h.p.} \tag{6.18}$$

Moreover, note that for all $z := E + \mathrm{i}\eta \in \mathcal{D}^{\mathrm{sub}}$, we have the estimates (recall (6.11))

$$\varkappa(z)\Delta^{1/3} \gtrsim N^{-1+4\varepsilon}, \quad N\eta \sim N^{1/2}\varkappa(z)^{1/4}\Delta^{1/12}\sqrt{\rho(z)N\eta} \gtrsim N^{1/4+\varepsilon-\zeta/4}. \tag{6.19}$$

As in Sect. 5, we conduct the proof along the vertical truncations of the domain $\mathcal{D}^{\mathrm{sub}}$, defined as

$$\mathcal{D}_\gamma^{\mathrm{sub}} \equiv \mathcal{D}_{t_k,\gamma}^{\mathrm{sub}} := \left\{z \in \mathcal{D}^{\mathrm{sub}} \equiv \mathcal{D}_{t_k}^{\mathrm{sub}} \,:\, \Im z \geq N^{-1+\gamma}\right\}, \quad 0 < \gamma \leq 1. \tag{6.20}$$

In particular, we assert that if for some constant $\gamma_0 > 0$, the resolvent G^s satisfies the estimate

$$\left\langle \Im G^s(z)\right\rangle \lesssim \rho(z), \tag{6.21}$$

with very high probability uniformly in $z \in \mathcal{D}_{\gamma_0}^{\mathrm{sub}} \cup \mathcal{D}^{\mathrm{abv}}$ and in time $s \in [0, s_k]$, then the estimate (6.9) holds uniformly in $z \in \mathcal{D}_{\gamma_1}^{\mathrm{sub}}$ for any fixed $\gamma_1 \leq \gamma_0 - (\zeta \wedge \frac{1}{2}\mu)$, and uniformly in time $s \in [0, s_k]$ with very high probability.

To this end, we show that the quantity $R_s(z) := \langle G^s(z) - M(z)\rangle$ satisfies

$$\left|\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}|R_s(z)|^p\right| \lesssim \left(1 + \frac{N^{3\zeta}}{\sqrt{\vartriangle t_k}}\right)\left[\mathbb{E}|R_s(z)|^p + \left(\frac{N^{-\zeta}}{N|\Im z|}\right)^p\right], \quad z \in \mathcal{D}_{\gamma_1}^{\mathrm{sub}}, \tag{6.22}$$

where $\vartriangle t_k := t_k - t_{k-1}$ and t_k are defined in (3.23). Note that $N^{3\zeta}\sqrt{\vartriangle t_k} \leq N^{3\zeta}T^{1/2} \lesssim N^{-\zeta}$, using that $T \sim N^{-\xi/4}$ from (3.22).

The proof of (6.22) is analogous to that of Proposition 5.5. The main difference is that for the most critical term (5.21), we use the bound

$$N^{-5/2}\left|\sum_{\alpha_1,\alpha_2,\alpha_3}\kappa_s(\alpha_1,\alpha_2,\alpha_3)M_{b_1a_2}M_{b_2a_3}(G^sG^s)_{b_3a_1}\right|\le N^{-1}\|\!|\kappa\|\!|_3^{\mathrm{av}}\|M\|^2\|G^sG^s\|_{\mathrm{hs}}$$
$$\lesssim\frac{\langle\Im G^s\rangle^{1/2}}{N\eta^{3/2}}\lesssim\frac{N^{-\zeta}}{N\eta}\frac{N^{2\zeta}}{\varkappa^{1/4}\Delta^{1/12}}\lesssim\frac{N^{-\zeta}}{N\eta}\frac{N^{2\zeta}}{\sqrt{T-t_k}}\lesssim\frac{N^{-\zeta}}{N\eta}\frac{N^{\frac{5}{2}\zeta}}{\sqrt{\Delta t_k}},\tag{6.23}$$

where we used (6.21) together with the monotonicity of the map $\eta\mapsto\eta\langle\Im G^s(E+\mathrm{i}\eta)\rangle$ for any fixed $E\in\mathbb{R}$ to assert that $\langle\Im G^s(z)\rangle\lesssim N^{2\zeta}\rho(z)$ with very high probability, uniformly in $z\in\mathcal{D}^{\mathrm{sub}}_{\gamma_1}$.

The remainder of the proof follows analogously to Sect. 5 using the estimates (6.18) instead of the respective bounds in (5.2) and (5.11). □

6.3. Improved Local Laws away from the Spectrum. Proof of Theorem 2.8.

Proof of Theorem 2.8. Let $\varepsilon:=\min\{\frac{1}{5}\varepsilon_0,\frac{1}{2}\xi_0\}$ and $\xi:=\frac{1}{10}\varepsilon$. Let $z\in\mathbb{C}$ be a spectral parameter satisfying $N^{\varepsilon_0}\eta_{\mathfrak{f}}(E)\le\operatorname{dist}(z,\operatorname{supp}\rho)\le C$. Without loss of generality, we assume that $\|\boldsymbol{x}\|=\|\boldsymbol{y}\|=\|B\|_{\mathrm{hs}}=1$, and that $z:=E+\mathrm{i}\eta$ with $\eta\ge0$.

First, consider the case $\operatorname{dist}(z,\operatorname{supp}\rho)\le2\eta$, then it is straightforward to check using the universal shape of the density ρ (see, e.g., Remark 7.3 in [10]) that $\rho(z)N\eta\gtrsim N^{\varepsilon}$. Therefore, in this regime, Theorem 2.8 follows from Theorem 3.2 and Proposition 3.3.

It remains to consider the regime $\operatorname{dist}(z,\operatorname{supp}\rho)\ge2\eta$. Clearly, E lies outside of the support of ρ. Let $\mathfrak{e}^-$ and $\mathfrak{e}^+$ be the left and right end-points of the gap that contains E. The assumption $\operatorname{dist}(z,\operatorname{supp}\rho)\ge2\eta$ implies that $\varkappa:=\operatorname{dist}(E,\mathfrak{e}^{\pm})\gtrsim\eta$, hence $\Delta:=\mathfrak{e}^+-\mathfrak{e}^-\ge\varkappa\gtrsim N^{\varepsilon_0}\eta_{\mathfrak{f}}(E)=N^{-2/3+\varepsilon_0}\Delta^{1/9}$, and thus $\Delta\ge N^{-3/4+9\varepsilon_0/8}$.

Define a local domain $\mathcal{D}^{\mathrm{out}}\equiv\mathcal{D}^{\mathrm{out}}(E)$ as

$$\mathcal{D}^{\mathrm{out}}\equiv\mathcal{D}^{\mathrm{out}}(E):=\{z'\in\mathbb{C}\,:\,|\Re z'-E|\le\tfrac{1}{2}\varkappa,\ |\Im z'|\le\varkappa\},\quad\varkappa:=\operatorname{dist}(E,\mathfrak{e}^{\pm}),\tag{6.24}$$

and observe that $z\in\mathcal{D}^{\mathrm{out}}$. Moreover, by Theorem 2.9 with $\theta_0:=\frac{1}{2}\varepsilon_0$, there exists a very-high-probability event Ω, such that $\operatorname{spec}(H)\cap\mathcal{D}^{\mathrm{out}}=\emptyset$ on Ω.

Therefore, on the very-high-probability event Ω, the matrix-valued map $z'\mapsto G(z')-M(z')$ is analytic in the interior of $\mathcal{D}^{\mathrm{out}}$. Using the Cauchy formula, we obtain the contour integral representation

$$G(z)-M(z)=\frac{1}{2\pi\mathrm{i}}\oint_\Gamma\frac{G(z')-M(z')}{z-z'}\mathrm{d}z',\tag{6.25}$$

where $\Gamma\subset\mathcal{D}^{\mathrm{out}}$ is the contour tracing the boundary of a rectangle centered at z with width $\frac{1}{4}\varkappa$ and height $\frac{3}{4}\varkappa$. Note that $|z'-z|\gtrsim\varkappa$ for all $z'\in\Gamma$. Using a monotonicity estimate analogous to Lemma 5.3 (see (A.11), (A.13) in Remark A.1), we conclude from Proposition 3.3 and Theorem 3.2 that on a very-high-probability event $\Omega'\subset\Omega$, the resolvent $G(z')$ satisfies

$$\left|\big\langle\big(G(z')-M(z')\big)B\big\rangle\right|\lesssim\frac{N^{\varepsilon}}{N|\Im z'|}\wedge\frac{1}{\varkappa},\quad\left|\big(G(z')-M(z')\big)_{\boldsymbol{x}\boldsymbol{y}}\right|$$

$$\lesssim N^{\varepsilon}\sqrt{\frac{\rho(z')}{N|\Im z'|}}+\frac{N^{\varepsilon}}{N|\Im z'|}\wedge\frac{1}{\varkappa},\tag{6.26}$$

uniformly in $z'\in\Gamma$, where the alternative $\varkappa^{-1}$ bound follows from the norm-bound on $\|G(z')\|$ and (2.15).

Plugging the bounds (6.26) into the representation (6.25) and using the comparison relation (6.1), we obtain (2.14b) and (2.14a) at the point z. Here we used (6.1) and $\varkappa\geq N^{\varepsilon_0}\eta_{\mathfrak{f}}(E)$ to assert that

$$\sqrt{\frac{\rho(z)}{N\eta}}\sim\sqrt{\frac{1}{N\varkappa^{1/2}\Delta^{1/6}}}\gtrsim\frac{1}{N\varkappa}.\tag{6.27}$$

Therefore, the local laws (2.14a) and (2.14b) hold for $\operatorname{dist}(z,\operatorname{supp}\rho)\leq C$.

In the complementary regime $\operatorname{dist}(z,\operatorname{supp}\rho)\geq C$, similar contour integration together with the global law (3.5b), can be used to obtain the faraway laws

$$\left|\big\langle\big(G(z)-M(z)\big)B\big\rangle\right|\lesssim\frac{N^{\xi_0}}{N\langle z\rangle^2}\|B\|_{\mathrm{hs}}\,,\quad\left|\big(G(z)-M(z)\big)_{\boldsymbol{x}\boldsymbol{y}}\right|\lesssim\frac{N^{\xi_0}}{\sqrt{N}\langle z\rangle^2}\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|\,,\tag{6.28}$$

in the regime $\operatorname{dist}(z,\operatorname{supp}\rho)\in[C,N^D]$ for some sufficiently large positive $C\sim 1$. Note that for such z, the proof requires only the global laws of Proposition 3.3 as an input, and is conducted without the use of the Zigzag dynamics. This concludes the proof of Theorem 2.8. □

7. Global Laws: Proof of Proposition 3.3

We prove Proposition 3.3 in two steps. First, in Sect. 7.2, we prove the isotropic local law (3.5a). Then, in Sect. 7.3, we conclude the proof of Proposition 3.3 by proving the averaged law (3.5b), using the isotropic law (3.5a) as an input. Before proceeding with the proof, we collect some preliminary bounds on the stability operator and define the appropriate norm for proving the isotropic local law.

7.1. Preliminaries for the global law. First, for any $z\in\mathbb{C}$, the *stability operator* $\mathcal{B}(z):\mathbb{C}^{N\times N}\to\mathbb{C}^{N\times N}$ is defined by its action on $X\in\mathbb{C}^{N\times N}$,

$$\mathcal{B}(z)[X]:=X-M(z)\mathcal{S}[X]M(z).\tag{7.1}$$

We control the inverse of the stability operator $\mathcal{B}$ using the following lemma.

Lemma 7.1 *(Proposition 4.4 in [11]). Let $M(z)$ be the solution to the MDE (2.3), and let $\mathcal{I}$ be the set of admissible energies defined in (2.10). Then the stability operator $\mathcal{B}(z)$, defined in (7.1) satisfies, for all $z\in\mathbb{C}$ with* $\operatorname{dist}(\Re z,\mathcal{I})\leq\frac{3}{4}c_M$,

$$\left\|\mathcal{B}^{-1}(z)\right\|_{\mathrm{hs}\to\mathrm{hs}}+\left\|\mathcal{B}^{-1}(z)\right\|_{\|\cdot\|\to\|\cdot\|}\lesssim 1+\beta(z)^{-1},\quad\beta(z):=\rho(z)^2+\rho(z)|\sigma(z)|+\rho(z)^{-1}|\Im z|,\tag{7.2}$$

where the function[13] $\sigma(z)$ *is defined as*

$$\sigma(z) := \left\langle \operatorname{sign}\big(\Re U(z)\big)\big(\rho(z)^{-1}\Im U(z)\big)^3 \right\rangle,$$
$$U := \frac{(\Im M)^{-1/2}(\Re M)(\Im M)^{-1/2} + \mathrm{i}}{\left|(\Im M)^{-1/2}(\Re M)(\Im M)^{-1/2} + \mathrm{i}\right|}, \quad z \in \mathbb{H}. \tag{7.3}$$

Note that by definition of $\mathcal{D}^{\mathrm{glob}}$ in (3.4), the stability *factor satisfies* $\beta(z) \geq N^{-\xi/4}$ for all $z \in \mathcal{D}^{\mathrm{glob}}$.

Remark 7.2 (Local Laws in the Stable Domain). In Sect. 7 we only use the bound $\beta(z) \geq \rho(z)^{-1}|\Im z|$. However, by Remark 10.4 in [10], there exists a function $\widetilde{\beta}(z)$ satisfying $\beta(z) \lesssim \widetilde{\beta}(z) \leq \beta(z)$, such that the map $\eta \mapsto \widetilde{\beta}(E + \mathrm{i}\eta)$ is non-decreasing in $\eta > 0$ for any fixed E. Therefore, the global domain, defined in (3.4), can be replaced by the *stable domain*, defined as

$$\mathcal{D}^{\mathrm{stab}} := \left\{ z := E + \mathrm{i}\eta \in \mathbb{H} \,:\, |E| \leq N^D,\ N^{-1+\varepsilon} \leq \eta \leq N^D,\ \widetilde{\beta}(z) \geq N^{-\xi/4} \right\}, \tag{7.4}$$

with our proof of Proposition 3.3 naturally extending to the larger *stable domain*. In particular, the stable domain extends down to the level $\eta \geq N^{-1+\varepsilon}$ in the bulk of spectrum, where $\rho(E) \gtrsim 1$. Therefore, we provide an independent proof of the local laws in Theorems 2.1 and 2.2 of [35] under the Assumptions 2.1–2.5 without the complicated graphical expansion machinery.

Next, for a fixed spectral parameter $z \in \mathcal{D}^{\mathrm{glob}}(\xi, D)$, and a fixed pair of vectors $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{C}^N$, define a family of sets of vectors,

$$\begin{aligned} \mathcal{V}_0 &\equiv \mathcal{V}_0(z) := \{\boldsymbol{e}_a\}_{a=1}^N \cup \{\boldsymbol{x}, \boldsymbol{y}\}, \\ \mathcal{V}_j &\equiv \mathcal{V}_j(z) := \mathcal{V}_{j-1} \cup \left\{ M\boldsymbol{u}, \kappa_{\mathrm{c}}\big((M\boldsymbol{u})a, \cdot b\big), \kappa_{\mathrm{d}}\big((M\boldsymbol{u})a, b\cdot\big) : \boldsymbol{u} \in \mathcal{V}_{j-1},\ a, b \in [N] \right\}, \\ &\qquad j \in \{1, \ldots, J\}, \end{aligned} \tag{7.5}$$

where $M := M(z)$, and J is an integer satisfying $J \geq 2/\xi$. We use the corresponding isotropic norm (Section 5.1 in [35])

$$\|X\|_* \equiv \|X\|_*^{\boldsymbol{x},\boldsymbol{y},J,z} := \sum_{j=0}^{J} N^{-\frac{j}{2J}} \|X\|_{(j)} + N^{-1/2} \max_{\boldsymbol{v} \in \mathcal{V}_J} \frac{\|X_{\cdot\boldsymbol{v}}\|}{\|\boldsymbol{v}\|},$$
$$\|X\|_{(j)} := \max_{\boldsymbol{u},\boldsymbol{v} \in \mathcal{V}_j} \frac{|X_{\boldsymbol{u}\boldsymbol{v}}|}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}. \tag{7.6}$$

Note that the cardinality of the sets $\mathcal{V}_j$ is bounded by N^{CJ}, hence we can take the maximum of very-high-probability bounds over these sets.

Finally, recall that for all z with $\Re z$ in the set of admissible energies $\mathcal{I}$ from Assumption 2.5, $M(z)$ satisfies the bound

$$\|M(z)\| \lesssim \langle z\rangle^{-1}. \tag{7.7}$$

[13] Roughly speaking, the quantity $|\sigma(z)|$ measures how close z is to a possible almost cusp, in particular, if x is an exact cusp of the density $\rho(x)$, then $\sigma(x) = 0$.

7.2. Proof of the isotropic bound in Proposition 3.3.

Proof. (Proof of the isotropic law in (3.5a)) Recall the definition of the domain $\mathcal{D}^{\mathrm{glob}}$ from (3.4). We conduct the proof iteratively along vertical truncations $\mathcal{D}^{\mathrm{glob}}_{\gamma}$ of the domain $\mathcal{D}^{\mathrm{glob}}$, defined as

$$\mathcal{D}^{\mathrm{glob}}_{\gamma} := \{z := E + \mathrm{i}\eta \in \mathcal{D}^{\mathrm{glob}} \; : \; \eta \geq N^{-1+\gamma}\}, \quad \gamma > 0. \tag{7.8}$$

Once the local law (3.5a) is established in the domain $\mathcal{D}^{\mathrm{glob}}_{\gamma_0}$ for some $\gamma_0 \geq 0$, a simple monotonicity argument analogous to Lemma 5.3 (see the proof of Lemma 5.3 in Appendix A) implies that the following bounds on the resolvent $G(z)$,

$$\left|G(z)_{\boldsymbol{u}\boldsymbol{v}}\right| \lesssim N^{\delta}\langle z\rangle^{-1}, \quad \left|\left(\Im G(z)\right)_{\boldsymbol{u}\boldsymbol{u}}\right| \lesssim N^{\xi+\delta}\left(\rho(z) + \frac{1}{N\eta}\right), \quad \text{w.v.h.p.,} \tag{7.9}$$

hold uniformly in $z \in \mathcal{D}^{\mathrm{glob}}_{\gamma_1}$ for any $\gamma_1 \geq \gamma_0 - \delta$ with $\delta \leq \frac{1}{20}\xi$, and for any deterministic $\boldsymbol{u}, \boldsymbol{v}$ with $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$. Therefore, the key step in the iteration is going from estimates on the resolvent $G(z)$ to a bound on $(G(z) - M(z))_{\boldsymbol{x}\boldsymbol{y}}$, that is, using the bounds (7.9) as an input to prove the isotropic local law (3.5a) in the domain $\mathcal{D}^{\mathrm{glob}}_{\gamma_1}$.

This crucial step is based on the following gap in the possible values of $\|G - M\|_*$.

Lemma 7.3 *(Gap in the Values of $G - M$). Fix a spectral parameter $z \in \mathcal{D}^{\mathrm{glob}}_{\gamma_1}$, with some $\gamma_1 > 0$ such that (7.9) holds on $\mathcal{D}^{\mathrm{glob}}_{\gamma_1}$, then*

$$\|G(z) - M(z)\|_* \lesssim N^{-\xi} \text{ w.v.h.p.} \quad \Longrightarrow \quad \|G(z) - M(z)\|_* \lesssim N^{\xi}\Psi(z) \text{ w.v.h.p.} \tag{7.10}$$

We initialize the iteration in the domain $\mathcal{D}^{\mathrm{glob}}_{2+\delta}$. Indeed, owing to the very high probability bound $|H_{\boldsymbol{u}\boldsymbol{v}}| \lesssim N^{1/2+\nu}$ for any $\nu > 0$, we have, for any deterministic $\boldsymbol{u}, \boldsymbol{v}$ with $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$,

$$\|G(z)\| \lesssim \langle z\rangle^{-1}, \quad \left|\left(\Im G(z)\right)_{\boldsymbol{u}\boldsymbol{u}}\right| \lesssim \frac{\eta}{\langle z\rangle^2} \sim \rho(z), \quad z \in \mathcal{D}^{\mathrm{glob}}_{2+\delta}, \quad \text{w.v.h.p.} \tag{7.11}$$

Note that the bound $\|G(z) - M(z)\|_* \leq N^{-\xi}$ holds trivially for all z with $\Im z \geq N^{\xi}$. After Lemma 7.3 is established, the proof of (3.5a) follows the standard continuity argument on a fine grid (see Section 5.4 in [35]).

This concludes the proof of the isotropic law in (3.5a). □

The remainder of this subsection is devoted to the proof of Lemma 7.3. A local law for random matrices with slow correlation decay away from the cusps was already proved in [35] and [11]. We present an independent proof under the Assumptions 2.1–2.5. We utilize the *minimalistic cumulant expansion*, that was used previously in [52] and [26]. This allows us to avoid the complicated graphical expansions.

Proof of Lemma 7.3. Since $z := E + \mathrm{i}\eta$ is fixed, we omit the argument of G, M, Ψ, ρ, β, and $\mathcal{B}$. Assume the very-high-probability bound

$$\|G - M\|_* \lesssim N^{-\xi}. \tag{7.12}$$

It suffices to show that $\|G - M\|_* \leq N^{\xi}\Psi$ with very high probability. Assume that for a deterministic control parameter ψ, the quantity $\Psi^{-1}\|G - M\|_*$ satisfies

$$\Psi^{-1}\|G - M\|_* \lesssim \psi, \quad \text{w.v.h.p.} \tag{7.13}$$

By definition of the resolvent $G := (H - z)^{-1}$ and the MDE (2.3), we difference $G - M$ satisfies

$$G - M = -M\underline{WG} + M\mathcal{S}[G - M]G, \tag{7.14}$$

where the matrix[14] $\underline{WG}$ is defined as

$$\underline{WG} := WG + \mathcal{S}[G]G. \tag{7.15}$$

Therefore, subtracting $M\mathcal{S}[G - M]M$ from both sides and the inverse of the stability operator $\mathcal{B}$, defined in (7.1), yields the equation

$$G - M = -\mathcal{B}^{-1}\big[M\underline{WG}\big] + \mathcal{B}^{-1}\big[M\mathcal{S}[G - M](G - M)\big], \tag{7.16}$$

Observe, that for any $X \in \mathbb{C}^{N\times N}$, (Eq. (5.4c) in [35])

$$\begin{aligned}\left\|\mathcal{B}^{-1}[X]\right\|_{(j)} &\le \|X\|_{(j)} + \left(\|M\|^2 \,|||\mathcal{S}||| + \|M\|^4 \,|||\mathcal{S}|||^2 \left\|\mathcal{B}^{-1}\right\|_{\text{hs}\to\text{hs}}\right) \|X\|_{\max} \\ &\lesssim \|X\|_{(j)} + \left(1 + \beta^{-1}\right) \|X\|_{(0)}\,,\end{aligned} \tag{7.17}$$

where in the last step we used (7.2). Here we denote

$$|||\mathcal{S}||| := \|\mathcal{S}\|_{\max\to\|\cdot\|} \vee \|\mathcal{S}\|_{\text{hs}\to\|\cdot\|}\,. \tag{7.18}$$

To control the norm $\|G - M\|_*$, we first bound the $\|\cdot\|_{(j)}$ individually, and then estimate the contribution coming from the last summand in (7.6) later. Fix an index $j \in \{0, \ldots, J\}$ and fix a pair of vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}_j$. We compute the p-th (for even p) moment of

$$S_j \equiv S_j^{\boldsymbol{uv}} := N^{\frac{-j}{2J}} (G - M)_{\boldsymbol{uv}}, \tag{7.19}$$

using the equation (7.16) for a single factor,

$$\begin{aligned}\mathbb{E}\big[|S_j|^p\big] \le \mathbb{E}\big[&N^{\frac{-j}{2J}} \left(\mathcal{B}^{-1}[M\underline{WG}]\right)_{\boldsymbol{uv}} \overline{S_j} |S_j|^{p-2}\,\big] \\ +\mathbb{E}\big[&N^{\frac{-j}{2J}} \left(\mathcal{B}^{-1}[M\mathcal{S}[G - M](G - M)]\right)_{\boldsymbol{uv}} \overline{S_j} |S_j|^{p-2}\,\big].\end{aligned} \tag{7.20}$$

First, we estimate the size of the second term on the right-hand side of (7.20). We observe that (Eq. (5.5a), (5.5b) in [35])

$$\|M\mathcal{S}[X]X\|_{(j)} \lesssim |||\kappa|||_2^{\text{iso}} \,\|M\| \min\left\{\, \|X\|_{(j+1)}\,, \sqrt{N}\, \|X\|_{(0)} \right\} \|X\|_*\,. \tag{7.21}$$

We only use the second mode of the min bound (i.e. use $\min\{A, B\} \le B$) when $j = J$. Combining (7.12), (7.17) and (7.21) (in particular, leading to the subscript (1) instead of (0) at the $(1 + \beta^{-1})\|G - M\|_{(1)}$-term), we deduce that

$$Q_j := N^{\frac{-j}{2J}} \left(\mathcal{B}^{-1}[M\mathcal{S}[G - M](G - M)]\right)_{\boldsymbol{uv}}$$

[14] The underline $\underline{WG}$ is a renormalization of WG; for renormalization of general products $f(W)Wg(W)$, see Section 4 in [25].

satisfies

$$
\begin{aligned}
\|Q_j\|_{(j)} &\lesssim \frac{\|G-M\|_*}{\langle z\rangle N^{\frac{j}{2J}}} \\
&\quad \Big(\|G-M\|_{(j+1)}\,\mathbf{1}_{j<J} + \sqrt{N}\,\|G-M\|_{(0)}\,\mathbf{1}_{j=J} + \big(1+\beta^{-1}\big)\,\|G-M\|_{(1)} \Big) \\
&\lesssim N^{\frac{1}{2J}-\xi}\langle z\rangle^{-1}\big(1+\beta^{-1}\big)\psi\Psi, \quad \text{w.v.h.p.,}
\end{aligned}
\tag{7.22}
$$

where in the last step we used the estimate (7.2), the definition of $\mathcal{D}^{\mathrm{glob}}$ in (3.4), assumptions (7.12–7.13), and the bound $\|X\|_{(j)} \le N^{\frac{j}{2J}}\,\|X\|_*$ that follows from the definition of $\|\cdot\|_*$ in (7.6).

Next, we estimate the first term in (7.20). For any $j \in \{0,\dots,J\}$ and any $\boldsymbol{u},\boldsymbol{v} \in \mathcal{V}_j$, using the multivariate cumulant expansion formula from Proposition 5.2, we obtain

$$
\begin{aligned}
&\Big|\mathbb{E}\big[(M\underline{WG})_{\boldsymbol{uv}}\overline{S_j}|S_j|^{p-2}\big]\Big| \\
&\le \bigg|\mathbb{E}\bigg[\frac{1}{N}\sum_{ab}\sum_{\alpha_1} M_{\boldsymbol{u}a}G_{b\boldsymbol{v}}\kappa(ab,\alpha_1)\partial_{\alpha_1}\big\{\overline{S_j}|S_j|^{p-2}\big\}\bigg]\bigg| \\
&\quad + \sum_{k=2}^{L-1}\bigg|\mathbb{E}\bigg[\sum_{ab}\sum_{\boldsymbol{\alpha}\in\mathcal{N}(ab)^k} M_{\boldsymbol{u}a}\frac{\kappa(ab,\boldsymbol{\alpha})}{N^{(k+1)/2}k!}\partial_{\boldsymbol{\alpha}}\big\{G_{b\boldsymbol{v}}\overline{S_j}|S_j|^{p-2}\big\}\bigg]\bigg| \\
&\quad + N^{\frac{j}{2J}}\big|\Omega_{j,L}^{\boldsymbol{uv}}\big|.
\end{aligned}
\tag{7.23}
$$

Similarly to (5.17), we can choose L large enough such that $|\Omega_{j,L}| \lesssim \big(\Psi\,\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\big)^p$. We note that the $N^{\frac{-j}{2J}}$ factors in (7.19) are only relevant for the quadratic term Q_j estimated above, therefore, we do not follow it in the sequel. Moreover, we drop the norms $\|\boldsymbol{u}\|$ and $\|\boldsymbol{v}\|$ for brevity.

First, we estimate the term involving second-order cumulants on the right-hand side of (7.23). Here we estimate the contribution coming from the cross part of the second cumulants κ_{c}, the estimate for the direct part κ_{d} is completely analogous. Ignoring the difference between S_j and $\overline{S_j}$, and dropping the overall $|S_j|^{p-2}$ factor, we obtain the bound

$$
\begin{aligned}
\bigg|\frac{1}{N}\sum_{ab}\sum_{\alpha_1}\kappa_{\mathrm{c}}(ab,\alpha_1)M_{\boldsymbol{u}a}G_{b\boldsymbol{v}}\partial_{\alpha_1}S_j\bigg| &\lesssim \frac{1}{N}\sum_{bb_1}\bigg|\sum_{a_1}\kappa_{\mathrm{c}}\big((M\boldsymbol{u})\,b, a_1b_1\big)G_{\boldsymbol{u}a_1}\bigg|\big|G_{b\boldsymbol{v}}G_{b_1\boldsymbol{v}}\big| \\
&\lesssim N^{-1+\delta}\langle z\rangle^{-1}\big\|\,\|\kappa_{\mathrm{c}}((M\boldsymbol{u})*,\cdot*)\|\,\big\|\,\|G_{\cdot\boldsymbol{v}}\|^2 \\
&\lesssim |||\kappa|||_2^{\mathrm{iso}}N^{\xi+2\delta}\Psi^2, \quad \text{w.v.h.p.}
\end{aligned}
\tag{7.24}
$$

In the ultimate step, we used (7.7) and (7.9) to assert that, with very high probability,

$$
\frac{1}{\langle z\rangle\sqrt{N}}\,\|G_{\cdot\boldsymbol{v}}\| = \sqrt{\frac{(\Im G)_{\boldsymbol{vv}}}{\langle z\rangle^2 N\eta}} \lesssim N^{\frac{\xi+\delta}{2}}\Psi.
\tag{7.25}
$$

Next, we bound term involving third and higher order cumulants in (7.23). Consider, for example,

$$\begin{aligned}&\Bigg|\sum_{ab}\sum_{\alpha_1,\alpha_2}\frac{\kappa(ab,\alpha_1,\alpha_2)}{N^{3/2}}M_{\boldsymbol{u}a}G_{b\boldsymbol{v}}(\partial_{\alpha_1}S_j)(\partial_{\alpha_2}S_j)\Bigg|\\ &\qquad\lesssim N^{-3/2}\Bigg|\sum_{ab}\sum_{a_1b_1a_2b_2}\kappa(ab,a_1b_1,a_2b_2)M_{\boldsymbol{u}a}G_{b\boldsymbol{v}}G_{\boldsymbol{u}a_1}G_{b_1\boldsymbol{v}}G_{\boldsymbol{u}a_2}G_{b_2\boldsymbol{v}}\Bigg|\\ &\qquad\lesssim N^{3\xi/2+7\delta/2}\Psi^3|||\kappa|||_3,\quad \text{w.v.h.p.}\end{aligned}\tag{7.26}$$

Note that the structure of the term (7.26) is identical to that of (5.36). Indeed, the only difference is that the resolvent $G_{\boldsymbol{x}a}$ is replaced by the deterministic approximation $M_{\boldsymbol{u}a}$ ($\boldsymbol{u}$ and $\boldsymbol{v}$ in (7.26) play the role of $\boldsymbol{x}$ and $\boldsymbol{y}$ in (5.36)). Consequently, the summation over a is bounded using

$$\Big(\sum_a|M_{\boldsymbol{u}a}|^2\Big)^{1/2}\le\|M\|\lesssim\frac{1}{\langle z\rangle}\quad\text{instead of}\quad\Big(\sum_a|G_{\boldsymbol{u}a}|^2\Big)^{1/2}\lesssim N^{\frac{\xi+\delta}{2}}\sqrt{\frac{\rho+\frac{1}{N\eta}}{\eta}},\tag{7.27}$$

yielding a saving of a $\sqrt{\rho/\eta}$ factor in terms of the (ρ/η)-power on the right-hand side of (7.26) compared to the bound in (5.36). All other terms in (7.23) with cumulant of order three and higher are bounded analogously to their counterparts in the proof of Proposition 5.4, with the additional saving of $\sqrt{\rho/\eta}$ coming from (7.27).

Therefore, using a weighted Young inequality to handle the separated $|S_j|^{p-k}$ terms, we deduce that for all $j\in\{0,\dots,J\}$,

$$\mathbb{E}\big[N^{\frac{-j}{2J}}\big(\mathcal{B}^{-1}[M\underline{WG}]\big)_{\boldsymbol{uv}}\overline{S_j}|S_j|^{p-2}\big]\le\big(N^{\xi/2+4\delta}(1+\beta^{-1})\Psi\big)^p+N^{-p\delta}\mathbb{E}\big[|S_j|^p\big].\tag{7.28}$$

It follows from (7.20), (7.22), (7.23), and (7.28) that

$$\mathbb{E}\big[|S_j|^p\big]\lesssim(\Psi)^p\bigg(N^{\xi/2+4\delta}(1+\beta^{-1})+N^{-\delta}\psi+N^{\frac{1}{2J}-\xi+\delta}\langle z\rangle^{-1}\big(1+\beta^{-1}\big)\psi\bigg)^p.\tag{7.29}$$

Since $J\ge 2/\xi$, and $\delta\le\xi/20$, we have $(1+\beta^{-1})N^{\frac{1}{2J}-\xi+\delta}\le N^{-\delta}$, and we conclude that

$$|S_j|\lesssim N^{\nu}\Psi\big(N^{\xi/2+4\delta}(1+\beta^{-1})+N^{-\delta}\psi\big),\quad\text{w.v.h.p.}\tag{7.30}$$

Next, we estimate the contribution of the last summand in (7.6) to $\|G-M\|_*$. We fix a vector $\boldsymbol{v}\in\mathcal{V}_J$ and compute a the p-th (for even p) moment of

$$S\equiv S^{\boldsymbol{v}}:=N^{-1}\,\|(G-M)_{\cdot\boldsymbol{v}}\|^2=N^{-1}\big((G-M)^*(G-M)\big)_{\boldsymbol{vv}}.\tag{7.31}$$

Using the equation (7.14) for a single S factor, we obtain

$$\begin{aligned}\mathbb{E}\big[|S|^p\big]\le N^{-1}\big|\,\mathbb{E}\big[\big((G-M)^*M\underline{WG}\big)_{\boldsymbol{vv}}\overline{S}|S|^{p-2}\big]\,\big|\\ +N^{-1}\big|\,\mathbb{E}\big[\big((G-M)^*M\mathcal{S}[G-M]G\big)_{\boldsymbol{vv}}\overline{S}|S|^{p-2}\big]\,\big|.\end{aligned}\tag{7.32}$$

To estimate the term in the second line of (7.32), we note the following bound,

$$\left|\big(X^* M \mathcal{S}[X] Y\big)_{\boldsymbol{v}\boldsymbol{v}}\right| \le \|X_{\cdot \boldsymbol{v}}\| \, \|M\| \, \|\mathcal{S}\|_{\max \to \|\cdot\|} \, \|X\|_{(0)} \, \|Y_{\cdot \boldsymbol{v}}\| \, . \tag{7.33}$$

Therefore, using (7.6), (7.12), (7.13), (7.25), and (7.33), we obtain the very-high-probability bound

$$\frac{1}{N}\left|\big((G-M)^* M \mathcal{S}[G-M] G\big)_{\boldsymbol{v}\boldsymbol{v}}\right| \lesssim \frac{1}{\langle z\rangle \sqrt{N}} \, \|G-M\|_*^2 \, \|G_{\cdot \boldsymbol{v}}\| \lesssim N^{-\xi+\delta} \Psi^2 \psi. \tag{7.34}$$

Next, we turn to estimating the first term on the right-hand side of (7.32) using the multivariate cumulant expansion formula,

$$\begin{aligned} &\frac{1}{N}\left| \mathbb{E}\big[\big((G-M)^* M \underline{WG}\big)_{\boldsymbol{v}\boldsymbol{v}} \overline{S} |S|^{p-2}\big] \right| \\ &\quad \lesssim \frac{1}{N^2}\left| \sum_{abc} \sum_{\alpha_1} \kappa(ab, \alpha_1) G_{b\boldsymbol{v}} M_{ca} \partial_{\alpha_1} \big\{ (G-M)^*_{\boldsymbol{v}c} \overline{S} |S|^{p-2} \big\} \right| \\ &\qquad + \frac{1}{N} \sum_{k=2}^{L} \left| \sum_{abc} \sum_{\boldsymbol{\alpha} \in \mathcal{N}(ab)^k} \frac{\kappa(ab, \boldsymbol{\alpha})}{N^{(k+1)/2} k!} M_{ca} \partial_{\boldsymbol{\alpha}} \big\{ G_{b\boldsymbol{v}} (G-M)^*_{\boldsymbol{v}c} \overline{S} |S|^{p-2} \big\} \right| \\ &\qquad + \Omega_L^{\boldsymbol{v}}, \end{aligned} \tag{7.35}$$

where for sufficiently large integer L, the error term $\Omega_L^{\boldsymbol{v}}$ admits the bound $\Omega_L^{\boldsymbol{v}} \lesssim \Psi^{2p}$, and is therefore negligible.

We bound the term involving the second cumulants in (7.35). First, for the term containing $\partial_{\alpha_1}(G-M)^*_{\boldsymbol{v}c}$, completely analogously to (7.24), we obtain

$$\frac{1}{N^2} \sum_c \left| \sum_{b a_1 b_1} \kappa\big((M\boldsymbol{e}_c) b, a_1 b_1\big) G_{b\boldsymbol{v}} G^*_{\boldsymbol{v} a_1} G^*_{b_1 c} \right| \lesssim \|\kappa\|_2^{\mathrm{iso}} \, N^{\xi+2\delta} \Psi^2, \quad \text{w.v.h.p.}, \tag{7.36}$$

where the additional summation over the index c is compensated by the N^{-1} prefactor. Next, we estimate the terms arising from $\partial_{\alpha_1} S$. We focus on the term containing $((G-M)^* \partial_{\alpha_1} G)_{\boldsymbol{v}\boldsymbol{v}}$, other terms are estimated similarly. For the cross part κ_{c}, we obtain (ignoring the factor $|S|^{p-2}$ temporarily)

$$\begin{aligned} &\frac{1}{N^3}\left| \sum_{cb} \sum_{\alpha_1} \kappa_{\mathrm{c}}\big((M\boldsymbol{e}_c) b, \alpha_1\big) G_{b\boldsymbol{v}} (G-M)^*_{\boldsymbol{v}c} \big((G-M)^* \partial_{\alpha_1} G\big)_{\boldsymbol{v}\boldsymbol{v}} \right| \\ &\quad \lesssim \frac{1}{N^2} \sum_{cd} \left| (G-M)^*_{\boldsymbol{v}c} (G-M)^*_{\boldsymbol{v}d} \right| \frac{1}{N} \sum_{bb_1} \left| \sum_{a_1} \kappa_{\mathrm{c}}\big((M\boldsymbol{e}_c) b, a_1 b_1\big) G_{d a_1} \right| \left| G_{b\boldsymbol{v}} G_{b_1 \boldsymbol{v}} \right| \\ &\quad \lesssim |||\kappa|||_2^{\mathrm{iso}} N^{\xi+2\delta} \Psi^2 \frac{1}{N^2} \sum_{cd} \left| (G-M)^*_{\boldsymbol{v}c} (G-M)^*_{\boldsymbol{v}d} \right| \\ &\quad \lesssim |||\kappa|||_2^{\mathrm{iso}} N^{\xi+2\delta} \Psi^2 \, \|G-M\|_*^2 \lesssim \|\kappa\|_2^{\mathrm{iso}} \, N^{\xi+2\delta} \Psi^4 \psi^2, \quad \text{w.v.h.p.}, \end{aligned} \tag{7.37}$$

where in the second step we used the bound analogous to (7.24) for each c, d, and in the last step we used (7.13).

Similar estimates hold for terms involving higher order cumulants in (7.35). For example, identifying $\alpha_i := (a_i, b_i)$,

$$\begin{aligned} N^{-7/2}&\Big|\sum_{abc}\sum_{\alpha_1,\alpha_2}\kappa(ab,\alpha_1,\alpha_2)M_{ca}(G-M)^*_{\boldsymbol{v}c}\big(\partial_{\alpha_1}G\big)_{b\boldsymbol{v}}\big((G-M)^*\partial_{\alpha_2}G\big)_{\boldsymbol{v}\boldsymbol{v}}\Big| \\ &\lesssim N^{-7/2}\sum_{cd}\big|(G-M)^*_{\boldsymbol{v}d}(G-M)^*_{\boldsymbol{v}c}\big|\Big|\sum_{ab}\sum_{\alpha_1,\alpha_2}\kappa(ab,\alpha_1,\alpha_2)M_{ca}G_{ba_1}G_{b_1\boldsymbol{v}}G_{da_2}G_{b_2\boldsymbol{v}}\Big| \\ &\lesssim |||\kappa|||_3\,\|G-M\|_*^2\,N^{\xi+3\delta}\Psi^2 \lesssim |||\kappa|||_3 N^{\xi+3\delta}\langle z\rangle^{-1}\Psi^4\psi^2, \quad \text{w.v.h.p.} \end{aligned} \tag{7.38}$$

Therefore, we obtain, using the very-high-probability bound $S \lesssim \psi\Psi$ by (7.13),

$$\frac{1}{N}\big|\,\mathbb{E}\big[\big((G-M)^*M\underline{WG}\big)_{\boldsymbol{v}\boldsymbol{v}}\overline{S}|S|^{p-2}\big]\,\big| \lesssim (\Psi)^{2p}\Big(N^{\xi}N^{8\delta}+N^{-\delta}\psi^2\Big)^p, \tag{7.39}$$

hence, using (7.32) and (7.34), we deduce that with very high probability,

$$\sqrt{S} \lesssim N^{\nu}\Psi\big(N^{\xi/2+4\delta}+N^{-\delta/2}\psi\big). \tag{7.40}$$

It follows from (7.6), (3.6), (7.30) and (7.40), that

$$\Psi^{-1}\|G-M\|_* \lesssim \psi \text{ w.v.h.p.} \Longrightarrow \Psi \lesssim N^{\xi/2+4\delta+\nu}(1+\beta^{-1}) + N^{-\delta/2+\nu}\psi \text{ w.v.h.p.} \tag{7.41}$$

By iteration, this implies that $\Psi^{-1}\|G-M\|_* \lesssim N^{\xi/2+4\delta+\nu}(1+\beta^{-1}) \lesssim N^{3\xi/4+4\delta+\nu}$ with very high probability, since $\beta \geq N^{-\xi/4}$ in $\mathcal{D}^{\mathrm{glob}}$.

This concludes the proof of Lemma 7.3. □

7.3. Proof of the averaged bound in Proposition 3.3. We conclude this section by proving the averaged law in Proposition 3.3 using the isotropic law (3.5a), proved in Sect. 7.2 above, as an input.

Proof. (Proof of the averaged law in (3.5b)) Fix a deterministic matrix B and a spectral parameter $z \in \mathcal{D}^{\mathrm{glob}}$, and let $R := \langle (G-M)B\rangle$. Using the equation (7.16), we compute the p-th (for even p) moment of R,

$$\mathbb{E}\big[|R|^p\big] \leq \big|\mathbb{E}\big[\big\langle M\mathcal{S}[G-M](G-M)\widetilde{B}\big\rangle\overline{R}|R|^{p-2}\big]\big| + \big|\mathbb{E}\big[\big\langle M\underline{WG}\widetilde{B}\big\rangle\overline{R}|R|^{p-2}\big]\big|, \tag{7.42}$$

where we denote $\widetilde{B} := \big((\mathcal{B}^{-1})^*[B^*]\big)^*$. By (7.2) and Lemma 7.1, the observable $\widetilde{B}$ satisfies

$$\big\|\widetilde{B}\big\|_{\mathrm{hs}} \lesssim \big(1+\beta^{-1}\big)\,\|B\|_{\mathrm{hs}}\,. \tag{7.43}$$

To bound the first term on the right-hand side of (7.42), we employ the polar decomposition $\widetilde{B} = \sum_j \sigma_j \boldsymbol{v}_j\boldsymbol{u}_j^*$, where $\sigma_j := \sigma_j(\widetilde{B})$ and $\boldsymbol{u}_j := \boldsymbol{u}_j(\widetilde{B})$, $\boldsymbol{v}_j := \boldsymbol{v}_j(\widetilde{B})$ are the singular values and corresponding left and right, respectively, singular vectors of $\widetilde{B}$. It follows from (3.5a), (7.21), and (7.43), that with very high probability,

$$\big|\big\langle M\mathcal{S}[G-M](G-M)\widetilde{B}\big\rangle\big| \leq \frac{1}{N}\sum_j |\sigma_j|\big|\big(M\mathcal{S}[G-M](G-M)\big)_{\boldsymbol{u}_j\boldsymbol{v}_j}\big|$$

$$\lesssim N^{2\xi}\big(1+\beta^{-1}\big)\Psi^2\,\|B\|_{\mathrm{hs}}\,, \tag{7.44}$$

where $\Psi := \Psi(z)$ is defined in (3.6).

Next, we bound the second term on the right-hand side of (7.42) using the multivariate cumulant expansion formula from Proposition 5.2,

$$\begin{aligned}
&\Big|\mathbb{E}\big[\langle M\underline{WG}\widetilde{B}\rangle\overline{R}|R|^{p-2}\big]\Big| \\
&\quad\leq \bigg|\mathbb{E}\bigg[\frac{1}{N^2}\sum_{ab}\sum_{\alpha_1}\kappa(ab,\alpha_1)\big(G\widetilde{B}M\big)_{ba}\partial_{\alpha_1}\big\{\overline{R}|R|^{p-2}\big\}\bigg]\bigg| \\
&\quad\quad+\sum_{k=2}^{L}\bigg|\mathbb{E}\bigg[\frac{1}{N}\sum_{ab}\sum_{\boldsymbol{\alpha}\in\mathcal{N}(ab)^k}\frac{\kappa(ab,\boldsymbol{\alpha})}{N^{(k+1)/2}k!}\partial_{\boldsymbol{\alpha}}\big\{\big(G\widetilde{B}M\big)_{ba}\overline{R}|R|^{p-2}\big\}\bigg]\bigg| \\
&\quad\quad+\big|\Omega_L^B\big|.
\end{aligned} \tag{7.45}$$

Here, once again Ω_L^B is an error term satisfying $|\Omega_L^B| \lesssim (\sqrt{\langle z\rangle/(N\eta)}\Psi\,\|B\|_{\mathrm{hs}})^p$ for large enough L, controlled similarly to (5.17). The terms involving second order cumulants admit the bound (ignoring the common $|R|^{p-2}$ factor)

$$\begin{aligned}
&\bigg|\frac{1}{N^2}\sum_{ab}\sum_{\alpha_1}\kappa(ab,\alpha_1)\big(G\widetilde{B}M\big)_{ba}\partial_{\alpha_1}R\bigg| \\
&\quad\leq \bigg|\frac{1}{N^3}\sum_{ab}\sum_{a_1b_1}\kappa(ab,a_1b_1)\big(G\widetilde{B}M\big)_{ba}\big(GBG\big)_{b_1a_1}\bigg| \\
&\quad\leq \frac{1}{\langle z\rangle N^2\eta^2}\big\|\,|\kappa(*,*)|\,\big\|\big\langle\widetilde{B}\widetilde{B}^*\Im G\big\rangle^{1/2}\big\langle BB^*\Im G\big\rangle^{1/2} \\
&\quad\lesssim N^{\xi}\big(1+\beta^{-1}\big)|||\kappa|||_2\frac{\langle z\rangle}{N\eta}\Psi^2\,\|B\|_{\mathrm{hs}}^2\,, \quad \text{w.v.h.p.,}
\end{aligned} \tag{7.46}$$

where in the second step we used the norm bound (7.7). Here, in the last step, we used the established isotropic law (3.5a), the spectral decomposition of $\widetilde{B}\widetilde{B}^*$ and (7.43) to assert that, with very high probability,

$$\begin{aligned}
\frac{\langle\widetilde{B}\widetilde{B}^*\Im G\rangle}{N\eta} &= \frac{1}{N^2\eta}\sum_j|\sigma_j|^2\big(\Im G\big)_{\boldsymbol{u}_j\boldsymbol{u}_j} \lesssim \frac{N^{\xi}}{N^2\eta}\sum_j|\sigma_j|^2\bigg(\rho+\sqrt{\frac{\rho}{N\eta}}+\frac{1}{N\eta}\bigg) \\
&\lesssim N^{\xi}\big(1+\beta^{-1}\big)^2\langle z\rangle^2\Psi^2\|B\|_{\mathrm{hs}}^2,
\end{aligned} \tag{7.47}$$

where σ_j and $\boldsymbol{u}_j$ are the singular values and left singular vectors of $\widetilde{B}$. Similar bound without the factor $(1+\beta^{-1})^2$ holds for B instead of $\widetilde{B}$. Note that, unlike for the isotropic law (3.5a), for the current proof of the average law there is no need to split the second order cumulant into direct and cross terms, the simpler bound (2.6) suffices.

Next, we estimate the terms in (7.45) involving third order cumulants. Consider the term containing a single (∂R). Dropping $|R|^{p-2}$, we obtain

$$\begin{aligned}
&\bigg|N^{-5/2}\sum_{ab}\sum_{\alpha_1,\alpha_2}\kappa(ab,\alpha_1,\alpha_2)\big(\partial_{\alpha_1}G\widetilde{B}M\big)_{ba}(\partial_{\alpha_2}R)\bigg| \\
&\quad\lesssim N^{-7/2}\,\|M\|\max_{cd}|G_{cd}|\sum_{ab}\sum_{\alpha_1\alpha_2}\big|\kappa(ab,\alpha_1,\alpha_2)\big|\big|\big(G\widetilde{B}\widetilde{B}^*G^*\big)_{b_1b_1}\big|^{1/2}\big(GBG\big)_{\alpha_2}
\end{aligned}$$

$$\lesssim \langle z\rangle^{-2}\Big\|\sum_{ab}|\kappa(ab,*,*)|\Big\|\sqrt{\frac{\langle \widetilde{B}\widetilde{B}^*\Im G\rangle}{N\eta}}\sqrt{\frac{\langle BB^*\Im G\rangle}{N^3\eta^3}} \lesssim N^{\xi}\big(1+\beta^{-1}\big)\frac{\lvert\lvert\lvert\kappa\rvert\rvert\rvert_3}{N\eta}\Psi^2\,\|B\|_{\mathrm{hs}}^2\,, \tag{7.48}$$

with very high probability, where we used (3.5a), (7.47), and the bounds

$$\big|(G\widetilde{B}M)_{ab}\big| \le \|M\|\,\big|\big(G\widetilde{B}\widetilde{B}^*G^*\big)_{aa}\big|^{1/2}, \quad \frac{1}{N}\sum_{ab}|(GBG)_{ab}|^2 \le \frac{1}{\eta^3}\langle BB^*\Im G\rangle. \tag{7.49}$$

The term containing $(\partial^2 R)$ admits a completely analogous estimate.

For the term containing $(\partial R)^2$, we obtain, dropping $|R|^{p-3}$,

$$\begin{aligned}
&\Big|N^{-5/2}\sum_{ab}\sum_{\alpha_1,\alpha_2}\kappa(ab,\alpha_1,\alpha_2)\big(G\widetilde{B}M\big)_{ba}(\partial_{\alpha_1}R)(\partial_{\alpha_2}R)\Big| \\
&\quad\lesssim N^{-9/2}\max_{\alpha}\big|(GBG)_{\alpha}\big|\sum_{ab,\alpha_2}\sum_{\alpha_1}\big|\kappa(ab,\alpha_1,\alpha_2)\big|\big|\big(G\widetilde{B}M\big)_{ba}\big(GBG\big)_{\alpha_2}\big| \\
&\quad\lesssim N^{\xi}\langle z\rangle^2\Psi^2\,\|B\|_{\mathrm{hs}}\,\Big\|\sum_{\alpha_1}|\kappa(*,\alpha_1,*)|\Big\|\sqrt{\frac{\langle\widetilde{B}MM^*\widetilde{B}^*\Im G\rangle}{N\eta}}\sqrt{\frac{\langle BB^*\Im G\rangle}{N^3\eta^3}} \\
&\quad\lesssim N^{2\xi}\big(1+\beta^{-1}\big)\frac{\lvert\lvert\lvert\kappa\rvert\rvert\rvert_3}{N\eta}\langle z\rangle^3\Psi^4\|B\|_{\mathrm{hs}}^3, \quad \text{w.v.h.p.,}
\end{aligned} \tag{7.50}$$

where we used the local law (3.5a) to assert that, with very high probability,

$$\frac{1}{N^{3/2}}\big|(GBG)_{ab}\big| \lesssim \frac{\|B\|}{\sqrt{N}}\frac{\sqrt{\big(\Im G\big)_{aa}\big(\Im G\big)_{bb}}}{N\eta} \lesssim N^{\xi}\langle z\rangle^2\Psi^2\,\|B\|_{\mathrm{hs}}\,. \tag{7.51}$$

Note that in estimating $\max_\alpha|(GBG)_\alpha|$, we need to use the operator norm $\|B\|$ since no summation on indices is available. We convert it into $\|B\|_{\mathrm{hs}}$ at a costs of an extra $\sqrt{N}$ factor, as $\|B\| \le \sqrt{N}\,\|B\|_{\mathrm{hs}}$, but this is affordable since we collected sufficiently many powers of $N^{-1/2}$ in the third cumulant term.

Finally, we estimate the term with no (∂R), namely, dropping $|R|^{p-1}$

$$\Big|N^{-5/2}\sum_{ab}\sum_{\alpha_1,\alpha_2}\kappa(ab,\alpha_1,\alpha_2)G_{ba_1}G_{b_1a_2}\big(G\widetilde{B}M\big)_{b_2a}\Big|. \tag{7.52}$$

For both G_{ba_1} and $G_{b_1a_2}$, we write $G_{ab} = M_{ab} + (G-M)_{ab}$ and use the bound $|M_{ab}| \lesssim \langle z\rangle^{-1}$, $|(G-M)_{ab}| \lesssim N^{\xi}\Psi$, w.v.h.p., that follow from (7.7) and (3.5a), respectively, to estimate the contributions coming from the deterministic and the fluctuating part separately. In particular, we obtain the very-high-probability bound,

$$\begin{aligned}
&\Big|N^{-5/2}\sum_{ab}\sum_{\alpha_1,\alpha_2}\kappa(ab,\alpha_1,\alpha_2)(G-M)_{ba_1}M_{b_1a_2}\big(G\widetilde{B}M\big)_{b_2a}\Big| \\
&\qquad\lesssim N^{-1/2+\xi}\Psi\langle z\rangle^{-2}\Big\|\sum_{\alpha_1}|\kappa(*,\alpha_1,*)|\Big\|\big\langle G\widetilde{B}\widetilde{B}^*G^*\big\rangle^{1/2} \\
&\qquad\lesssim \langle z\rangle^{-1}N^{2\xi}\big(1+\beta^{-1}\big)\lvert\lvert\lvert\kappa\rvert\rvert\rvert_3\Psi^2\,\|B\|_{\mathrm{hs}}\,.
\end{aligned} \tag{7.53}$$

The contributions coming from $M_{ba_1}(G-M)_{b_1a_2}$ and $(G-M)_{ba_1}(G-M)_{b_1a_2}$ admit analogous estimates. Therefore, it remains to bound the contribution coming from $M_{ba_1}M_{b_1a_2}$. Using (2.8), we estimate

$$\begin{aligned}&\left|N^{-5/2}\sum_{ab}\sum_{\alpha_1,\alpha_2}\kappa(ab,\alpha_1,\alpha_2)M_{ba_1}M_{b_1a_2}\big(G\widetilde{B}M\big)_{b_2a}\right|\\&\quad\le N^{-1}\vert\kern-0.25ex\vert\kern-0.25ex\vert\kappa\vert\kern-0.25ex\vert\kern-0.25ex\vert_3^{\mathrm{av}}\,\|M\|^2\,\big\|G\widetilde{B}M\big\|_{\mathrm{hs}}\\&\quad\lesssim N^{\xi/2}\big(1+\beta^{-1}\big)\langle z\rangle^{-2}N^{-1/2}\Psi\,\|B\|_{\mathrm{hs}}\,,\quad \text{w.v.h.p.}\end{aligned}\tag{7.54}$$

Putting back the dropped $|R|$ factors into the estimates (7.46), (7.48), (7.50), (7.53) and (7.54) and using the Young's inequality to separate these factors into an additive $|R|^p$ term with a small multiplicative constant, we see that the second and third order cumulant terms in (7.45) can be estimated by $\big(N^{3\xi/2}\langle z\rangle^{1/2}(N\eta)^{-1/2}\Psi\,\|B\|_{\mathrm{hs}}\big)^p+N^{-p\xi/4}|R|^p$. Here we used $\beta\ge N^{-\xi/4}$ from (3.4).

Estimating the terms involving fourth and higher order cumulants using simple power counting, similarly to (5.30)–(5.31), we deduce that

$$\mathbb{E}\big[|R|^p\big]\lesssim\left(N^{3\xi/2}\langle z\rangle^{1/2}(N\eta)^{-1/2}\Psi\,\|B\|_{\mathrm{hs}}\right)^p+N^{-p\xi/4}\mathbb{E}\big[|R|^p\big].\tag{7.55}$$

This concludes the proof of (3.5b). □

Acknowledgement We thank Giorgio Cipolloni for many productive discussions and the anonymous referees for several useful suggestions and spotting some typos.

Funding Open access funding provided by Institute of Science and Technology (IST Austria).

Data Availability There is no data associated to this work.

Declarations

Conflict of interest The authors have no Conflict of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A. Technical Lemmas

In this appendix, we collect the proofs of several technical lemmas used throughout this paper.

Proof of Lemma 3.4. Observe that for any $s\ge 0$ and any initial condition H, the distribution of the random matrix $\mathfrak{F}^s_{\mathrm{zag}}[H]$ satisfies

$$\mathfrak{F}^s_{\mathrm{zag}}\big[H\big]\overset{d}{=}\mathbb{E}[H]+\mathrm{e}^{-s/2}\big(H-\mathbb{E}H\big)+\sqrt{1-\mathrm{e}^{-s}}\,\Sigma_H^{1/2}\big[W_{\mathrm{G}}\big],\tag{A.1}$$

where W_{G} is a standard GUE/GOE random matrix (in the same symmetry class as H) independent of H. Moreover, if $\Sigma_H \geq c\Sigma_{\mathrm{G}}$ for some constant $0 < c < 1$, then there exists a random matrix $\widehat{W}$ with $\mathbb{E}\widehat{W} = 0$, such that

$$\Sigma_H^{1/2}\big[W_{\mathrm{G}}\big] \overset{d}{=} \widehat{W} + \sqrt{c}\,\widetilde{W}_{\mathrm{G}}, \tag{A.2}$$

where $\widetilde{W}_{\mathrm{G}}$ is a GUE/GOE matrix independent of $\widehat{W}$. Therefore,

$$\mathfrak{F}^s_{\mathrm{zag}}\big[H\big] \overset{d}{=} \widehat{H}^s + \sqrt{c}\sqrt{1-\mathrm{e}^{-s}}\,\widetilde{W}_{\mathrm{G}}, \quad \widehat{H}^s \overset{d}{:=} \mathbb{E}[H] + \mathrm{e}^{-s/2}\big(H - \mathbb{E}H\big) + \sqrt{1-\mathrm{e}^{-s}}\,\widehat{W}, \tag{A.3}$$

where $\widetilde{W}_{\mathrm{G}}$ is independent of $\widehat{H}^s$. Hence, (3.13) follows immediately from (3.8) and (A.3) for $\mathfrak{H}_{c,t}(H)$ defined as

$$\mathfrak{H}_{c,t}(H) := \mathrm{e}^{t/2}\bigg(\mathbb{E}H + \mathrm{e}^{-s(t)/2}\big(H - \mathbb{E}H\big) + \sqrt{1-\mathrm{e}^{-s(t)}}\,\widehat{W}\bigg), \tag{A.4}$$

where the random matrix $\widehat{W}$ independent of H satisfies (A.2), and $s(t) \equiv s_c(t)$ is defined in (3.14).

The estimate (3.15) is a direct consequence of (3.14). This concludes the proof of Lemma 3.4. □

Proof of Lemma 3.5. Fix a time $0 \leq t \leq T$ and let z_s denote the solution of (3.18) that satisfies $z_t \in \mathcal{D}^{\mathrm{abv}}_t$. It follows from (3.18) and (3.20) that for all $s \in [0, t]$,

$$\mathrm{d}\big(\eta_s \rho_s(z_s)\big) = -\pi\rho_s(z_s)^2\mathrm{d}s \leq 0, \tag{A.5}$$

where we denote $\eta_s := \Im z_s$. A similar computation reveals that

$$\mathrm{d}\big(\rho_s(z_s)^{-1}\eta_s\big) = -\big(\rho_s(z_s)^{-1}\eta_s + \pi\big)\mathrm{d}s \leq -\pi\,\mathrm{d}s, \tag{A.6}$$

since $\rho_s^{-1}(z)\Im z \geq 0$ for all $z \in \mathbb{C}$. Moreover, it follows from Assumption 2.5 that $|\mathrm{d}z_s/\mathrm{d}s| \lesssim C'$ for all $0 \leq s \leq T$ and all $z_t \in \mathcal{D}^{\mathrm{abv}}_t$, hence, using the estimates (A.5) and (A.6), we deduce that $z_s \in \mathcal{D}^{\mathrm{abv}}_s$ for all $s \in [0, t]$. This concludes the proof of Lemma 3.5. □

Proof of Lemma 4.1. Clearly, for terminal times $0 \leq T \lesssim 1$, the solutions to (3.17) satisfy $\|A_t - A_T\| \lesssim T - t$ and $\|\mathcal{S}_t - \mathcal{S}_T\|_{\|\cdot\|\to\|\cdot\|} \lesssim T - t$, for all $0 \leq t \leq T$. Therefore, for some sufficiently small threshold $T_* \sim 1$, the first bound in (4.1) follows immediately from Assumption 2.5 and the stability of the MDE against small perturbations of the data pair, see Section 10 in [10]. Moreover, it follows from the fullness Assumption 2.4, that, by possibly shrinking the threshold T_*, we can guarantee that $\mathcal{S}_t[X] \sim \langle X\rangle$ for any Hermitian matrix $X \geq 0$. Hence, the second bound in (4.1) follows from Proposition 3.5 in [10] and the first bound in (4.1). This concludes the proof of Lemma 4.1. □

Proof of Lemma 5.3. Throughout the proof, we consider the time $s \in [0, s_{\mathrm{final}}]$ to be fixed, and drop it from the superscript of G^s. The uniformity of all estimates in s follows trivially from the assumptions of Lemma 5.3.

First, we prove the second estimate in (5.6). The map $\eta \mapsto \eta^2/(x^2+\eta^2)$ is increasing in $\eta > 0$ for any $x \in \mathbb{R}$, hence it follows by spectral decomposition of $\Im G$ that

$$\eta_1\Im G(E + \mathrm{i}\eta_1) \leq \eta_0\Im G(E + \mathrm{i}\eta_0), \tag{A.7}$$

in the sense of quadratic forms. Therefore, the second estimate in (5.6) follows immediately from (5.2).

Next, we prove the first estimate in (5.6). Using the Schwarz inequality and the Ward identity, we deduce that for all $0 < \eta < \eta_0$,

$$\left|\frac{\mathrm{d}}{\mathrm{d}\eta}\big(G(E+\mathrm{i}\eta)\big)_{\boldsymbol{uv}}\right| \lesssim \frac{\left|\big(\Im G(E+\mathrm{i}\eta)\big)_{\boldsymbol{uu}}\big(\Im G(E+\mathrm{i}\eta)\big)_{\boldsymbol{vv}}\right|^{1/2}}{\eta} \lesssim \frac{\eta_0}{\eta^2}\rho(E+\mathrm{i}\eta_0), \quad \text{(A.8)}$$

where in the second step we used the monotonicity of the maps $\eta \mapsto \eta\Im G(E+\mathrm{i}\eta)$ and $\eta \mapsto \eta\rho(E+\mathrm{i}\eta)$, and the second bound in (5.6) established above. Integrating the bound (A.8) from η_1 to η_0, we obtain

$$\left|\big(G(E+\mathrm{i}\eta_1)\big)_{\boldsymbol{uv}}\right| \lesssim \left|\big(G(E+\mathrm{i}\eta_0)\big)_{\boldsymbol{uv}}\right| + \frac{\eta_0}{\eta_1}\rho(E+\mathrm{i}\eta_0). \quad \text{(A.9)}$$

Since $\rho(E+\mathrm{i}\eta_0) \lesssim 1$, the first estimate in (5.6) follows immediately from (5.2) and (A.9). This concludes the proof of Lemma 5.3. □

Remark A.1 (Local Laws below the Scale) Assume that $\rho(E+\mathrm{i}\eta_1)N\eta_1 \leq N^\varepsilon$ and $\rho(E+\mathrm{i}\eta_0)N\eta_0 = N^\varepsilon$, in particular $\eta_1 \leq \eta_0$. Using (A.7) with (3.1) at $z := E+\mathrm{i}\eta_0$ as an input, we obtain the very-high-probability bound

$$\big(\Im G(E+\mathrm{i}\eta_1)\big)_{\boldsymbol{uu}} \lesssim \frac{\eta_0}{\eta_1}\rho(E+\mathrm{i}\eta_0) \lesssim \frac{\rho(E+\mathrm{i}\eta_0)N\eta_0}{N\eta_1} \lesssim \frac{N^\varepsilon}{N\eta_1}. \quad \text{(A.10)}$$

Using Lemma 7.1 and the identity $\mathrm{d}M(z)/\mathrm{d}z = \mathcal{B}^{-1}(z)[M(z)^2]$, that follows by taking the z-derivative of (2.3), we conclude that $\|\mathrm{d}M(z)/\mathrm{d}z\| \lesssim |\rho(z)/\Im z|$. Hence, differentiating $(G(E+\mathrm{i}\eta) - M(E+\mathrm{i}\eta))_{\boldsymbol{uv}}$ with respect to η, similarly to (A.8), we can deduce that

$$\left|\big(G(E+\mathrm{i}\eta_1) - M(E+\mathrm{i}\eta_1)\big)_{\boldsymbol{uv}}\right| \lesssim \frac{N^\varepsilon}{N\eta_1}, \quad \text{w.v.h.p.} \quad \text{(A.11)}$$

Analogous reasoning also applies to averaged bounds. Indeed,

$$\begin{aligned}&\left|\frac{\mathrm{d}}{\mathrm{d}\eta}\big\langle\big(G(E+\mathrm{i}\eta) - M(E+\mathrm{i}\eta)\big)B\big\rangle\right| \\ &\quad\lesssim \frac{\left|\big\langle\Im G(E+\mathrm{i}\eta)\big\rangle\big\langle\Im G(E+\mathrm{i}\eta)BB^*\big\rangle\right|^{1/2} + \rho(E+\mathrm{i}\eta)\,\|B\|_{\mathrm{hs}}}{\eta}. \end{aligned} \quad \text{(A.12)}$$

Therefore, by integrating (A.12) in η and using (3.1), we can deduce that

$$\left|\big\langle\big(G(E+\mathrm{i}\eta_1) - M(E+\mathrm{i}\eta_1)\big)B\big\rangle\right| \lesssim \frac{N^\varepsilon}{N\eta_1}\|B\|_{\mathrm{hs}}, \quad \text{w.v.h.p.} \quad \text{(A.13)}$$

These results show that the local laws (3.1) hold at $z = E+\mathrm{i}\eta_1$, for any $0 < \eta_1 \leq \eta_0$, once they hold at $E+\mathrm{i}\eta_0$ with η_0 satisfying $\rho(E+\mathrm{i}\eta_0)N\eta_0 = N^\varepsilon$.

Proof of Lemma 6.2. First, we prove (6.2). Let σ_t be function defined in (7.3), corresponding to the solution M_t of the time-dependent MDE (3.16). It follows from Lemma 5.5 in [10] that σ_t admits a uniformly 1/3-Hölder regular extension $\overline{\mathbb{H}}$. Moreover, it follows from Lemma 7.16 in [10] that $|\sigma_t(\mathfrak{e}_t^-)| \sim |\sigma_t(\mathfrak{e}_t^+)| \sim \Delta_t^{1/3}$ and it follows from Theorem 7.7 (ii.b) in [10] that $\sigma_t(\mathfrak{e}_t^-) < 0$ and $\sigma_t(\mathfrak{e}_t^+) > 0$. Therefore, there exists a point $x_t \in (\sigma_t(\mathfrak{e}_t^-), \sigma_t(\mathfrak{e}_t^+))$ satisfying $\sigma_t(x_t) = 0$. For any $0 \le s \le t$, let $x_s := \varphi_{s,t}(x_t)$ as defined in (3.19). It follows from (3.20) that $\sigma_s(x_s) = 0$ for any $s \in [0, t]$.

Furthermore, $1/3$-Hölder regularity of ρ_t in t implies that there exists $c \sim 1$ such that for times s satisfying $0 \le t - s \le c\Delta_t^{1/3}$, the density ρ_s has a gap in the support around x_s of size $\Delta_s > 0$, let $\mathfrak{e}_s^-$ and $\mathfrak{e}_s^+$ denote its endpoints. From $1/3$-Hölder regularity of σ_s, we infer that $\mathrm{dist}(x_s, \mathfrak{e}_s^\pm) \sim \Delta_s$.

On the other hand, the map $h_s : x \mapsto \lim_{\eta\to+0} \rho_s(x + \mathrm{i}\eta)^{-1}\eta$ is also $1/3$-Hölder regular, uniformly in s, hence $h_s(x_s) \sim \mathrm{dist}(x_s, \mathfrak{e}_s^\pm)^{1/2}\Delta_s^{1/6} \sim \Delta_s^{2/3}$ by (6.1). Along the trajectories of (3.18), $h_s(x_s) \sim h_t(x_t) + (t - s)$ for all s satisfying $0 \le t - s \le c\Delta_t^{1/3}$, therefore (6.2) holds for all $0 \le t - s \le c\Delta_t^{1/3}$. In particular, $\Delta_{t-c\Delta^{1/3}} \gtrsim \Delta_t + \Delta_t^{1/2}$, which implies that (6.2) holds for all $0 \le s \le t$. This concludes the proof of (6.2).

Next, we prove (6.3). A similar relation for the evolution of the gaps under the free semicircular flow was studied in Section 5.1 of [36]. To keep the present paper reasonably self-contained, we present a complete proof for the evolution under the characteristic flow (3.18).

Observe that it follows immediately from (3.16) that the density $\rho_s(x)$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}x}\rho_s(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x}{2}\rho_s(x) + \langle \Re M_s(x)\rangle \rho_s(x)\right), \quad x \in \mathbb{R}, \quad 0 \le s \le t. \tag{A.14}$$

Consider the mass of ρ_s that lies to the left of the point x_s. Equation (A.14) implies that

$$\frac{\mathrm{d}}{\mathrm{d}s}\int_{-\infty}^{x_s} \rho_s(x)\mathrm{d}x = 0, \quad 0 \le s \le t, \tag{A.15}$$

where we used that $\rho_s(x_s) = 0$. Therefore, the mass of the band of ρ_s to the left of $\mathfrak{e}_s^-$ is constant $0 \le s \le t$. For any $r > 0$, define $\gamma_s(r)$ implicitly by

$$\int_{-\infty}^{\gamma_s(r)} \rho_s(x)\mathrm{d}x = \int_{-\infty}^{x_s} \rho_s(x)\mathrm{d}x - r. \tag{A.16}$$

Note that by the definition of the edge point $\mathfrak{e}_s^-$ and the structure theorem [10, Theorem 7.2 (ii)] for ρ_s, there exists a constant $\widetilde{c} > 0$ such that $\rho_s(\gamma_s(r)) > 0$ for all $0 \le r \le \widetilde{c}$ and all $0 \le s \le t$. Moreover, $\gamma_s(r) < \mathfrak{e}_s^- \le x_s$. Therefore, it follows from (A.14) that (a similar equation for the free semicircular flow was obtained in Section 4.1 of [32])

$$\frac{\mathrm{d}}{\mathrm{d}s}\gamma_s(r) = -\frac{1}{2}\gamma_s(r) - \big\langle \Re M_s\big(\gamma_s(r)\big)\big\rangle, \quad 0 \le r \le \widetilde{c}, \quad 0 \le s \le t. \tag{A.17}$$

The evolution equation (6.3) for $\mathfrak{e}_s^-$ follows by taking the limit $r \to 0$ in (A.17). An analogous argument that considers the mass of ρ_s to the right of x_s implies (6.3) for $\mathfrak{e}_s^+$.

Next, we prove the first estimate in (6.4). By taking the imaginary parts of (3.18) and (3.20), we obtain

$$\eta_s \sim \eta_t + \rho_t(t - s), \quad \rho_s \sim \rho_t, \quad 0 \le s \le t. \tag{A.18}$$

Moreover, it follows form the comparison relation for ρ_t from (6.1), that

$$\rho_t \sim \eta_t(\varkappa_t + \eta_t)^{-1/2}(\Delta_t + \varkappa_t + \eta_t)^{-1/6} \lesssim \eta_t^{1/2}\Delta_t^{-1/6}, \tag{A.19}$$

where we used $0 \le \varkappa_t \le \Delta_t$ and the assumption that $\eta_t \lesssim N^{-\nu}\Delta_t$. Therefore, using (6.2), (A.18) and (A.19), we obtain

$$\eta_s \lesssim \eta_t^{1/2}\Delta_t^{-1/6}\big(\Delta_t+(t-s)^{3/2}\big)^{2/3} \lesssim N^{-\nu/2}\Delta_t^{1/3}\Delta_s^{2/3} \lesssim N^{-\nu/2}\Delta_s, \quad 0\le s\le t, \tag{A.20}$$

hence the first bound in (6.4) is established. To prove the second relation in (6.4), observe that it suffices to show that for all $0 \le s \le t$ and all $\eta \lesssim N^{-\nu/2}\Delta_s$, we have the bound

$$\pm\Re\big[F_s^{\pm}(\eta)-F_s^{\pm}(+0)\big]>0, \quad F_s^{\pm}(\eta):=-\frac{1}{2}(\mathfrak{e}_s^{\pm}+\mathrm{i}\eta)-\langle M_s(\mathfrak{e}_s^{\pm}+\mathrm{i}\eta)\rangle. \tag{A.21}$$

Indeed, (A.21) implies that along (3.18), for all points at level $\eta \lesssim N^{-\nu/2}\Delta_s$ above the ends $\mathfrak{e}_s^{\pm}$ of the gap, their projection onto the real line moves away from the gap, for all times $0 \le s \le t$. Hence no trajectory $z_s = E_s + \mathrm{i}\eta_s$ satisfying $E_t \in (\mathfrak{e}_t^-, \mathfrak{e}_t^+)$ and $\eta_t \lesssim N^{-\nu}\Delta_t$ can violate (6.4).

To see that (A.21) holds, note that the Stieltjes representation for $\langle M_s\rangle$ and the universal shape (see, e.g., [36, Eqs. (2.4a)–(2.4e)] for precise formulas)
of the density ρ_s in the vicinity of its singularities $\mathfrak{e}_s^{\pm}$ yields

$$\begin{aligned}-\Re\big[F_s^-(\eta)-F_s^-(+0)\big] &= \frac{1}{\pi}\int_{\mathbb{R}}\frac{-\eta^2}{x(x^2+\eta^2)}\rho_s(\mathfrak{e}_s^-+x)\mathrm{d}x\\ &\ge C\Delta_s^{-1/6}\eta^{1/2}+\mathcal{O}\big(\Delta_s^{-5/3}\eta^2\big)+\mathcal{O}(\eta^2)>0,\end{aligned} \tag{A.22}$$

where in the last line we used the assumption $\eta \lesssim N^{-\nu/2}\Delta_s \lesssim N^{-\nu/2}$. The computation for F_s^+ is completely analogous. This concludes the proof of (6.4).

Next, we prove (6.5). Using the comparison relations (A.18) for ρ_s and $\rho_s^{-1}\eta_s$, together with the bound $\eta_s \lesssim N^{-\nu/2}\Delta_s$, and the assumption $\varkappa_t(z_t) \gtrsim N^{\nu}\eta_t$, we deduce that

$$1+\frac{\varkappa_s}{\eta_s} \sim \frac{\rho_t^{-2}\eta_t+(t-s)}{\Delta_t^{1/3}+(t-s)^{1/2}} \gtrsim \min\left\{1+\frac{\varkappa_t}{\eta_t}, \frac{\varkappa_t^{1/2}\Delta_t^{1/2}}{\eta_t}\right\} \gtrsim \frac{\varkappa_t}{\eta_t}, \tag{A.23}$$

which implies (6.5) immediately.

Finally, we prove (6.6). Without loss of generality, we can assume $z_s = \mathfrak{e}_s^- + y_s + \mathrm{i}\eta_s$ with $0 \le y_s \le (1-C_1)\Delta_s$ for some $1 \lesssim C_1 < 3/4$. Considering the difference of the real part of (3.18) and (6.3), we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}y_s+\frac{1}{2}y_s &= \Re\big\langle M_s(\mathfrak{e}_s^-)-M_s(\mathfrak{e}_s^-+y_s)+M_s(\mathfrak{e}_s^-+y_s)-M_s(z_s)\big\rangle\\ &= -y_s\frac{1}{\pi}\int_{\mathbb{R}}\frac{\rho_s(\mathfrak{e}_s^-+x)\mathrm{d}x}{x(x-y_s)}-\eta_s^2\frac{1}{\pi}\int_{\mathbb{R}}\frac{\rho_s(\mathfrak{e}_s^-+x)\mathrm{d}x}{(y_s-x)\big((y_s-x)^2+\eta_s^2\big)},\end{aligned} \tag{A.24}$$

where in the second line we used the Stieltjes representation for $\langle M_s\rangle$. Using the universal shape of the density ρ_s near the singularities $\mathfrak{e}_s^{\pm}$, we conclude that uniformly in $0 \le s \le t \le T$,

$$y_s\int_{\mathbb{R}}\frac{\rho_s(\mathfrak{e}_s^-+x)\mathrm{d}x}{x(x-y_s)} \gtrsim \frac{y_s^{1/2}}{\Delta_s^{1/6}}, \quad \eta_s^2\left|\int_{\mathbb{R}}\frac{\rho_s(\mathfrak{e}_s^-+x)\mathrm{d}x}{(y_s-x)\big((x-y_s)^2+\eta_s^2\big)}\right| \lesssim \frac{\eta_s^2}{\Delta_s^{1/6}(y_s+\eta_s)^{3/2}}. \tag{A.25}$$

Since $\eta_s \lesssim N^{-\nu/2}\varkappa_s$ and $\varkappa_s \le y_s$, we conclude from (A.24) and (A.25) that

$$\frac{\mathrm{d}}{\mathrm{d}s} y_s \lesssim -\frac{y_s^{1/2}}{\Delta_s^{1/6}}, \tag{A.26}$$

which, together with (6.2), implies (6.6) for some constant $\mathfrak{c} > 0$. This concludes the proof of Lemma 6.2. □

Proof of Lemma 6.5. We restrict our considerations to the event Ω on which the assumed bound (6.9) holds.

First, we prove (6.10). Assume, to the contrary, that there is an eigenvalue λ of H such that $\lambda \in [\mathfrak{e}_t^- + f(t), \mathfrak{e}_t^+ - f(t)]$, then

$$\big\langle \Im G(\lambda + \mathrm{i}\eta) \big\rangle \ge \frac{1}{N\eta}, \quad \eta > 0. \tag{A.27}$$

On the other hand, choosing η implicitly such that $\rho_t(\lambda + \mathrm{i}\eta)N\eta = N^{-\zeta/2}$, implies that $\lambda + \mathrm{i}\eta \in \mathcal{D}_t^{\mathrm{sub}}$ and $\eta \sim \varkappa_t(\lambda)^{1/4}\eta_{\mathfrak{f},t}^{3/4}N^{-\zeta/4}$. Therefore, using the assumed bound (6.9) with data $(\mathcal{D}_t^{\mathrm{sub}}, \zeta + \ell\nu, \Omega)$ yields

$$\big\langle \Im G(\lambda + \mathrm{i}\eta) \big\rangle \lesssim \frac{\rho_t(\lambda + \mathrm{i}\eta)N\eta + N^{-\zeta}(\log N)^{\gamma}}{N\eta} \lesssim \frac{N^{-\zeta/2}}{N\eta}. \tag{A.28}$$

Therefore, we conclude by contradiction that (6.10) holds on Ω. This concludes the proof of Lemma 6.5. □

Appendix B. Polynomially Decaying Metric Correlation Structure

In this section, we verify the last condition in Assumption 2.3 (i) for the ensemble in Example 2.6. More precisely, we show that (2.8) holds under the assumption that (recall (2.11c) from Example 2.6)

$$\big|\kappa(\alpha_1, \alpha_2, \alpha_3)\big| \le C_3 \prod_{e \in \mathfrak{T}_{\min}} \frac{1}{1 + d(e)^s}, \tag{B.1}$$

for some[15] $s > 2$, where $\mathfrak{T}_{\min}$ is a minimal spanning tree in a complete graph with vertices $\alpha_1, \alpha_2, \alpha_3$ and edge weights induced by distance d, defined in (2.11b). That is, out goal is to show that, for all $X, Y, Z \in \mathbb{C}^{N\times N}$, the estimate

$$\begin{aligned} N^{-3/2} \sum_{\alpha_1,\alpha_2,\alpha_3} &\big|\kappa(\alpha_1, \alpha_2, \alpha_3)\big| |X_{b_1a_2}|\,|Y_{b_2a_3}|\,|Z_{b_3a_1}| \\ &\lesssim C_3\, \|X\|\, \|Y\|\, \|Z\|_{\mathrm{hs}}, \qquad \alpha_j := (a_j, b_j), \end{aligned} \tag{B.2}$$

holds for some absolute implicit constant, where C_3 is the constant from (B.1).

We estimate the contribution of the case when $(\alpha_1, \alpha_3) \in \mathfrak{T}_{\min}$ and $d(\alpha_1, \alpha_3) = |a_1 - b_3| + |b_1 - a_3|$ in full detail. It is straightforward to check that in all other cases, using the trivial bounds $|X_{b_1a_2}| \le \|X\|$, $|Y_{b_2a_3}| \le \|Y\|$ is sufficient. Indeed, since $s > 2$,

[15] The estimate (B.2) below can be proved under the relaxed summability condition $s > 3/2$. However, $s > 2$ in (2.11c) is still necessary for (2.6)–(2.7).

the indices b_1, a_2, b_2, a_3 can be summed up after using the norm bounds on X and Y; then for the remaining (a_1, b_3) sum, we use $N^{-3/2} \sum_{a_1,b_3} |Z_{a_1 b_3}| \lesssim \|Z\|_{\mathrm{hs}}$. Therefore, it suffices to bound

$$\begin{aligned} \mathfrak{X} \equiv \mathfrak{X}(X, Y, Z) &:= C_3 N^{-3/2} \sum_{\alpha_1,\alpha_3} \frac{|Z_{b_3 a_1}|}{1 + \big(|a_1 - b_3| + |b_1 - a_3|\big)^s} \\ &\quad \times \sum_{\alpha_2} \frac{|X_{b_1 a_2}|\,|Y_{b_2 a_3}|}{1 + \big(|a_1 - a_2| + |b_1 - b_2|\big)^s}, \end{aligned} \tag{B.3}$$

where we assumed for concreteness that $\mathfrak{T}_{\min} = \{(\alpha_1, \alpha_2), (\alpha_1, \alpha_3)\}$ and $d(\alpha_1, \alpha_3) = |a_1 - a_2| + |b_1 - b_2|$ (other cases are identical). First, we use the Schwarz inequality in the b_2 summation, to obtain

$$\begin{aligned} \sum_{b_2} \frac{|Y_{b_2 a_3}|}{1 + \big(|a_1 - a_2| + |b_1 - b_2|\big)^s} &\lesssim \frac{1}{1 + |a_1 - a_2|^{s-1/2}} \sqrt{\sum_{b_2} |Y_{b_2 a_3}|^2} \\ &\lesssim \frac{\|Y\|}{1 + |a_1 - a_2|^{s-1/2}}. \end{aligned} \tag{B.4}$$

Plugging (B.4) into the expression for $\mathfrak{X}$ in (B.3) and performing the summation in a_3, we obtain the estimates

$$\begin{aligned} \mathfrak{X} &\lesssim C_3 \|Y\| N^{-3/2} \sum_{a_1,b_3} \frac{|Z_{b_3 a_1}|}{1 + |b_3 - a_1|^{s-1}} \sum_{a_2} \frac{1}{1 + |a_1 - a_2|^{s-1/2}} \sum_{b_1} |X_{b_1 a_2}| \\ &\lesssim C_3 \|X\| \|Y\| N^{-1} \sum_{a_1,b_3} \frac{|Z_{b_3 a_1}|}{1 + |b_3 - a_1|^{s-1}} \sum_{a_2} \frac{1}{1 + |a_1 - a_2|^{s-1/2}} \\ &\lesssim C_3 \|X\| \|Y\| \|Z\|_{\mathrm{hs}}, \end{aligned} \tag{B.5}$$

where in the second step we used Schwarz inequality in b_1, and in the ultimate step we use the fact that $s > 2$ to first sum the convergent series in a_2, and then apply Schwarz in (a_1, b_3). This yields the desired (B.2).

References

1. Adhikari, A., Huang, J.: Dyson Brownian motion for general β and potential at the edge. Probab. Theory Relat. Fields **178**, 893–950 (2020). https://doi.org/10.1007/s00440-020-00992-9
2. Adhikari, A., Landon, B.: Local law and rigidity for unitary Brownian motion. Probab. Theory Relat. Fields **187**, 753–815 (2023). https://doi.org/10.1007/s00440-023-01230-8
3. Adhikari, A., Lemm, M.: Universal eigenvalue statistics for dynamically defined matrices. J. Anal. Math. (2023). https://doi.org/10.1007/s11854-023-0314-z
4. Adlam, B. and Che, Z. *Spectral Statistics of Sparse Random Graphs with a General Degree Distribution*. (2015). eprint: 1509.03368
5. Adler, M., Ferrari, P.L., van Moerbeke, P.: Airy Processes with wanderers and new universality classes. Ann. Probab. **38**, 714–769 (2010). https://doi.org/10.1214/09-AOP493
6. Aggarwal, A. Huang, J.: *Edge Rigidity of Dyson Brownian Motion with General Initial Data*. (2023). eprint: 2308.04236
7. Ajanki, O., Erdős, L., Krüger, T.: Universality for general Wigner-type matrices. Probab. Theory Relat. Fields **169**, 667–727 (2016). https://doi.org/10.1007/s00440-016-0740-2
8. Ajanki, O., Erdős, L., Krüger, T.: *Quadratic vector equations on complex upper half-plane*. Vol. 261, pp. 1261. American Mathematical Society, 2019. https://doi.org/10.1090/memo/1261

9. Ajanki, O.H., Erdős, L., Krüger, T.: Stability of the matrix Dyson equation and random matrices with correlations. Probab. Theory Relat. Fields **173**, 293–373 (2019). https://doi.org/10.1007/s00440-018-0835-z
10. Alt, J., Erdős, L., Krüger, T.: The Dyson equation with linear self-energy: spectral bands, edges and cusps. Doc. Math. **25**, 1421–1539 (2020). https://doi.org/10.4171/dm/780
11. Alt, J., Erdős, L., Krüger, T., Schröder, D.: Correlated random matrices: band rigidity and edge universality. Ann. Probab. **48**(2), 963–1001 (2020). https://doi.org/10.1214/19-AOP1379
12. Anderson, P.W.: Absence of diffusion in certain random lattices. Phys. Rev. **109**, 1492 (1958). https://doi.org/10.1103/PhysRev.109.1492
13. Baik, J., Kriecherbauer, T., McLaughlin, K. D.-R., Miller, P. D.: Discrete Orthogonal Polynomials.(AM-164): Asymptotics and Applications (AM-164). Vol. 164. Princeton University Press (2007). https://doi.org/10.1515/9781400837137
14. Bauerschmidt, R., Huang, J., Knowles, A., Yau, H.-T.: Bulk eigenvalue statistics for random regular graphs. Ann. Probab. **45**, 3626–3663 (2017). https://doi.org/10.1214/16-AOP1145
15. Bekerman, F., Figalli, A., Guionnet, A.: Transport maps for β-matrix models and universality. Commun. Math. Phys. **338**, 589–619 (2015). https://doi.org/10.1007/s00220-015-2384-y
16. Borodin, A., Okounkov, A., Olshanski, G.: Asymptotics of Plancherel measures for symmetric groups. J. Am. Math. Soc. **13**, 481–515 (2000). https://doi.org/10.1090/S0894-0347-00-00337-4
17. Bourgade, P.: Extreme gaps between eigenvalues of Wigner matrices. J. Eur. Math. Soc. **24**, 2823–2873 (2021). https://doi.org/10.4171/JEMS/1141
18. Bourgade, P., Erdős, L., Yau, H.-T.: Universality of general β-ensembles. Duke Math. J. **163**, 1127–1190 (2014). https://doi.org/10.1215/00127094-2649752
19. Bourgade, P., Erdös, L., Yau, H.-T.: Edge universality of beta ensembles. Commun. Math. Phys. **332**, 261–353 (2014). https://doi.org/10.1007/s00220-014-2120-z
20. Bourgade, P., Erdős, L., Yau, H.-T., Yin, J.: Universality for a class of random band matrices. Adv. Theor. Math. Phys. **21**, 739–800 (2017). https://doi.org/10.4310/ATMP.2017.v21.n3.a5
21. Bourgade, P., Yau, H.T., Yin, J.: Random band matrices in the delocalized phase i: quantum unique ergodicity and universality. Commun. Pure Appl. Math. **73**, 1526–1596 (2020). https://doi.org/10.1002/cpa.21895
22. Brézin, E., Hikami, S.: Level spacing of random matrices in an external source. Phys. Rev. E **58**, 7176 (1998). https://doi.org/10.1103/physreve.58.7176
23. Brézin, E., Hikami, S.: Universal singularity at the closure of a gap in a random matrix theory. Phys. Rev. E **57**, 4140 (1998). https://doi.org/10.1103/PhysRevE.57.4140
24. Campbell, A., Cipolloni, G., Erdős, L., Ji, H.C.: *On the spectral edge of non-Hermitian random matrices*. (2024). eprint: 2404.17512
25. Cipolloni, G., Erdős, L., Schröder, D.: Eigenstate thermalization hypothesis for Wigner matrices. Commun. Math. Phys. **388**, 1005–1048 (2021). https://doi.org/10.1007/s00220-021-04239-z
26. Cipolloni, G., Erdős, L., Schröder, D.: Optimal multi-resolvent local laws for Wigner matrices. Electron. J. Probab. **27**, 1–38 (2022). https://doi.org/10.1214/22-EJP838
27. Cipolloni, G., Erdős, L., Henheik, J.: *Eigenstate thermalisation at the edge for Wigner matrices*. (2023). eprint: 2309.05488
28. Cipolloni, G., Erdős, L., Henheik, J.: Out-of-time-ordered correlators for Wigner matrices. Adv. Theor. Math. Phys. **28**, 2025–2083 (2024). https://doi.org/10.4310/ATMP.241031013250
29. Cipolloni, G., Erdős, L., Schröder, D.: Mesoscopic central limit theorem for non-Hermitian random matrices. Probab. Theory Relat. Fields **188**, 1131–1182 (2024). https://doi.org/10.1007/s00440-023-01229-1
30. Cipolloni, G., Erdős, L., Xu, Y.: Optimal decay of eigenvector overlap for non-Hermitian random matrices. (2024). https://arxiv.org/pdf/2411.16572
31. Cipolloni, G., Erdős, L., Xu, Y.: Universality of extremal eigenvalues of large random matrices. (2023). eprint: 2312.08325
32. Cipolloni, G., Erdős, L., Krüger, T., Schröder, D.: Cusp universality for random matrices, II: the real symmetric case. Pure Appl. Anal **1**, 615–707 (2019). https://doi.org/10.2140/paa.2019.1.615
33. Deift, P., Gioev, D.: Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices. Commun. Pure Appl. Math. **60**, 867–910 (2007). https://doi.org/10.1002/cpa.20164
34. Deift, P., Kriecherbauer, T., McLaughlin, K.T.-R., Venakides, S., Zhou, X.: Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. Commun. Pure Appl. Math. **52**(11), 1335–1425 (1999). https://doi.org/10.1002/(SICI)1097-0312(199911)52:11<1335::AID-CPA1>3.0.CO;2-1
35. Erdős, L., Krüger, T., Schröder, D.: Random matrices with slow correlation decay. Forum Math. Sigma **7**, E8 (2019). https://doi.org/10.1017/fms.2019.2

36. Erdős, L., Krüger, T., Schröder, D.: Cusp universality for random matrices I: local law and the complex Hermitian case. Commun. Math. Phys. **378**, 1203–1278 (2018). https://doi.org/10.1007/s00220-019-03657-4
37. Erdős, L., Knowles, A., Yau, H.-T., Yin, J.: Spectral statistics of Erdős-Rényi graphs II: eigenvalue spacing and the extreme eigenvalues. Commun. Math. Phys. **314**, 587–640 (2012). https://doi.org/10.1007/s00220-012-1527-7
38. Erdős, L., Péché, S., Ramírez, J.A., Schlein, B., Yau, H.-T.: Bulk universality for Wigner matrices. Commun. Pure Appl. Math. **63**, 895–925 (2010). https://doi.org/10.1002/cpa.20317
39. Erdős, L., Riabov, V.: Eigenstate thermalization hypothesis for Wigner-type matrices. Commun. Math. Phys. **405**, 282 (2024). https://doi.org/10.1007/s00220-024-05143-y
40. Erdős, L., Schlein, B., Yau, H.-T.: Universality of random matrices and local relaxation flow. Invent. Math. **185**, 75–119 (2011). https://doi.org/10.1007/s00222-010-0302-7
41. Erdős, L., Yau, H.-T.: *A dynamical approach to random matrix theory*. Vol. 28. American Mathematical Society (2017). https://doi.org/10.1090/cln/028
42. Guionnet, A., Huang, J.: Rigidity and edge universality of discrete β-ensembles. Commun. Pure Appl. Math. **72**, 1875–1982 (2019). https://doi.org/10.1002/cpa.21818
43. He, Y., Knowles, A.: Mesoscopic eigenvalue statistics of Wigner matrices. Ann. Appl. Probab. **27**, 1510–1550 (2017). https://doi.org/10.1214/16-AAP1237
44. Huang, J., Landon, B.: Rigidity and a mesoscopic central limit theorem for Dyson Brownian motion for general β and potentials. Probab. Theory Relat. Fields **175**, 209–253 (2019). https://doi.org/10.1007/s00440-018-0889-y
45. Huang, J., Landon, B., Yau, H.-T.: Bulk universality of sparse random matrices. J. Math. Phys. (2015). https://doi.org/10.1063/1.4936139
46. Johansson, K.: Discrete orthogonal polynomial ensembles and the Plancherel measure. Ann. Math. **153**, 259–296 (2001). https://doi.org/10.2307/2661375
47. Knowles, A., Yin, J.: Anisotropic local laws for random matrices. Probab. Theory Relat. Fields **169**, 257–352 (2017). https://doi.org/10.1007/s00440-016-0730-4
48. Krishnapur, M., Rider, B., Virág, B.: Universality of the stochastic airy operator. Commun. Pure Appl. Math. **69**, 145–199 (2016). https://doi.org/10.1002/cpa.21573
49. Landon, B., Lopatto, P., Sosoe, P.: Single eigenvalue fluctuations of general Wigner-type matrices. Probab. Theory Relat. Fields **188**, 1–62 (2024). https://doi.org/10.1007/s00440-022-01181-6
50. Landon, B., Sosoe, P.: *Almost-optimal bulk regularity conditions in the CLT for Wigner matrices*. (2022). eprint: 2204.03419
51. Lee, J.O., Schnelli, K.: Edge universality for deformed Wigner matrices. Rev. Math. Phys. **27**, 1550018 (2015). https://doi.org/10.1142/S0129055X1550018X
52. Lee, J.O., Schnelli, K.: Local law and Tracy–Widom limit for sparse random matrices. Probab. Theory Relat. Fields **171**, 543–616 (2018). https://doi.org/10.1007/s00440-017-0787-8
53. Lee, J., Schnelli, K., Stetler, B., Yau, H.-T.: Bulk universality for deformed Wigner matrices. Ann. Probab. **44**, 2 (2016). https://doi.org/10.1214/15-AOP1023
54. Mehta, M.L.: Random Matrices and the Statistical Theory of Energy Levels. Academic Press (1967). https://doi.org/10.1016/C2013-0-12505-6
55. Pastur, L., Shcherbina, M.: Bulk universality and related properties of Hermitian matrix models. J. Stat. Phys. **130**, 205–250 (2008). https://doi.org/10.1007/s10955-007-9434-6
56. Pastur, L., Shcherbina, M.: On the edge universality of the local eigenvalue statistics of matrix models. Mat. Fiz. Anal. Geom. **10**, 335–365 (2003)
57. Shcherbina, M.: Edge universality for orthogonal ensembles of random matrices. J. Stat. Phys. **136**, 35–50 (2009). https://doi.org/10.1007/s10955-009-9766-5
58. Shcherbina, M.: Change of variables as a method to study general β-models: bulk universality. J. Math. Phys. **55**, 043504 (2014). https://doi.org/10.1063/1.4870603
59. Shorack, G.R., Wellner, J.A.: Empirical Processes with Applications to Statistics. Society for Industrial and Applied Mathematics (2009). https://doi.org/10.1137/1.9780898719017
60. Sodin, S.: The spectral edge of some random band matrices. Ann. Math. (2010). https://doi.org/10.4007/ANNALS.2010.172.2223
61. Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. Commun. Math. Phys. **207**, 697–733 (1999). https://doi.org/10.1007/s002200050743
62. Tao, T., Vu, V.: Random matrices: sharp concentration of eigenvalues. Rand. Mat.: Theor. Appl. **2**(03), 1350007 (2013). https://doi.org/10.1142/S201032631350007X
63. Tao, T., Vu, V.: Random matrices: universality of local eigenvalue statistics. Acta Math. **206**, 127–204 (2011). https://doi.org/10.1007/s11511-011-0061-3
64. Tao, T., Vu, V.: Random matrices: universality of local eigenvalue statistics up to the edge. Commun. Math. Phys. **298**, 549–572 (2010). https://doi.org/10.1007/s00220-010-1044-5

65. Tracy, C.A., Widom, H.: Level-spacing distributions and the Airy kernel. Commun. Math. Phys. **159**, 151–174 (1994). https://doi.org/10.1007/BF02100489

66. Tracy, C.A., Widom, H.: On orthogonal and symplectic matrix ensembles. Commun. Math. Phys. **177**, 727–754 (1996). https://doi.org/10.1007/BF02099545

67. Tracy, C.A., Widom, H.: The Pearcey process. Commun. Math. Phys. **263**, 381–400 (2006). https://doi.org/10.1007/s00220-005-1506-3

68. Valkó, B., Virág, B.: Continuum limits of random matrices and the Brownian carousel. Invent. Math. **177**, 463–508 (2009). https://doi.org/10.1007/s00222-009-0180-z

69. Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. **62**, 548–564 (1955). https://doi.org/10.2307/1970079

Communicated by R. Seiringer.