Journal of Functional Analysis 290 (2026) 111180

Contents lists available at ScienceDirect

Journal of Functional Analysis

journal homepage: www.elsevier.com/locate/jfa

Regular Article

Optimal decay of eigenvector overlap for non-Hermitian random matrices

Giorgio Cipolloni [a], László Erdős [b,*,1], Yuanyuan Xu [c,1,2]

[a] University of Arizona, United States of America
[b] IST Austria, Austria
[c] AMSS, CAS, China

A R T I C L E I N F O

Article history:
Received 7 December 2024
Accepted 13 August 2025
Available online 1 September 2025
Communicated by Ivan Corwin

MSC:
60B20
15B52

Keywords:
Zigzag strategy
Multi-resolvent local law
Characteristic flow
Dyson Brownian motion

A B S T R A C T

We consider the standard overlap $\mathcal{O}_{ij} := \langle \boldsymbol{r}_j, \boldsymbol{r}_i \rangle \langle \boldsymbol{l}_j, \boldsymbol{l}_i \rangle$ of any bi-orthogonal family of left and right eigenvectors of a large random matrix X with centred i.i.d. entries and we prove that it decays as an inverse second power of the distance between the corresponding eigenvalues. This extends similar results for the complex Gaussian ensemble from Bourgade and Dubach [15], as well as Benaych-Georges and Zeitouni [13], to any i.i.d. matrix ensemble in both symmetry classes. As a main tool, we prove a two-resolvent local law for the Hermitisation of X uniformly in the spectrum with optimal decay rate and optimal dependence on the density near the spectral edge.

* Corresponding author.
E-mail addresses: gcipolloni@arizona.edu (G. Cipolloni), lerdos@ist.ac.at (L. Erdős), yyxu2023@amss.ac.cn (Y. Xu).

1 Partially supported by ERC Advanced Grant “RMTBeyond” No. 101020331.
2 Partially supported by National Key R&D Program of China No. 2024YFA1013503.

https://doi.org/10.1016/j.jfa.2025.111180

1. Introduction

We consider large $n \times n$ non–Hermitian matrices X with independent, identically distributed (i.i.d.) centred entries, with the standard normalization $\boldsymbol{E}|X_{ij}|^2 = \frac{1}{n}$ (*i.i.d. matrices*). While the eigenvalue statistics of fairly general large Hermitian random matrices have been very well understood in the last fifteen years, those of i.i.d. matrices have only been understood very recently both in the bulk [42,61,65] and at the edge [24] of the spectrum; see also [2,20,33,59,60,75] for recent results for *deformed* models of the form $A + X$, where A is a deterministic matrix. One of the main difficulties in understanding the eigenvalue statistics in the non–Hermitian setting is the high instability of the spectrum, which is partly manifested by the behaviour of their eigenvectors. While in the Hermitian case, e.g. for GOE/GUE[3] matrices for simplicity, the eigenvectors are distributed according to the Haar measure, for Ginibre[4] matrices the left/right eigenvectors are strongly correlated and difficult to grasp even for the Gaussian case (the general i.i.d. case is expected to be even much harder). Recall that in general, a non–Hermitian matrix X has a two sets of eigenvectors $\boldsymbol{l}_i, \boldsymbol{r}_i$, which, assuming that each eigenvalue σ_i of X is simple, are defined by

$$X\boldsymbol{r}_i = \sigma_i \boldsymbol{r}_i, \qquad \boldsymbol{l}_i^t X = \sigma_i \boldsymbol{l}_i^t, \qquad i = 1, 2, \dots, n. \tag{1.1}$$

We normalize the left/right eigenvectors to form a bi-orthogonal basis, i.e. $\langle \bar{\boldsymbol{l}}_j, \boldsymbol{r}_i \rangle = \boldsymbol{l}_j^t \boldsymbol{r}_i = \delta_{ij}$. Instead of studying the eigenvectors directly, one usually considers the scale-invariant quantity

$$\mathcal{O}_{ij} := \langle \boldsymbol{r}_j, \boldsymbol{r}_i \rangle \langle \boldsymbol{l}_j, \boldsymbol{l}_i \rangle, \tag{1.2}$$

called *eigenvector overlaps*. For random matrices, overlaps emerge in many problems both in mathematics [13,15,46,37,62,72] and physics [5,6,12,21–23,52,48,51,47,49,58,68]. Two of their most prominent motivations are:

(1) $\mathcal{O}_{ij}$ determine the correlation of the evolution of the eigenvalues σ_i under a *non–Hermitian* analog of the *Dyson Brownian motion*, which was rigorously derived in [15, Appendix A] (see also [19,54]).
(2) For $i = j$, the quantity $\sqrt{\mathcal{O}_{ii}}$ is known as the *eigenvalue condition number* in numerical analysis, as it determines how sensitive σ_i is to small perturbation (especially relevant in the context of *smoothed analysis* [69] via random matrices, see e.g. [10,27]).

[3] We say that an $n \times n$ random matrix W is a GUE (resp. GOE) matrix if it is Hermitian $W = W^*$, the entries above the diagonal W_{ij} are such that $\sqrt{n}W_{ij}$ are independent standard complex (real) Gaussian random variables, and the diagonal entries are real Gaussian random variables of variance $1/n$ (resp. $2/n$).
[4] We say that an $n \times n$ random matrix is a complex (resp. real) Ginibre matrix if its entries are centred complex (resp. real) Gaussian random variables with variance $1/n$.

Despite the central role of the overlaps, rigorous results about their size and distribution have appeared only recently. Inspired by the seminal work of Chalker and Mehlig [21] computing the expectations of the $\mathcal{O}_{ij}$ for the complex Ginibre ensemble, Bourgade and Dubach [15] also computed the second moment of the off-diagonal overlap and diagonal overlap,

$$\boldsymbol{E}\big[|\mathcal{O}_{ij}|^2\big|\sigma_i = z_1,\ \sigma_j = z_2\big] \approx \frac{(1-|z_1|^2)(1-|z_2|^2)}{|z_1-z_2|^4}, \tag{1.3}$$

$$\boldsymbol{E}\big[\mathcal{O}_{ii}\mathcal{O}_{jj}\big|\sigma_i = z_1,\ \sigma_j = z_2\big] \approx n^2(1-|z_1|^2)(1-|z_2|^2), \tag{1.4}$$

identifying a power law (quadratic) decay of $\mathcal{O}_{ij}$ for far away eigenvalues. Beyond Ginibre ensembles, Benaych-Georges and Zeitouni [13] showed that the closely related quantity $\sqrt{n}(\sigma_i - \sigma_j)\langle \boldsymbol{r}_i, \boldsymbol{r}_j\rangle/\|\boldsymbol{r}_i\|\|\boldsymbol{r}_j\|$ is sub-Gaussian (uniformly in n) for more general non-normal matrices defined by an invariant distribution, and explicitly compute its distribution in the complex Ginibre case. Both results identify a quadratic decay, however their proofs rely on the invariance (or even Gaussianity) of the underlying matrix ensemble.

Is this quadratic decay ubiquitous, in particular does it extend to general real and complex i.i.d. matrix ensembles without invariance properties? In this paper we answer affirmatively by proving that the optimal bound (modulo the n^ξ factor) as indicated by (1.3)-(1.4)

$$\frac{|\mathcal{O}_{ij}|}{\sqrt{\mathcal{O}_{ii}\mathcal{O}_{jj}}} \le \frac{n^\xi}{n|\sigma_i - \sigma_j|^2 + 1} \tag{1.5}$$

holds uniformly in the spectrum with very high probability for any small $\xi > 0$. Previously, only two non-optimal bounds were available. In [30, Theorem 3.1] it was shown that $|\mathcal{O}_{ij}|/\sqrt{\mathcal{O}_{ii}\mathcal{O}_{jj}} \le n^{-\epsilon}$ for bulk eigenvectors, if $|\sigma_i - \sigma_j| \ge n^{-1/2+\epsilon'}$ which is off by almost an entire factor n for far away eigenvalues. Close to the edge of the spectrum, $|\sigma_i|, |\sigma_j| \approx 1$, half of the optimal decay, $1/(\sqrt{n}|\sigma_i - \sigma_j|)$, was obtained in [35, Corollary 3.4]. We remark that the quadratic decay bound obtained in (1.5) is better than $1/(\sqrt{n}|\sigma_i - \sigma_j|)$ if $|\sigma_i - \sigma_j| \ge n^{-1/2+\xi}$. Neither proof can be directly improved and the optimal bound (1.5) requires to resolve two new challenges.

We now explain the novelties in our proof of (1.5). The main inputs are the so-called *multi–resolvent local laws* for products of resolvents of the Hermitisation of $X - z$. Since the non-Hermitian eigenvectors $\boldsymbol{l}_i, \boldsymbol{r}_i$ are not accessible directly, we relate them to the eigenvectors $\boldsymbol{w}_i^z = ((\boldsymbol{u}_i^z)^*, (\boldsymbol{v}_i^z)^*)^* \in \mathbb{C}^{2n}$ of the $(2n)\times(2n)$ Hermitized matrix

$$H^z := \begin{pmatrix} 0 & X - z \\ (X-z)^* & 0 \end{pmatrix}. \tag{1.6}$$

Note that $\boldsymbol{u}_i^z, \boldsymbol{v}_i^z \in \mathbb{C}^n$ are also the left/right singular vectors corresponding to the singular value λ_i^z of a family of matrices $X - z$ parametrized by $z \in \mathbb{C}$. We assume that the singular values λ_i^z are labelled in an increasing order.

Eigenvectors of X and singular vectors of $X - z$ are not related in general, except in the special case when z is an eigenvalue. Simple algebra shows that

$$\lambda_1^z = 0 \iff 0 \in \mathrm{Spec}(H^z) \iff 0 \in \mathrm{Spec}(X - z) \iff z \in \mathrm{Spec}(X), \tag{1.7}$$

and in this case the eigenvectors of X with eigenvalue z and the singular vectors of $X - z$ with singular value zero coincide. This fundamental relation will be our starting point to study the eigenvalues and the eigenvectors of the non–Hermitian X via the study of the spectrum of the Hermitized matrix H^z. Our goal is thus to estimate quantities of the form $|\langle \boldsymbol{u}_1^{z_1}, \boldsymbol{u}_1^{z_2}\rangle|, |\langle \boldsymbol{v}_1^{z_1}, \boldsymbol{v}_1^{z_2}\rangle|$ with high probability simultaneously for every $|z_1|, |z_2| \leq 1$. In fact, if we have the desired bound on these quantities, by choosing $z_1 = \sigma_i$, $z_2 = \sigma_j$, we can readily conclude (1.5). By spectral decomposition, one can easily see that for the desired bound on $|\langle \boldsymbol{u}_1^{z_1}, \boldsymbol{u}_1^{z_2}\rangle|, |\langle \boldsymbol{v}_1^{z_1}, \boldsymbol{v}_1^{z_2}\rangle|$, it is actually enough to estimate

$$\langle \mathrm{Im}\, G^{z_1}(\mathrm{i}\eta_1)\mathrm{Im}\, G^{z_2}(\mathrm{i}\eta_2)\rangle \lesssim \frac{\rho_1\rho_2}{|z_1 - z_2|^2}, \qquad G^z(\mathrm{i}\eta) := (H^z - \mathrm{i}\eta)^{-1}, \tag{1.8}$$

for appropriate choices of η_1, η_2 (eventually η_i is chosen a bit above the local fluctuation scale of the eigenvalues of H^{z_i} close to zero). Here ρ_i denotes the local density of the eigenvalues of H^{z_i}.

Although we only need the upper bound (1.8) for (1.5), along the way in Theorem 3.5 we will prove a stronger result, called *two-resolvent local law*, that even identifies the leading term in the relevant regime of η_i's:

$$\Big|\langle \mathrm{Im}\, G^{z_1}(\mathrm{i}\eta_1)\mathrm{Im}\, G^{z_2}(\mathrm{i}\eta_2) - M_{12}\rangle\Big| = o\Big(\frac{\rho_1\rho_2}{|z_1 - z_2|^2}\Big), \tag{1.9}$$

where M_{12} is a deterministic matrix computed from the underlying *Dyson equation*.

To prove (1.9), we rely on the *zigzag strategy* [30,28], which involves three main steps: i) controlling $\langle \mathrm{Im}\, G^{z_1}(\mathrm{i}\eta_1)\mathrm{Im}\, G^{z_2}(\mathrm{i}\eta_2) - M_{12}\rangle$ for large values of $\eta_1, \eta_2 \sim 1$, ii) [*zig-step*] using the *method of characteristics* (introduced in this form in [1,14,56] for single resolvents, see [31, Section 1.4] for more history) to show that this bound can be propagated down to optimally small spectral parameters η_i at the price of adding a Gaussian component to the matrix X, iii) [*zag-step*] using a version of the Green's function comparison theorem (GFT) to remove the Gaussian component. While on a high level this approach has been used quite extensively, in the current paper we face a novel challenge: we need to extract an additional small factor of order ρ coming from $\mathrm{Im}\, G^z$ (compared to G^z) close to the edge of the spectrum in the context of two different spectral parameters z_1, z_2. This requires to study the evolution of $\langle \mathrm{Im}\, G_1 \mathrm{Im}\, G_2\rangle$, with $G_i := G^{z_i}(\mathrm{i}\eta_i)$, along the characteristics flow, which is more delicate than the one for $\langle G_1 G_2\rangle$ investigated in [30,28] and also requires a more delicate analysis at the GFT level. A bound on this quantity was estimated in [35, Part B of Theorem 3.3] close to the edge of the spectrum with a non-optimal decay $1/|z_1 - z_2|$. This non-optimality is a consequence of bounding $\langle G_1 G_2\rangle$ and then writing $2\mathrm{i}\mathrm{Im}\, G = G - G^*$, rather than analyzing $\langle \mathrm{Im}\, G_1 \mathrm{Im}\, G_2\rangle$ directly and extract a cancellation between G and G^* near the edge, as we do instead in the current paper.

The second challenge we resolve in this paper is to estimate optimally the deterministic M-terms, such as M_{12} in (1.9), even for longer resolvent chains that naturally arise along the zig-flow (we need three-resolvent chains). In fact, the precision of the zigzag method ultimately depends only on such estimates on M-terms. While an explicit recursion is available for expressing the M-terms, the formula is inherently unstable in the small η regime as it contains delicate cancellations. In [35] these cancellations were unravelled but only outside of the spectrum and without the extra ρ-factors. Now we find the optimal estimate uniformly in the spectrum and with the correct ρ factors.

We remark that the analogue of the quadratic decay in $z_1 - z_2$ in (1.8), and hence in (1.5), appears in a related Hermitian model as well. In [31] the authors study a Wigner matrix W with two different (deterministic) deformations, $W + D_1$ and $W + D_2$. They show that the product of their resolvents $\langle G_1 G_2 \rangle$, hence their eigenvector overlap, exhibits essentially a quadratic decay $1/\langle (D_1 - D_2)^2 \rangle$, although the precise statement is modulated by the difference in the spectral parameters as well. However, the results therein are optimal only within the bulk of the spectrum, again because the smallness coming from $\operatorname{Im} G^z$ was not exploited.

The central role that (1.8) plays in studying overlaps underlines the importance of multi-resolvent local laws in the theory of random matrices. For many years, single resolvent local laws have been the technical workhorses behind the proofs of Wigner-Dyson-Mehta and related universality results on local eigenvalue statistics since they provided almost optimal a priori bounds on individual eigenvalues and eigenvectors. These are the standard *rigidity* and *delocalisation* results used regularly in arguments based on the Dyson Brownian motion (DBM) and the Green function comparison theorems (GFT). A lot of relevant information on random matrices, however, are not accessible via a single resolvent. Key physical concepts such as eigenvector overlaps, thermalisation phenomena, out-of-time-order correlations etc. are inherently connected with multi-resolvent chains of the form $G_1 A_1 G_2 A_2 \ldots$, where G_i are random resolvents and A_i's are deterministic matrices. Besides these obvious motivations, the recent results on bulk universality of eigenvalues and eigenvectors for i.i.d. matrices, initiated by the breakthrough papers [61,65] using the partial Schur decomposition, also heavily rely on multi-resolvent local laws as a priori technical inputs. The Schur decomposition method is powerful to prove the universality of the distribution of the diagonal overlaps $\mathcal{O}_{ii}$ [67], but it cannot yet access to off-diagonal overlaps. Our result gives (1.5) an optimal a priori bound on $\mathcal{O}_{ij}$ and we also expect that multi-resolvent local laws will play an important role in resolving the open conjecture on the universality of $\mathcal{O}_{ij}$ for i.i.d. matrices.

We close this introduction with a quick overview of the recent vast literature about overlaps (1.2) for various random matrix ensembles. With the exception of [15], all these works concern only diagonal overlaps $\mathcal{O}_{ii}$. Long after the classical paper [21], the two main works that revived the interest in overlaps of non–Hermitian eigenvectors in the context of random matrix theory are [15,46] on Ginibre ensemble. Besides the results on the off-diagonal overlaps in [15] that we already described, the distribution of $\mathcal{O}_{ii}$ and the correlation of $\mathcal{O}_{ii}\mathcal{O}_{jj}$ for bulk eigenvalues in the complex case have also been computed in that paper. In [46] the author computes the distribution of $\mathcal{O}_{ii}$ both in the bulk and at the edge of the spectrum for complex matrices, as well as for overlaps corresponding to real eigenvalues of real Ginibre matrices. Expectation of overlaps for complex eigenvalues was covered in [73]. Beyond Gaussian ensembles, in the context of general i.i.d. matrices, the precise (in terms of n–dependence) size of $\mathcal{O}_{ii}$ was identified in [32,44] (see also for previous non-optimal results in terms of the n–dependence [10,57,71]), and more recently even the universality of its distribution has been established [67]. Two very recent results [43,66] also prove the universality of the Gaussian fluctuations directly for the eigenvectors' entries of i.i.d. matrices. Eigenvector overlaps have also been studied for several other models even beyond i.i.d. matrices, including deformed Ginibre matrices [11,69,76], symplectic Ginibre matrices [5,7], non–Hermitian Brownian motion [18,45], the elliptic ensemble [39,38,50,70], spherical and truncated unitary ensembles [40,41,64], and non-normal dynamical systems [63].

Acknowledgment. We thank Yan Fyodorov for pointing out the reference [58] and for useful comments. We also thank the anonymous referee for their several comments that substantially improved the presentation of this manuscript.

Conventions and notations. For integers $k,l \in \mathbb{N}$, with $k < l$, we use the notations $[k,l] := \{k,\dots,l\}$ and $[k] := [1,k]$. For positive quantities f,g we write $f \lesssim g$ and $f \sim g$ if $f \le Cg$ or $cg \le f \le Cg$, respectively, for some constants $c,C > 0$ which depend only on the constants appearing in (2.1). We set $f \wedge g = \min\{f,g\}$ and $f \vee g = \max\{f,g\}$. Furthermore, for n–dependent positive quantities f_n, g_n we use the notation $f_n \ll g_n$ to denote that $\lim_{n\to\infty}(f_n/g_n) = 0$. Throughout the paper $c,C > 0$ denote small and large constants, respectively, which may change from line to line. We denote vectors by bold-faced lower case Roman letters $\boldsymbol{x},\boldsymbol{y} \in \mathbb{C}^d$, for some $d \in \mathbb{N}$. Vector and matrix norms, $\|\boldsymbol{x}\|$ and $\|A\|$, indicate the usual Euclidean norm and the corresponding induced matrix norm. For any $d\times d$ matrix A we use $\langle A\rangle := d^{-1}\mathrm{Tr}A$ to denote the normalized trace of A. Moreover, for vectors $\boldsymbol{x},\boldsymbol{y} \in \mathbb{C}^d$ we define the usual scalar product $\langle \boldsymbol{x},\boldsymbol{y}\rangle := \sum_{i=1}^d \overline{x_i} y_i$, and use $(A)_{\boldsymbol{xy}}$ to denote the inner product $\langle \boldsymbol{x}, A\boldsymbol{y}\rangle$.

We define the following special $2n\times 2n$ matrices

$$E_1 := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad E_\pm := E_1 \pm E_2 = \begin{pmatrix} 1 & 0 \\ 0 & \pm 1 \end{pmatrix}, \qquad F := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{1.10}$$

in particular, E_+ is the $2n$–dimensional identity matrix.

We will often use the concept of "with very high probability" for an n-dependent event, meaning that for any fixed $D > 0$ the probability of the event is bigger than $1 - n^{-D}$ if $n \ge n_0(D)$. Moreover, we use the convention that $\xi > 0$ denotes an arbitrary small positive exponent which is independent of n. We recall the standard notion of *stochastic domination*: given two families of non-negative random variables

$$X = \left(X^{(n)}(u) \,:\, n \in \mathbb{N}, u \in U^{(n)}\right) \quad \text{and} \quad Y = \left(Y^{(n)}(u) \,:\, n \in \mathbb{N}, u \in U^{(n)}\right)$$

indexed by n (and possibly some parameter u in some parameter space $U^{(n)}$), we say that X is *stochastically dominated* by Y, if for any $\xi, D > 0$ we have

$$\sup_{u \in U^{(n)}} \boldsymbol{P}\left[X^{(n)}(u) > n^{\xi} Y^{(n)}(u)\right] \leq n^{-D} \tag{1.11}$$

for large enough $n \geq n_0(\xi, D)$. In this case we use the notation $X \prec Y$ or $X = O_{\prec}(Y)$ and we say that $X(u) \prec Y(u)$ *holds uniformly* in $u \in U^{(n)}$.

2. Main results

We consider *i.i.d. matrices* $X \in \mathbb{C}^{n \times n}$ (complex case) or $X \in \mathbb{R}^{n \times n}$ (real case), i.e. $n \times n$ matrices with independent and identically distributed (i.i.d.) entries $x_{ab} \stackrel{\mathrm{d}}{=} n^{-1/2}\chi$, where the distribution of χ satisfies the following assumptions:

Assumption 2.1. We assume that $\boldsymbol{E}\chi = 0$, $\boldsymbol{E}|\chi|^2 = 1$, additionally $\boldsymbol{E}\chi^2 = 0$ in the complex case. Furthermore, we require that for any $p \in \mathbb{N}$, there exists a constant $C_p > 0$ such that

$$\boldsymbol{E}|\chi|^p \leq C_p. \tag{2.1}$$

Temporarily assuming that the spectrum $\{\sigma_i\}_{i=1}^n$ of X is simple, X can be diagonalized with left eigenvectors $\{\boldsymbol{l}_i\}_{i=1}^n$ and right eigenvectors $\{\boldsymbol{r}_i\}_{i=1}^n$. For example, if χ is a continuous random variable, then the spectrum is simple with probability one. We normalize the left and right eigenvectors so that they form a bi-orthogonal basis $\langle \overline{\boldsymbol{l}}_i, \boldsymbol{r}_j \rangle = \delta_{ij}$, $\forall i, j = 1, \ldots, n$. Due to the non-Hermiticity of X, $\{\boldsymbol{l}_i\}_{i=1}^n$ and $\{\boldsymbol{r}_i\}_{i=1}^n$ typically differ and they are not orthogonal. To quantify their relation we consider the scale invariant quantity, called *overlap*,

$$\mathcal{O}_{ij} = \langle \boldsymbol{l}_i, \boldsymbol{l}_j \rangle \langle \boldsymbol{r}_i, \boldsymbol{r}_j \rangle. \tag{2.2}$$

While diagonal overlaps, i.e. $\mathcal{O}_{ii} = \|\boldsymbol{l}_i\|^2 \|\boldsymbol{r}_i\|^2$, have been understood fairly well by now [15,32,46,44,67], much less is known for general $\mathcal{O}_{ij}$. Previous to the current work, the only results for $\mathcal{O}_{ij}$, with $i \neq j$, were computation of their expectation [21] and variance [15] (see also [13] for the distributional limit of a related quantity) in the special case of the complex Ginibre ensemble, i.e. when χ is a standard complex Gaussian random variable in Assumption 2.1. Our main result is a high probability upper bound for the normalized version of $\mathcal{O}_{ij}$ for general i.i.d. matrices, which holds uniformly in the spectrum:

Theorem 2.2. *Assume that X is a real or complex i.i.d. matrix satisfying Assumption 2.1 and that, for simplicity, its spectrum is simple (see Remark 2.3). Let $\mathcal{O}_{ij}$ be the overlaps defined in (2.2). Then, with very high probability, for any small $\xi > 0$, we have*

$$\sup_{i,j\in[n]} \left(n|\sigma_i - \sigma_j|^2 + 1\right) \left[\frac{\left|\langle \boldsymbol{r}_i, \boldsymbol{r}_j\rangle\right|^2}{\|\boldsymbol{r}_i\|^2\|\boldsymbol{r}_j\|^2} + \frac{\left|\langle \boldsymbol{l}_i, \boldsymbol{l}_j\rangle\right|^2}{\|\boldsymbol{l}_i\|^2\|\boldsymbol{l}_j\|^2} \right] \le n^{\xi}. \tag{2.3}$$

In particular, by a Cauchy-Schwarz inequality, this implies

$$\sup_{i,j\in[n]} \left(n|\sigma_i - \sigma_j|^2 + 1\right) \frac{|\mathcal{O}_{ij}|}{\sqrt{\mathcal{O}_{ii}\mathcal{O}_{jj}}} \le n^{\xi}. \tag{2.4}$$

Remark 2.3. For simplicity of notation we stated Theorem 2.2 for matrices X with simple spectrum. However, we do not need this assumption in the proof (see the end of Section 3.3). In fact, from the same proof it follows that for any left/right eigenvectors $\boldsymbol{l}, \boldsymbol{r}$ and $\boldsymbol{l}', \boldsymbol{r}'$ with corresponding eigenvalues σ, σ', respectively, for any small $\xi > 0$, we have

$$\left[\frac{\left|\langle \boldsymbol{r}, \boldsymbol{r}'\rangle\right|^2}{\|\boldsymbol{r}\|^2\|\boldsymbol{r}'\|^2} + \frac{\left|\langle \boldsymbol{l}, \boldsymbol{l}'\rangle\right|^2}{\|\boldsymbol{l}\|^2\|\boldsymbol{l}'\|^2} \right] \le \frac{n^{\xi}}{n|\sigma - \sigma'|^2 + 1}. \tag{2.5}$$

This bound holds with very high probability simultaneously for all eigenvectors of X.

The bound (2.4) is optimal (up to the n^{ξ} error), as it can be seen from the variance computation in [15] in the complex Gaussian case. The optimality of the stronger bound (2.3) follows from [13], where the authors proved that $\sqrt{n}(\sigma_i - \sigma_j)\langle \boldsymbol{r}_i, \boldsymbol{r}_j\rangle / \|\boldsymbol{r}_i\|\|\boldsymbol{r}_j\|$ is uniformly (in n) sub-Gaussian for a large class of invariant non-normal matrix ensembles and even compute its distributional limit in the complex Ginibre case. While our result gives only an upper bound, it holds for more general i.i.d. matrices satisfying Assumption 2.1. In fact, (2.3) is the analog of the upper bound in the display below [13, Eq. (5)]. While these results from [13,15] are similar to ours, the proof methods are completely different.

Remark 2.4. In Theorem 2.2 we proved an (almost) optimal bound for the overlaps of matrices X with i.i.d. entries as in Assumption 2.1. However, inspecting our proof it is clear that this argument can easily be extended to more general models, i.e. to deformed models $X + A$, where X is an i.i.d. matrix and A is a deterministic deformation belonging to a large class of matrices (see e.g. [32, Section 2.1], [20, Assumption 2.8] for bulk and edge eigenvectors, respectively). More precisely, the arguments of Sections 4–5 work verbatim in this more general case as well, once the deterministic bounds in Lemma 3.3 are established.

In Theorem 2.2 we show that the *normalized overlap* $\mathcal{O}_{ij}/\sqrt{\mathcal{O}_{ii}\mathcal{O}_{jj}}$ can be estimated in high probability sense, in particular its tail probability is very small. The same is not true for $\mathcal{O}_{ij}$, because overlaps themselves are expected to have a heavy tail, similarly to what it has been shown for the diagonal overlaps $\mathcal{O}_{ii}$ [15,46,67]. In the following corollary we present a bound on $\mathcal{O}_{ij}$ which holds only in a first moment sense using the analogous bounds on the diagonal overlaps $\mathcal{O}_{ii}$ given in [44, Theorems 2.7, 2.9] to account for the normalisation.

Corollary 2.5. *Consider an i.i.d. matrix X satisfying Assumption 2.1. Furthermore, we assume that χ has a bounded density g and in the complex case $\operatorname{Re}\chi, \operatorname{Im}\chi$ are independent. Fix any small constants $\tau, \epsilon, \xi > 0$ with $\xi \le \epsilon/10$ and any large constant $D > 0$. Set $\mathcal{D}_z^\xi \subset \mathbb{C}$ to be the disk centred at $z \in \mathbb{C}$ with radius $n^{-1/2+\xi}$. Then for any $z_i \in \mathbb{C}$ $(i = 1, 2)$ with $|z_i| \le 1 + n^{-1/2+\tau}$ and $|z_1 - z_2| \ge n^{-1/2+\epsilon}$, there is an event Ξ with very high probability, i.e. $\boldsymbol{P}(\Xi) \ge 1 - n^{-D}$, such that*

$$\boldsymbol{E}\Big(\sum_{\sigma_i \in \mathcal{D}^\xi_{z_1}, \sigma_j \in \mathcal{D}^\xi_{z_2}} |\mathcal{O}_{ij}| \cdot \mathbf{1}_\Xi\Big) \lesssim \frac{n^{8\xi}}{|z_1 - z_2|^2}, \qquad \textit{[Complex case]} \tag{2.6}$$

$$\boldsymbol{E}\Big(\sum_{\sigma_i \in \mathcal{D}^\xi_{z_1}, \sigma_j \in \mathcal{D}^\xi_{z_2}} |\mathcal{O}_{ij}| \cdot \mathbf{1}_\Xi\Big) \lesssim \frac{n^{8\xi}}{|z_1 - z_2|^2} \cdot \frac{1}{|\operatorname{Im} z_1| \wedge |\operatorname{Im} z_2|}, \qquad \textit{[Real case]} \tag{2.7}$$

Furthermore, in the complex case in the bulk regime, i.e. if $|z_1|, |z_2| \le 1 - \tau$, the event Ξ is trivial, i.e. (2.6) holds without $\mathbf{1}_\Xi$. In the real case we have a special bound for the real eigenvalues σ_i, σ_j near z_1, z_2:

$$\boldsymbol{E}\Big(\sum_{\substack{\sigma_i \in \mathcal{D}^\xi_{z_1} \cap \mathbb{R} \\ \sigma_j \in \mathcal{D}^\xi_{z_2} \cap \mathbb{R}}} |\mathcal{O}_{ij}|^{1/2}\Big) \lesssim \frac{n^{4\xi}}{|z_1 - z_2|}. \tag{2.8}$$

Remark 2.6. From the local circular laws (see [16, Theorem 2.2], [17, Theorem 1.3], [74, Theorem 1.2], and [9, Theorem 2.3]), we know that with very high probability there are eigenvalues close to z_1 and z_2. More precisely, fix any small $\tau, \xi > 0$ with $\tau < \xi/10$, then for any $|z_i| \le 1 + n^{-1/2+\tau}$, there exist eigenvalues $\sigma_i \in \mathcal{D}^\xi_{z_1}$ and $\sigma_j \in \mathcal{D}^\xi_{z_2}$ with very high probability. Hence, the summations in (2.6)–(2.7) are not empty, in fact they typically contain roughly $n^{2\xi} \times n^{2\xi}$ eigenvalue pairs. Our result holds for the corresponding overlap for any of these pairs.

We conclude this section explaining another interesting result one can obtain using the two–resolvent local law in (3.26) and using very similar techniques to the ones used to prove (3.26) in Section 4. We will not present the complete proof for brevity, but we indicate the key points and what steps need more work.

Mesoscopic Central Limit Theorem uniformly in the spectrum: Fix a reference point $|z_0| < 1$, an exponent $a \in [0, 1/2)$ and a smooth and compactly supported test function

$f : \mathbb{C} \to \mathbb{C}$. Then the (centred) linear statistics of the eigenvalues σ_i around z_0 at scale n^{-a},

$$L_n(f) := \sum_{i=1}^n f_{z_0,a}(\sigma_i) - \boldsymbol{E}\sum_{i=1}^n f_{z_0,a}(\sigma_i), \qquad f_{z_0,a}(z) := f\big(n^a(z-z_0)\big),$$

satisfy a Central Limit Theorem (CLT) [29,30,25]. In particular, while in the macroscopic case [29,25], i.e. $a = 0$, $f_{z_0,a}$ was allowed to be supported anywhere in the complex plane, in the mesoscopic case [30], i.e. $a \in (0, 1/2)$, $f_{z_0,a}$ had to be supported in the bulk of the spectrum. Thus the mesoscopic CLT close to the edge (when $a > 0$ and $|z_0| = 1$) was not covered in [29,30,25]. Inspecting the proofs in [29,30,25], it is clear that their main inputs are averaged local laws for products of two resolvents (see [30, Theorem 3.3]) as in (3.26) and singular vector overlaps bound (see [30, Theorem 3.1]) as in (3.28), below. We claim that using this bound as an input, after proving a Wick theorem exactly as in [29, Section 6] and replacing the Dyson Brownian motion analysis [29, Section 7] with the one presented in [34, Section 7], one can routinely prove a mesoscopic CLT for test function supported also close to the edge. More precisely, for reference points $|z_0| = 1$ one would get that $L_n(f)$, for $a \in (0, 1/2)$, converges to a centred Gaussian random variable $L(f)$ with the following variance (for simplicity we consider only the complex case)

$$\boldsymbol{E}\big|L(f)\big|^2 = \frac{1}{4\pi}\int_{\mathbb{D}} \big|\nabla f_{z_0,a}(z)\big|^2 \,\mathrm{d}^2 z + \frac{1}{2}\sum_{k\in\mathbb{Z}} |k|\big|\widehat{f_{z_0,a}}(k)\big|^2, \tag{2.9}$$

where $\widehat{f_{z_0,a}}$ denotes the Fourier transform of the restriction of $f_{z_0,a}$ to the unit circle, and $\mathbb{D}$ the open unit disk. In particular, one can see that compared to [30, Theorem 2.1] one gets an additional contribution from eigenvalues close to the edge, as it was already observed in the macroscopic case in [29, Theorem 2.2]. Similarly, one can extend the time–dependent version of the CLT [18, Theorem 2.3] down to the optimal mesoscopic scale at the edge.

3. Proof strategy: multi-resolvent local laws

In this section we recall some preliminary inputs and present the main new technical results to obtain the overlap bound in Theorem 2.2. More precisely, in Section 3.1 we recall the single resolvent local law and the rigidity estimate for the eigenvalues. Then, in Section 3.2 we present local laws for products of two resolvents. Finally, in Section 3.3 we show that the in the regime when the limiting density of states is small we can improve the averaged local law from Section 3.2, and use this result to prove Theorem 2.2 and Corollary 2.5.

3.1. Local laws for the resolvent

Note that $z \in \mathbb{C}$ lies in the spectrum of X if only and if the least singular value of $X - z$ is zero. Hence we link the eigenvalues (resp. eigenvectors) of X with the singular values (resp. singular vectors) of matrix $X - z$. As already mentioned in the Introduction, let $(\lambda_i^z)_{i=1}^n$ denote the singular values of $X - z$ labelled in a non-decreasing order, and $\{\boldsymbol{u}_i^z\}_{i=1}^n, \{\boldsymbol{v}_i^z\}_{i=1}^n$ denote the corresponding left and right singular vectors of $X - z$ normalised such that $\|\boldsymbol{u}_i^z\|^2 = \|\boldsymbol{v}_i^z\|^2 = 1/2$. In the same spirit of [53], we consider the Hermitisation of X, *i.e.*,

$$H^z := \begin{pmatrix} 0 & X - z \\ X^* - \overline{z} & 0 \end{pmatrix} \in \mathbb{C}^{2n \times 2n}, \qquad z \in \mathbb{C}. \tag{3.1}$$

Simple algebra (also called *chiral symmetry*) shows that H^z has a symmetric spectrum given by $(\lambda_{\pm i}^z)_{i=1}^n = (\pm\lambda_i^z)_{i=1}^n$, with the corresponding (orthonormal) eigenvectors given by $\boldsymbol{w}_{\pm i}^z = ((\boldsymbol{u}_i^z)^*, \pm(\boldsymbol{v}_i^z)^*)^* \in \mathbb{C}^{2n}$. Note that $\boldsymbol{u}_i^z, \boldsymbol{v}_i^z$ are the left and right singular vectors of $X - z$. Eigenvalues of H^z in general have no direct relation with the eigenvalues of X, however, from the trivial relation (1.7) it is clear that eigenvectors of X with eigenvalue z and the singular vectors of H^z with singular value zero coincide.

Let G^z be the resolvent of the Hermitisation of $X - z$, *i.e.*,

$$G^z(w) := (H^z - w)^{-1}, \qquad H^z = \begin{pmatrix} 0 & X - z \\ X^* - \overline{z} & 0 \end{pmatrix}, \qquad w \in \mathbb{C} \setminus \mathbb{R}, \quad z \in \mathbb{C}. \tag{3.2}$$

Define the deterministic approximation of $G^z(w)$ as an $(2n) \times (2n)$ block constant matrix given by

$$M^z(w) := \begin{pmatrix} m^z(w)I_n & -zu^z(w)I_n \\ -\overline{z}u^z(w)I_n & m^z(w)I_n \end{pmatrix}, \quad \text{with} \qquad u^z(w) := \frac{m^z(w)}{w + m^z(w)}, \tag{3.3}$$

where $m^z(w)$ is the unique solution of the scalar cubic equation with side condition

$$-\frac{1}{m^z(w)} = w + m^z(w) - \frac{|z|^2}{w + m^z(w)}, \qquad \mathrm{Im}[m^z(w)]\mathrm{Im} w > 0. \tag{3.4}$$

In fact, $M^z(w)$ is the unique solution of the *Matrix Dyson equation* (see e.g. [4])

$$-[M^z(w)]^{-1} = w + Z + \mathcal{S}[M^z(w)], \qquad Z := \begin{pmatrix} 0 & z \\ \overline{z} & 0 \end{pmatrix}, \tag{3.5}$$

satisfying the side condition $\mathrm{Im}\, M^z(w) \cdot \mathrm{Im}\, w > 0$, where

$$\mathcal{S}[\cdot] := \langle \cdot E_+ \rangle E_+ - \langle \cdot E_- \rangle E_- = 2\langle \cdot E_1 \rangle E_2 + 2\langle \cdot E_2 \rangle E_1, \qquad \mathcal{S} : \mathbb{C}^{2n \times 2n} \to \mathbb{C}^{2n \times 2n}, \tag{3.6}$$

is the *covariance operator* of the Hermitisation of X. Additionally, we define the *self-consistent eigenvalue density* of H^z by (see e.g. [8])

$$\rho^z(E) := \lim_{\eta\to 0^+} \frac{1}{\pi}\mathrm{Im}\,\langle M^z(E+\mathrm{i}\eta)\rangle, \quad E\in\mathbb{R}. \tag{3.7}$$

As summarised in [34, Section 3.1] (see references therein), the density $\rho^z(E)$ has the following properties:

1) if $|z|<1$, then $\rho^z(E)$ has a local minimum at $E=0$ of height $\rho^z(0)\sim(1-|z|^2)^{1/2}$;
2) if $|z|=1$, then $\rho^z(E)$ has an exact cubic cusp singularity at $E=0$;
3) if $|z|>1$, then there exists a small gap $[-\Delta/2,\Delta/2]$ in the support of the symmetric density $\rho^z(E)$, with $\Delta\sim(|z|^2-1)^{3/2}$.

Moreover, extending to the complex plane, if $|z|\le 1$, then for any $w=E+\mathrm{i}\eta$, we have

$$\rho^z(w) := \pi^{-1}|\mathrm{Im}\,\langle M^z(w)\rangle| \sim (1-|z|^2)^{1/2} + (|E|+\eta)^{1/3}, \qquad |E|\le c, \quad 0\le\eta\le 1, \tag{3.8}$$

for some small $c>0$, while in the complementary regime $|z|>1$, for any $w=\frac{\Delta}{2}+\kappa+\mathrm{i}\eta$, we have

$$\rho^z(w) \sim \begin{cases} (|\kappa|+\eta)^{1/2}(\Delta+|\kappa|+\eta)^{-1/6}, & \kappa\in[0,c] \\ \frac{\eta}{(\Delta+|\kappa|+\eta)^{1/6}(|\kappa|+\eta)^{1/2}}, & \kappa\in[-\Delta/2,0] \end{cases}, \quad 0\le\eta\le 1. \tag{3.9}$$

In this paper we focus on the imaginary axis, *i.e.*, $w=\mathrm{i}\eta$, and we may often drop the superscript z and the argument $w=\mathrm{i}\eta$ for brevity. For instance, we may write $m=m^z(\mathrm{i}\eta)$ for brevity. In particular,

$$\rho := \frac{|\mathrm{Im}\, m|}{\pi} \sim \begin{cases} \eta^{1/3} + |1-|z|^2|^{1/2}, & |z|\le 1 \\ \frac{\eta}{|1-|z|^2|+\eta^{2/3}}, & |z|>1 \end{cases}, \qquad 0\le\eta\le 1. \tag{3.10}$$

Now we state the local law for the resolvent; see [16, Theorem 3.4] for $|z|\le 1-\tau$ and [34, Theorem 3.1] for $||z|-1|\le\tau$ (see also [17]), where $\tau>0$ is a small constant.

Theorem 3.1. *There is sufficiently small constants $\tau,\tau'>0$ such that for any $z\in\mathbb{C}$ with $|z|\le 1+\tau$ and for any $w=E+\mathrm{i}\eta$ with $|E|\le\tau'$ we have*

$$\big|\langle \boldsymbol{x}, (G^z(w)-M^z(w))\boldsymbol{y}\rangle\big| \prec \|\boldsymbol{x}\|\|\boldsymbol{y}\|\left(\sqrt{\frac{\rho^z(w)}{n|\eta|}} + \frac{1}{n|\eta|}\right), \tag{3.11}$$

$$\big|\big\langle \big(G^z(w)-M^z(w)\big)A\big\rangle\big| \prec \frac{\|A\|}{n|\eta|}, \tag{3.12}$$

uniformly in spectral parameter $|\eta| > 0$*, deterministic vectors* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{2n}$*, and matrices* $A \in \mathbb{C}^{2n \times 2n}$*.*

Using standard arguments in e.g. [3, Corollary 1.10-1.11], we obtain the following rigidity estimates on the eigenvalues from Theorem 3.1; see also [34, Corollary 3.2] for $|z| > 1$ only.

Corollary 3.2. *Fix a small constant* $\tau > 0$ *and pick* z *such that* $|z| \leq 1 + \tau$*. Then there exists a small constant* $c > 0$ *such that*

$$|\lambda_i^z - \gamma_i^z| \prec \max\left\{ \frac{1}{|i|^{1/4} n^{3/4} + \sqrt{(1 - |z|^2)_+} n}, \frac{\Delta^{1/9}}{n^{2/3} |i|^{1/3}} \right\}, \qquad |i| \leq cn \tag{3.13}$$

with Δ *denoting the size of the gap around zero in the support of* ρ^z *(in particular* $\Delta = 0$ *if* $|z| \leq 1$*), where* γ_i^z *is the* i*-th quantile of* ρ^z *given by*

$$\int_0^{\gamma_i^z} \rho^z(x)\, \mathrm{d}x = \frac{i}{2n}, \qquad \gamma_{-i}^z := -\gamma_i^z, \qquad i \in [n]. \tag{3.14}$$

In addition, there exists a small constant $c' > 0$ *such that*

$$\left| \#\{j : E_1 \leq \lambda_j^z \leq E_2\} - 2n \int_{E_1}^{E_2} \rho^z(x) \mathrm{d}x \right| \prec 1, \qquad |E_i| \leq c', \quad i = 1, 2. \tag{3.15}$$

3.2. Local laws for multiple resolvents

The main technical result of this manuscript is local laws for the product of two Green functions

$$G^{z_1}(\mathrm{i}\eta_1) A G^{z_2}(\mathrm{i}\eta_2), \qquad A \in \mathbb{C}^{2n \times 2n}, \quad G^{z_i}(\mathrm{i}\eta_i) \in \mathbb{C}^{2n \times 2n},\ i = 1, 2,$$

with different parameters $z_i \in \mathbb{C}$ and $\eta_i \in \mathbb{R} \setminus \{0\}$ $(i = 1, 2)$, which yield the overlap bounds of the eigenvectors stated in Theorem 2.2. Before we state the local laws, we first introduce the deterministic approximation matrices of $G^{z_1}(\mathrm{i}\eta_1) A G^{z_2}(\mathrm{i}\eta_2)$, *i.e.*,

$$M_{12}^A := M_{12}^A(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2) = \mathcal{B}_{12}^{-1}\Big[M^{z_1}(\mathrm{i}\eta_1) A M^{z_2}(\mathrm{i}\eta_2) \Big], \tag{3.16}$$

where the two–body stability operator $\mathcal{B}_{12} : \mathbb{C}^{2n \times 2n} \to \mathbb{C}^{2n \times 2n}$ is given by

$$\mathcal{B}_{12} := \mathcal{B}_{12}(z_1, \omega_1, z_2, \omega_2) = 1 - M^{z_1}(\omega_1) \mathcal{S}[\cdot] M^{z_2}(\omega_2), \tag{3.17}$$

with the covariance operator $\mathcal{S}[\cdot]$ defined in (3.6); see also (4.10) later for the definition of $\mathcal{B}_{ij}$ with general $i, j \in \mathbb{N}$. Using the identity $\operatorname{Im} G = \frac{1}{2\mathrm{i}}(G - G^*)$ twice and the definition of M_{12}^A in (3.16), we further define the deterministic approximating matrix of $\operatorname{Im} G^{z_1}(\mathrm{i}\eta_1) A \operatorname{Im} G^{z_2}(\mathrm{i}\eta_2)$ as a fourfold linear combination of $M_{12}^A(\pm\mathrm{i}\eta_1, z_1, \pm\mathrm{i}\eta_2, z_2)$ as defined in (3.16), *i.e.*,

$$\begin{aligned}\widehat{M}_{12}^A :=& \widehat{M}_{12}^A(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2)\\ =& \frac{1}{4}\big(M_{12}^A(\mathrm{i}\eta_1, z_1, -\mathrm{i}\eta_2, z_2) + M_{12}^A(-\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2) - M_{12}^A(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2)\\ &- M_{12}^A(-\mathrm{i}\eta_1, z_1, -\mathrm{i}\eta_2, z_2)\big).\end{aligned} \tag{3.18}$$

To approximate the product $G^{z_1}(\mathrm{i}\eta_1)A_1 G^{z_2}(\mathrm{i}\eta_2)A_2 G^{z_1}(\mathrm{i}\eta_1)$, we also define

$$M_{121}^{A_1,A_2} := \mathcal{B}_{11}^{-1}\Big[M^{z_1}(\mathrm{i}\eta_1)A_1 M_{21}^{A_2} + M^{z_1}(\mathrm{i}\eta_1)\mathcal{S}\big[M_{12}^{A_1}\big]M_{21}^{A_2}\Big], \tag{3.19}$$

with $M_{21}^A = M_{21}^A(\mathrm{i}\eta_2, z_2, \mathrm{i}\eta_1, z_1)$ defined as in (3.16).

The proof of the following lemma is presented in Appendix A.

Lemma 3.3. *For any bounded deterministic matrices $A_1, A_2 \in \mathbb{C}^{2n\times 2n}$, we have*

$$\|M_{12}^{A_1}(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2)\| \lesssim \frac{1}{\gamma}, \tag{3.20}$$

$$\big|\langle \widehat{M}_{12}^{A_1}(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2)A_2\rangle\big| \lesssim \frac{\rho_1\rho_2}{\widehat{\gamma}}, \tag{3.21}$$

$$\|M_{121}^{A_1,A_2}(\mathrm{i}\eta_1, A_1, \mathrm{i}\eta_2, A_2, \mathrm{i}\eta_1)\| \lesssim \frac{1}{\eta_*\gamma}, \tag{3.22}$$

uniformly for any $|z_i| < 10$ and $0 < |\eta_i| < 10$, where $\eta_* := \min_{i=1,2}|\eta_i|$ and $\rho^* := \max_{i=1,2}|\rho_i|$ with $\rho_i := \pi^{-1}|\operatorname{Im} m^{z_i}(\mathrm{i}\eta_i)|$, and the following parameters:

$$\gamma = \gamma(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2) := \frac{|z_1 - z_2|^2 + \rho_1|\eta_1| + \rho_2|\eta_2| + \left(\frac{\eta_1}{\rho_1}\right)^2 + \left(\frac{\eta_2}{\rho_2}\right)^2}{|z_1 - z_2| + \rho_1^2 + \rho_2^2 + \frac{\eta_1}{\rho_1} + \frac{\eta_2}{\rho_2}}, \tag{3.23}$$

$$\widehat{\gamma} = \widehat{\gamma}(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2) := |z_1 - z_2|^2 + \rho_1|\eta_1| + \rho_2|\eta_2| + \left(\frac{\eta_1}{\rho_1}\right)^2 + \left(\frac{\eta_2}{\rho_2}\right)^2. \tag{3.24}$$

Theorem 3.4. *Fix small constants $\epsilon, \tau > 0$. Then, for deterministic matrices $\|A_i\| \lesssim 1$ and vectors $\|\boldsymbol{x}\|, \|\boldsymbol{y}\| \lesssim 1$, the following isotropic local law:*

$$\Big|\big\langle \boldsymbol{x}, \big(G^{z_1}(\mathrm{i}\eta_1)A_1 G^{z_2}(\mathrm{i}\eta_2) - M_{12}^{A_1}\big)\boldsymbol{y}\big\rangle\Big| \prec \frac{1}{\sqrt{n\gamma}\eta_*}, \tag{3.25}$$

and the averaged local law:

$$\left|\left\langle\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2)-M_{12}^{A_1}\right)A_2\right\rangle\right|\prec\frac{1}{\sqrt{n\ell}\gamma}\wedge\frac{1}{n\eta_*^2},\tag{3.26}$$

hold uniformly for any $|z_i|\le 1+\tau$ *and* $\ell\ge n^{-1+\epsilon}$ *with* $\eta_*:=\min_{i=1,2}|\eta_i|$, $\ell:=\min_{i=1,2}\rho_i|\eta_i|$, *and the deterministic matrix* $M_{12}^A=M_{12}^A(\mathrm{i}\eta_1,z_1,\mathrm{i}\eta_2,z_2)$ *is defined in (3.16).*

3.3. Improved local law at the edge

We state the following improved local law for $\operatorname{Im}G^{z_1}A\operatorname{Im}G^{z_2}$ than (3.26), with $\operatorname{Im}G=\frac{1}{2\mathrm{i}}(G-G^*)$, up to the edge regime $|z_i|\le 1+n^{-1/2+\tau}$.

Theorem 3.5. *Fix small constants* $\epsilon,\tau>0$ *with* $0<\tau<\epsilon/10$*. Then, for any deterministic matrices* $\|A_i\|\lesssim 1$*, the following local law*

$$\left|\langle\left(\operatorname{Im}G^{z_1}(\mathrm{i}\eta_1)A_1\operatorname{Im}G^{z_2}(\mathrm{i}\eta_2)-\widehat{M}_{12}^{A_1}\right)A_2\rangle\right|\prec\frac{1}{\sqrt{n\ell}}\frac{\rho_1\rho_2}{\widehat{\gamma}},\tag{3.27}$$

holds uniformly for any $z_i\in\mathbb{C}$ *and* $\eta_i\in\mathbb{R}$ $(i=1,2)$ *such that* $|z_i|\le 1+n^{-1/2+\tau}$ *and* $\ell=\min_{i=1,2}\rho_i|\eta_i|\ge n^{-1+\epsilon}$*, with* $\widehat{\gamma}$ *given by (3.24), and the deterministic matrix* $\widehat{M}_{12}^A=\widehat{M}_{12}^A(\mathrm{i}\eta_1,z_1,\mathrm{i}\eta_2,z_2)$ *defined in (3.18).*

By a standard argument, we obtain the following optimal bound for the overlaps between singular vectors of $X-z_i$ for different $z_1,z_2\in\mathbb{C}$. The proof is similar to [30, Theorem 3.1] for $|z|\le 1-\tau$ and [35, Corollary 3.4] for $1\le|z_i|\le 1+n^{-1/2+\tau}$. The only difference is to replace the non-optimal local law for $\operatorname{Im}G^{z_1}\operatorname{Im}G^{z_2}$ (*e.g.*, [35, Theorem 3.3]) with the optimal version stated in Theorem 3.5.

Corollary 3.6. *Fix a small constant* $\tau>0$*. Then there exists a small* $0<\omega_c<\tau/10$ *such that, for any* $z_i\in\mathbb{C}$ $(i=1,2)$ *with* $|z_i|\le 1+n^{-1/2+\tau}$*, the following holds with very high probability,*

$$|\langle\boldsymbol{u}_i^{z_1},\boldsymbol{u}_j^{z_2}\rangle|^2+|\langle\boldsymbol{v}_i^{z_1},\boldsymbol{v}_j^{z_2}\rangle|^2\lesssim\frac{n^{-1+2\tau}}{|z_1-z_2|^2+n^{-1}},\qquad |i|,|j|\le n^{\omega_c}.\tag{3.28}$$

We conclude this section by showing that the bound (3.27) readily implies those in Theorem 2.2 and Corollary 2.5. Actually, Corollary 3.6 is obtained as a by-product in the proof of Theorem 2.2 (see (3.32) below).

Proof of Theorem 2.2. Let $\eta_{\mathrm{f}}=\eta_{\mathrm{f}}(z)$ denote the local fluctuation scale of the eigenvalues of H^z close to zero, which, for $|z|\le 1$ is implicitly defined by

$$\int_{-\eta_{\mathrm{f}}}^{\eta_{\mathrm{f}}}\rho^z(x)\,\mathrm{d}x=\frac{1}{n}.$$

From (3.10), note that $\eta_{\mathrm{f}} \sim n^{-1}$ when $|z| < 1$ and $\eta_{\mathrm{f}} \sim n^{-3/4}$ when $|z| = 1$. For $|z| > 1$ the density ρ^z develops a gap Δ^z around zero, and one can see that in this case $\eta_{\mathrm{f}} \sim |\Delta^z|^{1/9}/n^{2/3}$ and $|\Delta^z| \sim (|z|^2-1)^{3/2}$ (see e.g. [34, Appendix A] and references therein). Fix a small $\xi > 0$, and define $\eta_i := n^\xi \eta_{\mathrm{f}}(z_i)$, for $z_i \in \mathbb{C}$. Then, by (3.27) and (3.21), we obtain

$$\sup_{|z_1|,|z_2|\le 1+n^{-1/2+\tau}} \frac{\widehat{\gamma}}{\rho_1\rho_2}\big|\langle \operatorname{Im} G^{z_1}(\mathrm{i}\eta_1)\operatorname{Im} G^{z_2}(\mathrm{i}\eta_2)\rangle\big| \lesssim 1, \tag{3.29}$$

with very high probability. Here $\rho_i = \pi^{-1}|\operatorname{Im} m^{z_i}(\mathrm{i}\eta_i)|$ and $\widehat{\gamma} = \widehat{\gamma}(\mathrm{i}\eta_1, z_1, \mathrm{i}\eta_2, z_2)$. We point out that to obtain the supremum bound in (3.29) we used (3.27) and a standard grid argument using the Lipschitz continuity of $z \mapsto \operatorname{Im} G^z$.

Next, by the rigidity estimate of the eigenvalues of H^z in (3.13), we have

$$\max_{|j|\le n^{\omega_c}} \frac{|\lambda_j^z|}{\eta_{\mathrm{f}}(z)} \lesssim 1, \tag{3.30}$$

with very high probability, for some small constant $0 < \omega_c < \xi/10$. Note that for any $j \in [n]$, λ_j^z is the j-th non-negative eigenvalue of H^z in the non-deceasing order and $\lambda_{-j}^z = -\lambda_j^z$.

We notice that by the spectral decomposition of $\operatorname{Im} G = (G - G^*)/(2\mathrm{i})$ and the chiral symmetry of H^z in (3.1), we have

$$\langle \operatorname{Im} G^{z_1}(\mathrm{i}\eta_1)\operatorname{Im} G^{z_2}(\mathrm{i}\eta_2)\rangle = \frac{4}{n}\sum_{j,k=1}^{n} \frac{\eta_1\eta_2\big(|\langle \boldsymbol{u}_j^{z_1}, \boldsymbol{u}_k^{z_2}\rangle|^2 + |\langle \boldsymbol{v}_j^{z_1}, \boldsymbol{v}_k^{z_2}\rangle|^2\big)}{\big((\lambda_j^{z_1})^2+\eta_1^2\big)\big((\lambda_k^{z_2})^2+\eta_2^2\big)}, \tag{3.31}$$

where $(\lambda_j^{z_i})_{j=1}^n$ are the non-decreasing singular values of $X - z_i$ with $i = 1, 2$, and $\{\boldsymbol{u}_j^{z_i}\}_{j=1}^n$, $\{\boldsymbol{v}_j^{z_i}\}_{j=1}^n$ are the corresponding left and right singular vectors normalized such that $\|\boldsymbol{u}_j^{z_i}\|^2 = \|\boldsymbol{v}_j^{z_i}\|^2 = 1/2$. Then, on the very high probability event on which (3.29)–(3.30) hold, we obtain (recall $\eta_i := n^\xi\eta_{\mathrm{f}}(z_i)$)

$$\begin{aligned}
&\sup_{|z_1|,|z_2|\le 1+n^{-1/2+\tau}} (n|z_1-z_2|^2+1)\big[|\langle \boldsymbol{u}_1^{z_1}, \boldsymbol{u}_1^{z_2}\rangle|^2 + |\langle \boldsymbol{v}_1^{z_1}, \boldsymbol{v}_1^{z_2}\rangle|^2\big] \\
&\quad\le \sup_{|z_1|,|z_2|\le 1+n^{-1/2+\tau}} (n|z_1-z_2|^2+1)(n\eta_1\eta_2)\langle \operatorname{Im} G^{z_1}(\mathrm{i}\eta_1)\operatorname{Im} G^{z_2}(\mathrm{i}\eta_2)\rangle \\
&\quad\lesssim \sup_{|z_1|,|z_2|\le 1+n^{-1/2+\tau}} \frac{|z_1-z_2|^2+n^{-1}}{\widehat{\gamma}}(n\eta_1\rho_1)(n\eta_2\rho_2) \lesssim n^{4\xi},
\end{aligned} \tag{3.32}$$

with very high probability, where in the last inequality we used (3.24) to show that $(|z_1-z_2|^2+n^{-1})/\widehat{\gamma} \lesssim 1$ and we used that $n\eta_i\rho_i \lesssim n^{2\xi}$ by the definition of η_{f} above (3.29) (note that η_{f} has the property that $n\eta_{\mathrm{f}}\rho^z(\mathrm{i}\eta_{\mathrm{f}}) \sim 1$).

Recall the basic relation (1.7) that $z \in \operatorname{Spec}(X)$ if and only if the least singular value λ_1^z is zero. Moreover, if $z \in \operatorname{Spec}(X)$, then the left and right eigenvector of X associated to the eigenvalue z are exactly the left and right singular vector of $X - z$. This implies that

$$\begin{aligned}&\sup_{i,j\in[n]} \left(n|\sigma_i - \sigma_j|^2 + 1\right)\left[\frac{\left|\langle \boldsymbol{r}_i, \boldsymbol{r}_j\rangle\right|^2}{\|\boldsymbol{r}_i\|^2\|\boldsymbol{r}_j\|^2} + \frac{\left|\langle \boldsymbol{l}_i, \boldsymbol{l}_j\rangle\right|^2}{\|\boldsymbol{l}_i\|^2\|\boldsymbol{l}_j\|^2}\right]\\ =&4 \sup_{z_1,z_2\in\mathrm{Spec}(X)} (n|z_1 - z_2|^2 + 1)\big[|\langle \boldsymbol{u}_1^{z_1}, \boldsymbol{u}_1^{z_2}\rangle|^2 + |\langle \boldsymbol{v}_1^{z_1}, \boldsymbol{v}_1^{z_2}\rangle|^2\big],\end{aligned} \tag{3.33}$$

where $\boldsymbol{u}_1^{z_i}$ and $\boldsymbol{v}_1^{z_i}$ $(i = 1, 2)$ are the left and right singular vectors associated to the least singular value of $X - z_i$, normalized such that $\|\boldsymbol{u}_1^{z_i}\|^2 = \|\boldsymbol{v}_1^{z_i}\|^2 = 1/2$. Note that with very high probability the spectrum of X lies inside the disk $|z| \le 1 + n^{-1/2+\tau}$ (see [9, Theorem 2.1]). Hence, with very high probability,

$$\begin{aligned}&\sup_{i,j\in[n]} \left(n|\sigma_i - \sigma_j|^2 + 1\right)\left[\frac{\left|\langle \boldsymbol{r}_i, \boldsymbol{r}_j\rangle\right|^2}{\|\boldsymbol{r}_i\|^2\|\boldsymbol{r}_j\|^2} + \frac{\left|\langle \boldsymbol{l}_i, \boldsymbol{l}_j\rangle\right|^2}{\|\boldsymbol{l}_i\|^2\|\boldsymbol{l}_j\|^2}\right]\\ \lesssim& \sup_{|z_1|,|z_2|\le 1+n^{-1/2+\tau}} (n|z_1 - z_2|^2 + 1)\big[|\langle \boldsymbol{u}_1^{z_1}, \boldsymbol{u}_1^{z_2}\rangle|^2 + |\langle \boldsymbol{v}_1^{z_1}, \boldsymbol{v}_1^{z_2}\rangle|^2\big].\end{aligned} \tag{3.34}$$

This, together with (3.32), immediately implies (2.3)–(2.4) and concludes the proof. □

Proof of Corollary 2.5. Fix any small $\xi > 0$. From (2.4), we know that for any $i, j \in [N]$, with very high probability,

$$\begin{aligned}|\mathcal{O}_{ij}| \le&\frac{n^\xi}{n|\sigma_i - \sigma_j|^2 + 1}\sqrt{\mathcal{O}_{ii}\mathcal{O}_{jj}}\\ \lesssim&\frac{n^\xi}{n|z_1 - z_2|^2 + 1}\sqrt{\mathcal{O}_{ii}\mathcal{O}_{jj}},\end{aligned} \tag{3.35}$$

where the last step follows if we assume that the corresponding eigenvalues $\sigma_i \in \mathcal{D}_{z_1}^\xi$ and $\sigma_j \in \mathcal{D}_{z_2}^\xi$, with $|z_1 - z_2| \ge n^{-1/2+\epsilon}$ and $\xi < \epsilon/10$. Moreover, from the local law mentioned in Remark 2.6, for any $z \in \mathbb{C}$, with very high probability,

$$\#\{\sigma_i \in \mathrm{Spec}(X) : \sigma_i \in \mathcal{D}_z^\xi\} \le n^{3\xi}. \tag{3.36}$$

We next recall various bounds on the expectation of the diagonal overlaps from [44, Theorems 2.7, 2.9]. For example, in the complex case, using [44, Eq (2.14)], there exists an event Ξ with very high probability, i.e. $\boldsymbol{P}(\Xi) \ge 1 - n^{-D}$, such that for any $z \in \mathbb{C}$,

$$\boldsymbol{E}\Big(\sum_{\sigma_i\in\mathcal{D}_z^\xi} |\mathcal{O}_{ii}| \cdot \boldsymbol{1}_\Xi\Big) \lesssim n^{1+3\xi}. \tag{3.37}$$

Combining the very high probability bound (3.35)-(3.36) with the expectation bound (3.37) and using a simple Cauchy-Schwarz inequality, we obtain the desired result in

(2.7) for the complex case. The remaining statements of Corollary 2.5 follow similarly by combining (3.35)-(3.36) with the corresponding expectation bounds from Eq (2.15), Eq (2.17), and Eq (2.18) of [44], respectively. This finishes the proof. □

4. Zig flow

In this section we show that if certain local laws (see [29, Theorem 5.2]) hold for $\eta \sim 1$ (for simplicity in this introductory paragraph we focus on $|z| < 1$), then they can be propagated down to $\eta \gg n^{-1}$ using the Ornstein–Uhlenbeck and characteristic flows. The main results of this section (Propositions 4.3–4.4) are the fundamental inputs to prove Theorem 3.4. In fact, they show that if the local laws from Theorem 3.4 hold on a global scale ($\eta \sim 1$), then they also hold down to the optimal η-scale ($\eta \gg 1/n$), at the price of adding a Gaussian component. This Gaussian component will then be removed in Section 5 using a Green's Function Comparison Theorem (GFT). We point out that some parts of this section, like the definition and properties of the characteristics flow and the analysis of the flow in the averaged case, are very close to [35, Section 5]; however, we decided to present the details here for the convenience of the reader. Consider the Ornstein–Uhlenbeck flow

$$\mathrm{d}X_t = -\frac{1}{2}X_t\mathrm{d}t + \frac{\mathrm{d}B_t}{\sqrt{N}}, \qquad X_0 = X, \tag{4.1}$$

where X is an i.i.d. matrix satisfying Assumption 2.1. Here B_t is an $n \times n$ matrix with entries being real or complex standard i.i.d. Brownian motions, depending on the symmetry class of X.

We now define the characteristics flow associated with (4.1). Define the Hermitisation of X_t by

$$W_t := \begin{pmatrix} 0 & X_t \\ X_t^* & 0 \end{pmatrix}, \tag{4.2}$$

and consider the spectral parameters $\Lambda \in \mathbb{C}^{2n\times 2n}$ defined by

$$\Lambda := \begin{pmatrix} \mathrm{i}\eta & z \\ \overline{z} & \mathrm{i}\eta \end{pmatrix}, \qquad \eta \in \mathbb{R}, \qquad z \in \mathbb{C}. \tag{4.3}$$

Throughout this section we consider spectral parameters being $2n \times 2n$ block–constant matrices, i.e. we consider Λ's consisting of four $n \times n$ blocks each one of them being a constant multiple of the n–dimensional identity matrix. The main object of interest within this section is the resolvent $(W_t - \Lambda_t)^{-1}$, with W_t from (4.2) and Λ_t being the solution of the *characteristic flow*:

$$\partial_t\Lambda_t = -\mathcal{S}[M(\Lambda_t)] - \frac{\Lambda_t}{2}, \tag{4.4}$$

with $\mathcal{S}$ being the covariance operator from (3.6) and $M(\Lambda_t) := M^{z_t}(\mathrm{i}\eta_t)$. Note that if Λ_0 is constant in each of its four blocks so is Λ_t for any $t \geq 0$, hence the equation (4.4) reduces to two scalar equations:

$$\partial_t \eta_t = -\mathrm{Im}\, m^{z_t}(\mathrm{i}\eta_t) - \frac{\eta_t}{2}, \qquad \partial_t z_t = -\frac{z_t}{2}. \tag{4.5}$$

We point out that there is a one to one correspondence between (η_t, z_t) and Λ_t. For this reason throughout this section we may refer to one or the other interchangeably. The η_t–characteristics are decreasing in time, for this reason, throughout this section, we only consider the characteristics up to the first time they touch the real axis

$$T^*(\eta_0, z_0) := \sup\big\{t \geq 0 : \mathrm{sgn}(\eta_t) = \mathrm{sgn}(\eta_0)\big\}. \tag{4.6}$$

Even if not stated explicitly, in the reminder of this section we always consider initial conditions such that $|\eta_0| + |z_0| \lesssim 1$, which is equivalent to $\|\Lambda_0\| \lesssim 1$. In the following (trivially checkable) lemma we summarize several properties of the characteristics (4.5) (cf. [35, Lemma 5.1]):

Lemma 4.1. *Fix Λ_0 with $\|\Lambda_0\| \lesssim 1$, and let Λ_t be the solution of (4.4) with initial condition Λ_0. Define $M_t := M^{z_t}(\mathrm{i}\eta_t)$ and $\rho_t := \pi^{-1}|\mathrm{Im}\, m_t|$, then for any $0 \leq s \leq t \leq T^*(\eta_0, z_0)$ we have*

$$z_t = e^{-(t-s)/2} z_s, \qquad M_t = e^{(t-s)/2} M_s, \qquad \frac{|\eta_t|}{\rho_t} = e^{s-t}\frac{|\eta_s|}{\rho_s} - \pi(1 - e^{s-t}). \tag{4.7}$$

Additionally, the maps $t \mapsto |\eta_t|$, $t \mapsto |\eta_t|\rho_t$, and $t \mapsto |\eta_t|/\rho_t$ are strictly decreasing, and for any $\alpha > 1$ there is a constant C_α such that

$$\int_s^t \frac{1}{|\eta_r|^\alpha}\,\mathrm{d}t \leq \frac{C_\alpha}{|\eta_t|^{\alpha-1}\rho_t}, \qquad \int_s^t \frac{\rho_r}{|\eta_r|}\,\mathrm{d}r = \frac{1}{\pi}\left[\log\left(\frac{\eta_s}{\eta_t}\right) - \frac{t-s}{2}\right]. \tag{4.8}$$

Additionally, by standard ODE theory we have (cf. [35, Lemma 5.2]):

Lemma 4.2. *Fix any $T \leq 1$ and pick Λ of the form (4.3) with $|\eta| > 0$ and $\|\Lambda\| \leq 1$. Then there exist Λ_0 with $\|\Lambda_0\| \lesssim 1$ and $T^*(\eta_0, z_0) > T$, and the solution Λ_t of (4.4) with initial condition Λ_0 is so that $\Lambda_T = \Lambda$. Furthermore, there exists an N–independent constant $c_* > 0$ such that* $\mathrm{dist}(\mathrm{i}\eta_0, \mathrm{supp}(\rho^{z_0})) \geq c_* T$.

In the rest of this section we will consider times $t \leq T^* := \min_i T_i^*(\eta_{i,0}, z_{i,0})$. Note that by the first two relations in (4.7), for $s, t \leq T^*$, it follows

$$|z_{1,s} - z_{2,s}| \sim |z_{1,t} - z_{2,t}|, \qquad \rho_s \sim \rho_t. \tag{4.9}$$

We are now ready to state the two main results of this section. We first state the averaged case, whose statement is very similar to [35, Proposition 5.4] (note the difference in the definition of γ in (4.13)), and then we state the isotropic case. For deterministic matrices $A_1, A_2 \in \mathbb{C}^{2n\times 2n}$, we define the deterministic approximations for the product of two and three resolvents (using the notation $M_i := M^{z_i}(\mathrm{i}\eta_i)$)

$$M_{12}^{A_1} := \mathcal{B}_{12}^{-1}\big[M_1 A_1 M_2\big], \qquad \mathcal{B}_{ij}[\cdot] := 1 - M_i \mathcal{S}[\cdot] M_j \tag{4.10}$$

$$M_{121}^{A_1,A_2} := \mathcal{B}_{11}^{-1}\Big[M_1 A_1 M_{21}^{A_2} + M_1 \mathcal{S}\big[M_{12}^{A_1}\big] M_{21}^{A_2}\Big], \tag{4.11}$$

where $\mathcal{S} : \mathbb{C}^{2n\times 2n} \to \mathbb{C}^{2N\times 2N}$ is the covariance operator defined in (3.6). If we replace M_i by $M_{i,t} = M^{z_{i,t}}(\mathrm{i}\eta_{i,t})$ in (4.10)–(4.11), we then denote the corresponding quantities by $M_{12,t}^{A_1}, M_{121,t}^{A_1,A_2}$. We are now ready to state the main results of this section:

Proposition 4.3. *Fix small n–independent constants $\tau, \omega_*, \epsilon > 0$, and for $i = 1,2$ fix spectral parameters*

$$\Lambda_{i,0} := \begin{pmatrix} \mathrm{i}\eta_{i,0} & z_{i,0} \\ \overline{z_{i,0}} & \mathrm{i}\eta_{i,0} \end{pmatrix}, \tag{4.12}$$

with $|z_{i,0}| \le 10$, $|\eta_{i,0}| \le \omega_*$. For $t \le T^*$ let $\Lambda_{i,t}$ be the solution of (4.4) with initial condition $\Lambda_{i,0}$. Let $G_{i,t} := (W_t - \Lambda_{i,t})^{-1}$, and let $M_{12,t}^{A}$ be defined in (4.10). Define the following control parameters and their time–dependent versions[5]*:*

$$\ell = \ell(\eta_1, z_1, \eta_2, z_2) := \min_{i=1,2} \rho_i |\eta_i|, \qquad \ell_t := \ell(\eta_{1,t}, z_{1,t}, \eta_{2,t}, z_{2,t}),$$

$$\gamma = \gamma(\eta_1, z_1, \eta_2, z_2) := \frac{|z_1 - z_2|^2 + \rho_1|\eta_1| + \rho_2|\eta_2| + \left(\frac{\eta_1}{\rho_1}\right)^2 + \left(\frac{\eta_2}{\rho_2}\right)^2}{|z_1 - z_2| + \rho_1^2 + \rho_2^2 + \frac{\eta_1}{\rho_1} + \frac{\eta_2}{\rho_2}}, \qquad \gamma_t := \gamma(\eta_{1,t}, z_{1,t}, \eta_{2,t}, z_{2,t}). \tag{4.13}$$

Assume that for some small $0 < \xi \le \epsilon/10$, with very high probability, with $G_0^z(\mathrm{i}\eta) := (H_0^z - \mathrm{i}\eta)^{-1}$, it holds

$$\big|\big\langle \big(G_0^{z_{1,0}}(\mathrm{i}\eta_1) A_1 G^{z_{2,0}}(\mathrm{i}\eta_2) - M_{12}^{A_1}(\eta_1, z_{1,0}, \eta_2, z_{2,0})\big) A_2 \big\rangle\big| \le \frac{n^\xi}{\sqrt{n\ell_0}\gamma_0} \wedge \frac{n^\xi}{n|\eta_1\eta_2|} \tag{4.14}$$

uniformly in $\|A_i\| \le 1$ and for any $|\eta_i| \in [|\eta_{i,0}|, n^{100}]$, where ℓ_0, γ_0 are evaluated at[6] *$(\eta_1, z_{1,0}, \eta_2, z_{2,0})$. Then,*

[5] Recall that $\rho_i = \rho^{z_i}(\mathrm{i}\eta_i)$, and so in the time dependent case we have $\rho_{i,t} = \rho^{z_{i,t}}(\mathrm{i}\eta_{i,t})$. For example, using these notations, $\ell_t := \min_{i=1,2} \rho_{i,t}\eta_{i,t}$.

[6] We point out that we require (4.14) for $(\eta_1, z_{1,0}, \eta_2, z_{2,0})$, with $|\eta_i| \in [|\eta_{i,0}|, n^{100}]$, and not only for $\eta_i = \eta_{i,0}$.

$$\left|\langle (G_{1,t}A_1G_{2,t} - M_{12,t}^{A_1})A_2\rangle\right| \le \frac{n^{2\xi}}{\sqrt{n\ell_t}\gamma_t} \wedge \frac{n^{2\xi}}{n|\eta_{1,t}\eta_{2,t}|}, \tag{4.15}$$

with very high probability uniformly in $\|A_i\| \le 1$ *and all* $t \le T^*$ *such that* $n\ell_t \ge n^\epsilon$.

We now briefly comment on the significance of the two control parameters (4.13). The condition $n\ell = 1$ denotes the local fluctuation scale of the eigenvalues. We prove local laws always in the regime $n\ell \gg 1$, i.e. above the fluctuation scale. The parameter γ represents a control on the stability operator. In fact, γ is chosen so that (see (3.20))

$$\|M_{12}^A\| \lesssim \frac{\|A\|}{\gamma}, \tag{4.16}$$

uniformly in the spectrum. Furthermore, we also point out that the explicit formulas for ℓ, γ simplify when Λ_i are in the bulk of the spectrum; in fact in this case we have

$$\ell \sim \min_{i=1,2} |\eta_i|, \qquad \gamma \sim |z_1 - z_2|^2 + |\eta_1| + |\eta_2|. \tag{4.17}$$

Close to the edge of the spectrum, i.e. when $|z_i| \approx 1$, we instead have $\gamma \sim |z_1 - z_2| + \eta_1^{2/3} + \eta_2^{2/3}$.

We now state the local law in the isotropic case. We state it in a slightly more general form as it will be needed later in the zag-part of the proof in Section 5, but we point that the most delicate regime is $b = 1$, since is the regime when we gain the most from the parameter γ.

Proposition 4.4. *Fix small n–independent constants $\tau, \omega_*, \epsilon > 0$, and for $i = 1, 2$ fix spectral parameters $\Lambda_{i,0}$ as in (4.12). Fix an exponent $b \in [0, 1]$. For $t \le T^*$ let $\Lambda_{i,t}$ be the solution of (4.4) with initial condition $\Lambda_{i,0}$. Let $G_{i,t} := (W_t - \Lambda_{i,t})^{-1}$, and let $M_{12,t}^A$, $M_{121,t}^{A_1,A_2}$ be defined in (4.10) and (4.11), respectively. Let ℓ_t, γ_t be from (4.13), and assume that for some small $0 < \xi \le \epsilon/10$, with very high probability, it holds*

$$\begin{aligned}
\left|\left\langle \boldsymbol{x}, \left(G_0^{z_{1,0}}(\mathrm{i}\eta_1)A_1G_0^{z_{2,0}}(\mathrm{i}\eta_2) - M_{12}^{A_1}(\eta_1, z_{1,0}, \eta_2, z_{2,0})\right)\boldsymbol{y}\right\rangle\right| &\le \frac{n^\xi (\rho^*)^{\frac{1-b}{2}}}{\sqrt{n}\eta_*^{\frac{3-b}{2}}\gamma_0^{\frac{b}{2}}}, \\
\left|\left\langle \boldsymbol{x}, \left(G_0^{z_{1,0}}(\mathrm{i}\eta_1)A_1G_0^{z_{2,0}}(\mathrm{i}\eta_2)A_2G_0^{z_{1,0}}(\mathrm{i}\eta_1) - M_{121}^{A_1,A_2}(\eta_1, z_{1,0}, \eta_2, z_{2,0})\right)\boldsymbol{y}\right\rangle\right| &\le \frac{n^\xi}{\eta_*^{2-b}\gamma_0^b},
\end{aligned} \tag{4.18}$$

where $\eta_* := |\eta_1| \wedge |\eta_2|$, uniformly in $\|A_i\| \le 1$, $\|\boldsymbol{x}\| + \|\boldsymbol{y}\| \lesssim 1$, and for any $|\eta_i| \in [|\eta_{i,0}|, n^{100}]$, where ℓ_0, γ_0 are evaluated at $(\eta_1, z_{1,0}, \eta_2, z_{2,0})$. Then,

$$\begin{aligned}
\left|\left\langle \boldsymbol{x}, \left(G_{1,t}A_1G_{2,t} - M_{12,t}^{A_1}\right)\boldsymbol{y}\right\rangle\right| &\le \frac{n^{2\xi}(\rho_t^*)^{\frac{1-b}{2}}}{\sqrt{n}\eta_{*,t}^{\frac{3-b}{2}}\gamma_t^{\frac{b}{2}}}, \\
\left|\left\langle \boldsymbol{x}, \left(G_{1,t}A_1G_{2,t}A_2G_{1,t} - M_{121,t}^{A_1,A_2}\right)\boldsymbol{y}\right\rangle\right| &\le \frac{n^{2\xi}}{(\eta_{*,t})^{2-b}\gamma_t^b},
\end{aligned} \tag{4.19}$$

where $\eta_{*,t} := |\eta_{1,t}| \wedge |\eta_{2,t}|$, *with very high probability uniformly in all* $t \leq T^*$ *such that* $n\ell_t \geq n^\epsilon$, $\|A_i\| \leq 1$, *and* $\|\boldsymbol{x}\| + \|\boldsymbol{y}\| \lesssim 1$.

Note that the estimate on the three-resolvent isotropic chain is far from optimal, it shows only that the random chain can be bounded by the size of its deterministic approximation M_{121}, see (3.22). This will, however, be sufficient for our purpose since for the quadratic variation we anyway need only a bound on the random chain itself. We are now ready to present the proofs of Propositions 4.3–4.4. We first present the proof in the averaged case and then consider the isotropic case.

Proof of Proposition 4.3. The proof of this proposition is very similar to [35, Proposition 5.4], in fact, given the bounds on the smallest eigenvalues of the stability operator from Lemma A.1, the proof is the basically the same as the proof of [35, Proposition 5.4]. While in the edge regime it is necessary to follow very precisely two bad directions, as a consequence of the instability of the cusp regime (see e.g. [35, Eq.s (5.46)–(5.47)]), in the bulk there is only one bad direction to follow rendering the argument easier and simpler to follow. For this reason we refer the reader to [35, Proposition 5.4] for a general proof which works everywhere in the spectrum, once the input Lemma A.1 is given, and here we present the simpler proof in the bulk for reader's convenience. Since there are still similarities with estimates performed in the proof of [35, Proposition 5.4], here we highlight the main differences and omit several technical details that can be found in [35, Section 5].

Recall the definition $G_{i,t} = (W_t - \Lambda_{i,t})^{-1}$, then for any $A_1, A_2 \in \mathbb{C}^{2n\times 2n}$, by Itô's formula, we have

$$
\begin{aligned}
&\mathrm{d}\langle (G_{1,t}A_1G_{2,t} - M^{A_1}_{12,t})A_2\rangle \\
&\quad= \sum_{a,b=1}^{2n} \partial_{ab}\langle G_{1,t}A_1G_{2,t}A_2\rangle \frac{\mathrm{d}B_{ab,t}}{\sqrt{n}} + \langle (G_{1,t}A_1G_{2,t} - M^{A_1}_{12,t})A_2\rangle \mathrm{d}t \\
&\qquad + 2\sum_{i\neq j} \langle (G_{1,t}A_1G_{2,t} - M^{A_1}_{12,t})E_i\rangle \langle M^{A_2}_{21,t}E_j\rangle \mathrm{d}t \\
&\qquad + 2\sum_{i\neq j} \langle M^{A_1}_{12,t}E_i\rangle \langle (G_{2,t}A_2G_{1,t} - M^{A_2}_{21,t})E_j\rangle \mathrm{d}t \\
&\qquad + 2\sum_{i\neq j} \langle (G_{1,t}A_1G_{2,t} - M^{A_1}_{12,t})E_i\rangle \langle (G_{2,t}A_2G_{1,t} - M^{A_2}_{21,t})E_j\rangle \mathrm{d}t \\
&\qquad + \langle (G_{1,t} - M_{1,t})\rangle \langle G^2_{1,t}A_1G_{2,t}A_2\rangle \mathrm{d}t + \langle (G_{2,t} - M_{2,t})\rangle \langle G^2_{2,t}A_2G_{1,t}A_1\rangle \mathrm{d}t.
\end{aligned}
\tag{4.20}
$$

Here ∂_{ab} denotes the directional derivative in the direction $(W_t)_{ab}$, $B_{ab,t}$ denotes the (a,b)–entry of the matrix Brownian motion B_t from (4.1), and $\sum_{i\neq j}$ denotes a summation over the pairs $(i,j) \in \{(1,2),(2,1)\}$. We point out that here we also used the following lemma describing the evolution of the deterministic approximation for the product of two resolvents:

Lemma 4.5 *(Lemma 5.5 [30]). For any matrices $A_1, A_2 \in \mathbb{C}^{2N\times 2N}$, we have*

$$\partial_t \langle M_{12,t}^{A_1} A_2\rangle = \langle M_{12,t}^{A_1} A_2\rangle + \langle \mathcal{S}[M_{12,t}^{A_1}] M_{21,t}^{A_2}\rangle. \tag{4.21}$$

For concreteness, in the remainder of the proof we assume $\eta_{1,t}\eta_{2,t} < 0$, the other case being completely analogous. Define

$$Y_t := \max_{B_1,B_2\in\{A_1,A_2,E_+,E_-\}} \big|\langle (G_{1,t}B_1G_{2,t} - M_{12,t}^{B_1})B_2\rangle\big|, \tag{4.22}$$

and the stopping time

$$\tau_1 := \inf\big\{t\geq 0 : Y_t = n^{2\xi}\alpha_t\big\} \wedge T^*, \qquad \alpha_t := \frac{1}{n|\eta_{1,t}\eta_{2,t}|} \wedge \frac{1}{\sqrt{n\ell_t}\gamma_t}. \tag{4.23}$$

Recall that by (4.17) in the bulk regime $\gamma_t \sim |z_{1,t} - z_{2,t}|^2 + |\eta_{1,t}| + |\eta_{2,t}|$. From the monotonicity of $t \mapsto |\eta_t|$ (see Lemma 4.1), and (4.9) it also follows that $\gamma_s \gtrsim \gamma_t$ for $s \leq t$. If $|z_{1,t} - z_{2,t}| \geq n^{-\delta}$, for some very small fixed $0 < \delta \leq \xi/10$, then the desired result follows by a simple (almost) global law (see e.g. [30, Lemma 5.4]). For the remainder of the proof we thus assume $|z_{1,t} - z_{2,t}| \leq n^{-\delta}$. Note that $|z_{1,t} - z_{2,t}| \sim |z_{1,s} - z_{2,s}|$ for any s, t, by (4.9), it thus makes no difference at which time we impose conditions on $|z_{1,t} - z_{2,t}|$.

Writing

$$E_1 = \frac{1}{2}\big(E_+ + E_-\big), \qquad E_2 = \frac{1}{2}\big(E_+ - E_-\big), \tag{4.24}$$

and proceeding similarly to [35, Eqs. (5.28)–(5.41)], from (4.20) we readily obtain

$$\begin{aligned}
\mathrm{d}\langle (G_{1,t}A_1G_{2,t} - M_{12,t}^{A_1})A_2\rangle
&= \langle G_{1,t}A_1G_{2,t} - M_{12,t}^{A_1}\rangle\langle M_{21,t}^{A_2}\rangle \mathrm{d}t - \langle (G_{1,t}A_1G_{2,t} - M_{12,t}^{A_1})E_-\rangle\langle M_{21,t}^{A_2}E_-\rangle \mathrm{d}t \\
&\quad + \langle M_{12,t}^{A_1}\rangle\langle G_{2,t}A_2G_{1,t} - M_{21,t}^{A_2}\rangle \mathrm{d}t - \langle M_{12,t}^{A_1}E_-\rangle\langle (G_{2,t}A_2G_{1,t} - M_{21,t}^{A_2})E_-\rangle \mathrm{d}t \\
&\quad + h_t\,\mathrm{d}t + \mathrm{d}e_t.
\end{aligned} \tag{4.25}$$

Here e_t, h_t are a martingale and a forcing term such that

$$\int_0^{t\wedge\tau_1} \big|h_s\big|\,\mathrm{d}s + \Bigg[\int_0^{\cdot} \mathrm{d}e_s\Bigg]_{t\wedge\tau_1}^{1/2} \lesssim n^{\xi}\alpha_{t\wedge\tau_1}, \tag{4.26}$$

with very high probability. Here $[\cdot]_t$ denotes the quadratic variation process.

By explicit but fairly tedious computations we find (see (A.35) in Appendix A) (recall $\|A_i\| \le 1$)

$$\left|\langle M_{12,t}^{A_1} E_- \rangle\right| \lesssim \frac{|z_1 - z_2|}{\gamma_t} \le \frac{n^{-\delta}}{\gamma_t}, \qquad \left|\langle M_{21,t}^{A_2} \rangle\right| \le \left|\langle M_{21,t}^{E_+} \rangle\right| + \mathcal{O}\left(\frac{n^{-\delta}}{\gamma_t}\right). \tag{4.27}$$

Integrating (4.25) in time (from 0 to $t \wedge \tau_1$) and using (4.27) to estimate the $M_{12,t}, M_{21,t}$–terms in (4.25), we thus obtain

$$Y_{t\wedge\tau_1} \le Y_0 + \int_0^{t\wedge\tau_1} \left(2\left|\langle M_{21,s}^{E_+} \rangle\right| + \frac{n^{-\delta}}{\gamma_s}\right) Y_s \, \mathrm{d}s + n^{\xi} \alpha_{t\wedge\tau_1}. \tag{4.28}$$

Furthermore, by [30, Eq. (5.31)], we have

$$\exp\left(2 \int_s^t \left|\langle M_{21,r}^{E_+} \rangle\right| \mathrm{d}r\right) \le \frac{\gamma_s^2}{\gamma_t^2}. \tag{4.29}$$

Finally, by (4.29), using Gronwall inequality and with (4.18) to estimate $Y_0 \le \alpha_{t\wedge\tau_1}$, we obtain $Y_{t\wedge\tau_1} \le n^{\xi}\alpha_{t\wedge\tau_1}$. We point out that here we also used (recall the definition of α_t from (4.23))

$$\int_0^t \alpha_s \frac{1}{\gamma_s} \frac{\gamma_s^2}{\gamma_t^2} \, \mathrm{d}s \lesssim \alpha_t, \tag{4.30}$$

which follows by simple explicit computations. Note that the exponent 2 in (4.30) is very important in the regime $\gamma_s \sim |\eta_{1,s}| + |\eta_{2,s}|$, in fact, in this regime, (4.30) would not be correct if (γ_s/γ_t) had any other exponent larger than 2. In the regime when $\gamma_s \sim |z_{1,s} - z_{2,s}|^2$, the exact power of (γ_s/γ_t) does not matter. This proves $\tau_1 = T^*$ with very high probability, concluding the proof. □

Proof of Proposition 4.4. Within this proof we use that the averaged local law from Proposition 4.3 is already proven. In fact the evolution of averaged quantities does not contain isotropic quantities, but the evolution of isotropic quantities contains both averaged and isotropic ones.

We now write the evolution of isotropic products of two and three resolvents $G_{i,t} = (W_t - \Lambda_{i,t})^{-1}$ along the flow (4.1). Let $A_1, A_2 \in \mathbb{C}^{2n\times 2n}$, then, by Itô's formula, we have the following equations. For products of two resolvents we have:

$$
\begin{aligned}
&\mathrm{d}(G_{1,t}A_1G_{2,t}-M_{12,t}^{A_1})_{\boldsymbol{xy}} \\
&\quad = \mathrm{d}\mathcal{E}_t + (G_{1,t}A_1G_{2,t}-M_{12,t}^{A_1})_{\boldsymbol{xy}}\mathrm{d}t + 2\sum_{i\neq j}\langle M_{12,t}^{A_1}E_i\rangle(G_{1,t}E_jG_{2,t}-M_{12,t}^{E_j})_{\boldsymbol{xy}}\mathrm{d}t \\
&\qquad + 2\sum_{i\neq j}(M_{12,t}^{E_j})_{\boldsymbol{xy}}\langle(G_{1,t}A_1G_{2,t}-M_{12,t}^{A_1})E_i\rangle\mathrm{d}t \\
&\qquad + 2\sum_{i\neq j}\langle(G_{1,t}A_1G_{2,t}-M_{12,t}^{A_1})E_i\rangle(G_{1,t}E_jG_{2,t}-M_{12,t}^{E_j})_{\boldsymbol{xy}}\mathrm{d}t \\
&\qquad + \langle G_{1,t}-m_{1,t}\rangle(G_{1,t}^2A_1G_{2,t})_{\boldsymbol{xy}}\mathrm{d}t + \langle G_{2,t}-m_{1,t}\rangle\mathrm{d}(G_{1,t}A_1G_{2,t}^2)_{\boldsymbol{xy}}\mathrm{d}t,
\end{aligned}
\tag{4.31}
$$

with

$$
\mathrm{d}\mathcal{E}_t := \frac{1}{\sqrt{n}}\sum_{a,b=1}^{2n}\partial_{ab}(G_{1,t}A_1G_{2,t})_{\boldsymbol{xy}}\mathrm{d}B_{ab,t}. \tag{4.32}
$$

Here we recall that ∂_{ab} denotes the directional derivative in the direction $(W_t)_{ab}$, $B_{ab,t}$ denotes the (a,b)–entry of the matrix Brownian motion B_t from (4.1), and $\sum_{i\neq j}$ denotes a summation over the pairs $(i,j)\in\{(1,2),(2,1)\}$. For products of three resolvents we have:

$$
\begin{aligned}
&\mathrm{d}(G_{1,t}A_1G_{2,t}A_2G_{1,t}-M_{121,t}^{A_1,A_2})_{\boldsymbol{xy}} \\
&\quad = \mathrm{d}\widehat{\mathcal{E}}_t + \frac{3}{2}(G_{1,t}A_1G_{2,t}A_2G_{1,t}-M_{121,t}^{A_1,A_2})_{\boldsymbol{xy}}\mathrm{d}t \\
&\qquad + (LT)+(EI)+(EII)+(EIII),
\end{aligned}
\tag{4.33}
$$

with

$$
\mathrm{d}\widehat{\mathcal{E}}_t := \frac{1}{\sqrt{n}}\sum_{a,b=1}^{2n}\partial_{ab}(G_{1,t}A_1G_{2,t}A_2G_{1,t})_{\boldsymbol{xy}}\mathrm{d}B_{ab,t}, \tag{4.34}
$$

and

$$
\begin{aligned}
(LT) :&= \sum_{i\neq j}\langle M_{21,t}^{A_2}E_i\rangle(G_{1,t}A_1G_{2,t}E_jG_{1,t}-M_{121,t}^{A_1,E_j})_{\boldsymbol{xy}}\mathrm{d}t \\
&\quad + \sum_{i\neq j}\langle M_{12,t}^{A_1}E_i\rangle(G_{1,t}E_jG_{2,t}A_2G_{1,t}-M_{121,t}^{E_j,A_2})_{\boldsymbol{xy}}\mathrm{d}t \\
(EI) :&= \sum_{i\neq j}\langle(G_{1,t}A_1G_{2,t}A_2G_{1,t}-M_{121,t}^{A_1,A_2})E_i\rangle\big[(M_{11,t}^{E_j})_{\boldsymbol{xy}}+(G_{1,t}E_jG_{1,t}-M_{11,t}^{E_j})_{\boldsymbol{xy}}\big]\mathrm{d}t \\
&\quad + \sum_{i\neq j}\langle M_{121,t}^{A_1,A_2}E_i\rangle(G_{1,t}E_jG_{1,t}-M_{11,t}^{E_j})_{\boldsymbol{xy}}\mathrm{d}t \\
(EII) :&= \sum_{i\neq j}\langle(G_{1,t}A_1G_{2,t}-M_{12,t}^{A_1})E_i\rangle\big[(M_{121,t}^{E_j,A_2})_{\boldsymbol{xy}}
\end{aligned}
$$

$$
\begin{aligned}
&+ (G_{1,t}E_jG_{2,t}A_2G_{1,t} - M_{121,t}^{E_j,A_2})_{\boldsymbol{xy}}\big]\mathrm{d}t \\
&+ \sum_{i\neq j}\langle (G_{2,t}A_2G_{1,t} - M_{21,t}^{A_2})E_i\rangle\big[(M_{121,t}^{A_1,E_j})_{\boldsymbol{xy}} \\
&+ (G_{1,t}A_1G_{2,t}E_jG_{1,t} - M_{121,t}^{A_1,E_j})_{\boldsymbol{xy}}\big]\mathrm{d}t \\
(EIII) :=\ & \langle G_{1,t}-m_{1,t}\rangle (G_{1,t}^2A_1G_{2,t}A_2G_{1,t})_{\boldsymbol{xy}}\mathrm{d}t + \langle G_{2,t}-m_{2,t}\rangle (G_{1,t}A_1G_{2,t}^2A_2G_{1,t})_{\boldsymbol{xy}}\mathrm{d}t \\
&+ \langle G_{1,t}-m_{1,t}\rangle (G_{1,t}A_1G_{2,t}A_2G_{1,t}^2)_{\boldsymbol{xy}}\mathrm{d}t. \qquad (4.35)
\end{aligned}
$$

We point out that in the derivation of (4.31)–(4.33) we also used the following lemma whose proof is postponed to Appendix A.

Lemma 4.6. *For any matrices $A_1, A_2 \in \mathbb{C}^{2N\times 2N}$ and vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{2N}$, we have*

$$
\begin{aligned}
\partial_t (M_{12,t}^{A_1})_{\boldsymbol{xy}} &= (M_{12,t}^{A_1})_{\boldsymbol{xy}} + (\mathcal{S}[M_{12,t}^{A_1}]M_{21,t}^{I})_{\boldsymbol{xy}}, \\
\partial_t (M_{121,t}^{A_1,A_2})_{\boldsymbol{xy}} &= (M_{121,t}^{A_1,A_2})_{\boldsymbol{xy}} + (\mathcal{S}[M_{121,t}^{A_1,A_2}]M_{11,t}^{I})_{\boldsymbol{xy}} + (\mathcal{S}[M_{12,t}^{A_1}]M_{121,t}^{I,A_2}))_{\boldsymbol{xy}} \\
&\quad + (M_{121,t}^{A_1,I})\mathcal{S}[M_{21,t}^{A_2}])_{\boldsymbol{xy}}.
\end{aligned} \tag{4.36}
$$

We now introduce the short–hand notations

$$
\begin{aligned}
Y_{12,t}(B_1) &:= (G_{1,t}B_1G_{2,t} - M_{12,t}^{B_1})_{\boldsymbol{xy}}, \\
Y_{121,t}(B_1,B_2) &:= (G_{1,t}B_1G_{2,t}B_2G_{1,t} - M_{121,t}^{B_1,B_2})_{\boldsymbol{xy}},
\end{aligned} \tag{4.37}
$$

where the vectors $\boldsymbol{x}, \boldsymbol{y}$ are fixed and omitted from the notation, and define the times

$$
T_i^\epsilon = T^\epsilon(\eta_{i,0}, z_{i,0}) := \min\{t > 0 : N\eta_{i,t}\rho_{i,t} = n^\epsilon\} \tag{4.38}
$$

such that $\eta_{i,t}\rho_{i,t}$ reaches the threshold $n\eta_{i,T_i^\epsilon}\rho_{i,T_i^\epsilon} = n^\epsilon$. For the deterministic approximations we have the bounds (see (3.20) and (3.22) of Lemma 3.3)

$$
\|M_{12,t}^{A_1}\| \lesssim \frac{\|A_1\|}{\gamma_t}, \qquad \|M_{121,t}^{A_1,A_2}\| \lesssim \frac{\|A_1\|\|A_2\|}{\eta_{*,t}\gamma_t}. \tag{4.39}
$$

From now on, for simplicity of the presentation, we assume that $b = 1$, which is anyway the most delicate regime, as we pointed out before the statement of Proposition 4.4. Define the stopping time

$$
\tau_2 := \inf\left\{ t \geq 0 : \sup_{\widetilde{\eta}_{i,0}}^{*} Y_{12,t} = \frac{n^{2\xi}}{\sqrt{n\widetilde{\gamma}_t\widetilde{\eta}_{*,t}}}, \qquad \sup_{\widetilde{\eta}_{i,0}}^{*} Y_{121,t} = \frac{n^{2\xi}}{\widetilde{\eta}_{*,t}\widetilde{\gamma}_t} \right\} \wedge T_1^\epsilon \wedge T_2^\epsilon, \tag{4.40}
$$

where

$$
Y_{12,t} := \max_{B_1\in\{A_1,E_+,E_-\}} \big|Y_{12,t}(B_1)\big|, \qquad Y_{121,t} := \max_{B_1,B_2\in\{A_1,A_2,E_+,E_-\}} \big|Y_{121,t}(B_1,B_2)\big|.
$$

Here $\widetilde{\gamma}_t := \gamma(\widetilde{\eta}_{1,t}, z_{1,t}, \widetilde{\eta}_{2,t}, z_{2,t})$, with $\widetilde{\eta}_{i,t}$ being the solution of (4.5) with initial condition $\widetilde{\eta}_{i,0}$, and $\widetilde{\ell}_t$, $\widetilde{\eta}_{*,t}$ are defined similarly. The supremum $\sup^*$ is taken over all the initial conditions $\widetilde{\eta}_{i,0}$ with $\widetilde{\eta}_{i,0} \in [\eta_{i,0}, \omega_*]$. We point out that in (4.40) we need to take the supremum over a larger set of initial conditions $\widetilde{\eta}_{i,0}$ only because at some point (see (4.60) below) we need to consider quantities involving $|G|$, and for $|G|$ we use an integral representation which requires integration over an entire segment of spectral parameters.

In the following, for $t \leq \tau_2$, we will often use the bound

$$\sup_{\eta_{i,t} \leq \check{\eta}_i \leq \omega_*/2} \left[\sqrt{n\check{\gamma}_t}\check{\eta}_{*,t} Y_{12,t} + \widehat{\eta}_{*,t}\check{\gamma}_t Y_{121,t} \right] \leq n^{2\xi}, \tag{4.41}$$

with very high probability. Here $\check{\gamma}_t := \gamma(\check{\eta}_1, z_{1,t}, \check{\eta}_2, z_{2,t})$, and $\check{\ell}_t$, $\check{\eta}_{*,t}$ are defined similarly. We point out that in (4.41) we also used that, as a consequence of (4.2), for any fixed $t \leq T_1^\epsilon \wedge T_2^\epsilon$ and any $\check{\eta}_i \in [\eta_{i,0}, \omega_*/2]$ there is a $\widetilde{\eta}_{i,0} \in [\eta_{i,0}, \omega_*]$ such that $\widetilde{\eta}_{i,t} = \check{\eta}_i$.

We now estimate the terms in the rhs. of (4.31)–(4.33) one by one. We first present the estimates for (4.31) and then for (4.33).

Estimates for (4.31)**:** We start computing the quadratic variation of the stochastic term

$$\begin{aligned} \mathrm{d}[\mathcal{E}_t, \mathcal{E}_t] = &\frac{1}{n\eta_{1,t}^2} (\operatorname{Im} G_{1,t})_{\boldsymbol{x}\boldsymbol{x}} (G_{2,t}^* A_1 \operatorname{Im} G_{1,t} A_1 G_{2,t})_{\boldsymbol{y}\boldsymbol{y}} \mathrm{d}t \\ &+ \frac{1}{n\eta_{2,t}^2} (G_{1,t} A_1 \operatorname{Im} G_{2,t} A_1 G_{1,t}^*)_{\boldsymbol{x}\boldsymbol{x}} (\operatorname{Im} G_{2,t})_{\boldsymbol{y}\boldsymbol{y}} \mathrm{d}t. \end{aligned} \tag{4.42}$$

Using (4.41) together with the second bound in (4.39), we estimate

$$\mathrm{d}[\mathcal{E}_t, \mathcal{E}_t] \lesssim \frac{n^{2\xi}\rho_{1,t}}{n\eta_{1,t}^2 \eta_{*,t}\gamma_t} + \frac{n^{2\xi}\rho_{2,t}}{n\eta_{2,t}^2 \eta_{*,t}\gamma_t} \mathrm{d}t. \tag{4.43}$$

We point out that in the following we will often use several properties of the characteristics from Lemma 4.1 even if not stated explicitly. By the Burkholder–Davis–Gundy inequality we thus obtain

$$\sup_{0 \leq r \leq t} \left| \int_0^{r \wedge \tau_2} \mathrm{d}\mathcal{E}_s \right| \lesssim n^{3\xi/2} \sqrt{\int_0^{t\wedge\tau_2} \frac{1}{n\eta_{*,s}\gamma_s} \left(\frac{\rho_{1,s}}{\eta_{1,s}^2} + \frac{\rho_{2,s}}{\eta_{2,s}^2} \right) \mathrm{d}s} \lesssim \frac{n^{3\xi/2}}{\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2}}, \tag{4.44}$$

where we used $\gamma_s \gtrsim \gamma_t$, for $s \leq t$, and the first relation in (4.8).

The second term in the rhs. of (4.31) can be neglected since, by considering the evolution of $e^{-t}(G_{1,t}A_1G_{2,t} - M_{12,t}^{A_1})_{\boldsymbol{x}\boldsymbol{y}}$, it amounts to a simple, negligible, rescaling of size $e^t \sim 1$. Next, by Proposition 4.3, we estimate the terms in the second line of (4.31) by

$$\begin{aligned}\left|\int_0^t (M_{12,s}^{E_j})_{\boldsymbol{xy}}\langle (G_{1,s}A_1G_{2,s}-M_{12,s}^{A_1})E_i\rangle \mathrm{d}s\right| &\lesssim \int_0^t \frac{1}{\gamma_s}\cdot\left(\frac{n^\xi}{\sqrt{n\ell_s}\gamma_s}\wedge\frac{n^\xi}{n\eta_{*,s}^2}\right)\mathrm{d}s \\ &\lesssim \frac{n^\xi}{\sqrt{n\ell_t}\gamma_t}\le\frac{1}{\sqrt{n\gamma_t}\eta_{*,t}},\end{aligned} \tag{4.45}$$

where in the last inequality we used that $\sqrt{\ell_t\gamma_t}\ge\eta_{*,t}$. In the second inequality we also used $\gamma_s\gtrsim\gamma_t$, for $s\le t$ and the first relation in (4.8). Similarly, using the definition of stopping time τ_2, we estimate

$$\begin{aligned}\left|\int_0^{t\wedge\tau_2}\langle (G_{1,s}A_1G_{2,s}-M_{12,s}^{A_1})E_i\rangle (G_{1,s}E_jG_{2,s}-M_{12,s}^{E_j})_{\boldsymbol{xy}}\mathrm{d}s\right| &\lesssim \frac{n^{2\xi}}{\sqrt{n\ell_s}}\cdot\frac{1}{\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2}} \\ &\le\frac{1}{\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2}},\end{aligned} \tag{4.46}$$

where in the last inequality we used that $\sqrt{n\ell_t}\ge n^{\epsilon/2}>n^{2\xi}$.

For the terms in the last line of (4.33), by spectral decomposition, for $t\le\tau$, we use

$$\left|(G_{1,t}^2A_1G_{2,t})_{\boldsymbol{xy}}\right|\le\frac{1}{\eta_{1,t}}(\operatorname{Im}G_{1,t})_{\boldsymbol{xx}}^{1/2}(G_{2,t}^*A_1\operatorname{Im}G_{1,t}A_1G_{2,t})_{\boldsymbol{yy}}^{1/2}\lesssim\frac{n^\xi\sqrt{\rho_{1,t}}}{\eta_{1,t}\sqrt{\eta_{*,t}\gamma_t}}. \tag{4.47}$$

We thus estimate those terms by

$$\left|\int_0^t\langle G_{1,s}-m_{1,s}\rangle (G_{1,s}^2A_1G_{2,s})_{\boldsymbol{xy}}\mathrm{d}s\right|\lesssim\int_0^t\frac{n^\xi}{n\eta_{1,s}}\cdot\frac{n^\xi\sqrt{\rho_{1,s}}}{\eta_{1,s}\sqrt{\eta_{*,s}\gamma_s}}\mathrm{d}s\lesssim\frac{n^\xi}{\sqrt{n\gamma_t}\eta_{*,t}}. \tag{4.48}$$

From (4.31), writing E_1,E_2 as in (4.24), combining (4.44)–(4.48), and using (4.27), we conclude

$$\begin{aligned}&(G_{1,t\wedge\tau_2}A_1G_{2,t\wedge\tau_2}-M_{12,t\wedge\tau_2}^{A_1})_{\boldsymbol{xy}} \\ &\quad=(G_{1,0}A_1G_{2,0}-M_{12,0}^{A_1})_{\boldsymbol{xy}}+\int_0^{t\wedge\tau_2}\langle M_{12,s}^{A_1}\rangle (G_{1,s}G_{2,s}-M_{12,s}^I)_{\boldsymbol{xy}}\,\mathrm{d}s \\ &\qquad+\mathcal{O}\left(\frac{n^{3\xi/2}}{\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2}}\right).\end{aligned} \tag{4.49}$$

From now on we assume that $A_1\in\{E_+,E_-\}$ without loss of generality, for presentational simplicity. In fact, if this is not the case then we can decompose

$$A_1=\langle A_1E_+\rangle E_++\langle A_1E_-\rangle E_-+A_c, \tag{4.50}$$

with A_c so that $\|M_{12,t}^{A_c}\| \lesssim 1$, uniformly in $t \geq 0$. This last inequality holds since the subspace corresponding to the smallest eigenvalues of the stability operator is spanned by E_-, E_+ (see e.g. Appendix A).

Finally, using (4.18) to estimate the initial condition in (4.49), we obtain

$$
\begin{aligned}
Y_{12,t\wedge\tau_2}(E_+) &= \int_0^{t\wedge\tau_2} \left[\langle M_{12,s}^{E_+}\rangle Y_{12,s}(E_+) + \langle M_{12,s}^{E_+}E_-\rangle Y_{12,s}(E_-)\right] \mathrm{d}s + \mathcal{O}\left(\frac{n^{3\xi/2}}{\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2}}\right), \\
Y_{12,t\wedge\tau_2}(E_-) &= \int_0^{t\wedge\tau_2} \left[\langle M_{12,s}^{E_-}\rangle Y_{12,s}(E_+) - \langle M_{12,s}^{E_-}E_-\rangle Y_{12,s}(E_-)\right] \mathrm{d}s + \mathcal{O}\left(\frac{n^{3\xi/2}}{\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2}}\right).
\end{aligned}
\tag{4.51}
$$

We are now ready to conclude the estimate for $Y_{12,t}$. In the following we omit some details and present only the main steps since this is very similar to the proof in [35, Eqs. (5.43)–(5.46) and Eqs. (5.58)–(5.59)]. First, we introduce the short–hand notations

$$
\mathcal{Y}_t := \begin{pmatrix} Y_{12,t\wedge\tau_2}(E_+) \\ Y_{12,t\wedge\tau_2}(E_-) \end{pmatrix}, \qquad \mathcal{M}_t := \begin{pmatrix} \langle M_{12,s}^{E_+}\rangle & \langle M_{12,s}^{E_+}E_-\rangle \\ \langle M_{12,s}^{E_-}\rangle & -\langle M_{12,s}^{E_-}E_-\rangle \end{pmatrix},
$$

and

$$
\mathcal{F}_t := \text{last three lines of (4.31)}.
$$

Then, from (4.31), combining (4.42)–(4.50), we can rewrite (4.51) in the more precise form

$$
\mathcal{Y}_{t\wedge\tau_2} = \int_0^{t\wedge\tau_2} \mathcal{M}_s\mathcal{Y}_s \,\mathrm{d}s + \int_0^{t\wedge\tau_2} \left[\mathcal{F}_s \,\mathrm{d}s + \mathrm{d}\mathcal{E}_s\right], \tag{4.52}
$$

with

$$
\left(\int_0^{t\wedge\tau_2} \|\mathcal{F}_s\| \,\mathrm{d}s\right)^2 + \int_0^{t\wedge\tau_2} \|\mathcal{C}_s\| \,\mathrm{d}s \leq \frac{n^{3\xi}}{n\gamma_{t\wedge\tau_2}\eta_{*,t\wedge\tau_2}^2}. \tag{4.53}
$$

Here $\mathcal{C}_s$ denotes the 2×2 covariance matrix of the martingale $\mathcal{E}_s$. Next, we recall that from [35, Eqs. (A.14)–(A.16)] it follows that $\langle M_{12,s}^{E_+}\rangle > -\langle M_{12,s}^{E_-}E_-\rangle$, for $\eta_{1,s}\eta_{2,s} < 0$ and that $\langle M_{12,s}^{E_\pm}E_\mp\rangle$ is purely imaginary.

We are now ready to conclude the proof by an integral Gronwall inequality as in [35, Lemma 5.6]. Define

$$
\alpha_t := \frac{1}{\sqrt{n\gamma_t}\eta_{*,t}}, \qquad f_t := \langle M_{12,t}^{E_+}\rangle. \tag{4.54}
$$

Then, by [35, Eqs. (A.14)–(A.16)], it is easy to see that (4.52) satisfies the hypothesis of [35, Lemma 5.6]. We can thus apply a matrix Gronwall inequality from [35, Lemma 5.6] and obtain $Y_{12,t\wedge\tau_2} = \|\mathcal{Y}_{t\wedge\tau_2}\| \lesssim n^{3\xi/2}/(\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2})$. We point out that in this last inequality we used

$$\exp\left(\int_s^t f_t \,\mathrm{d}r\right) \lesssim \frac{\eta_{*,s}}{\eta_{*,t}}, \tag{4.55}$$

which follows from $f_t \lesssim \sqrt{\rho_{1,r}\rho_{2,r}/(\eta_{1,r}\eta_{2,r})}$ and $\sqrt{\eta_{1,s}\eta_{2,s}/(\eta_{1,t}\eta_{2,t})} \sim \eta_{*,s}/\eta_{*,t}$ (for this second relation, see the similar proof [30, Eq. (5.32)]), and

$$\int_0^t \frac{1}{\sqrt{n\gamma_s}\eta_{*,s}} \cdot \frac{1}{\gamma_s} \cdot \frac{\eta_{*,s}}{\eta_{*,t}} \,\mathrm{d}s \lesssim \frac{1}{\sqrt{n\gamma_t}\eta_{*,t}}, \tag{4.56}$$

where we used that $\int(1/\gamma_s) \lesssim 1$. Note that in the averaged case we used the more involved estimate (4.29), instead here we used the weaker and simply obtainable bound (4.55) which is enough for the proof of the isotropic law. We conclude the estimate of the isotropic estimate of the product of two resolvents by pointing out that a similar proof goes through for the case $b \in [0,1)$, with the difference that (4.56) can be replaced with

$$\int_0^t \frac{(\rho_s)^*}{\sqrt{n}\eta_{*,s}^{(3-b)/2}\gamma_s^{b/2}} \cdot \frac{1}{\gamma_s} \cdot \frac{\eta_{*,s}}{\eta_{*,t}} \,\mathrm{d}s \lesssim \frac{(\rho_t)^*}{\sqrt{n}\eta_{*,t}^{3/2-b}\gamma_t^{b/2}}. \tag{4.57}$$

A similar comment applies to isotropic chains of three resolvents below, even if we will not point it out explicitly.

Estimates for (4.33): Similarly to the two resolvent case above, we notice that the second term in the rhs. of (4.33) can be neglected from the analysis, since it only amounts to a simple rescaling of size $e^{3t/2} \sim 1$. Several of the following estimates are similar to the estimates that we preformed in the rhs. of (4.31), for this reason we omit some easily checkable details. In particular, we point out that throughout we will use properties of the characteristics from Lemma 4.1 even if not stated explicitly.

We now compute the quadratic variation of the stochastic term in (4.33):

$$\begin{aligned} \mathrm{d}[\widehat{\mathcal{E}}_t, \widehat{\mathcal{E}}_t] = \frac{1}{n\eta_{1,t}^2}\Big[&(\mathrm{Im}\, G_{1,t})_{\boldsymbol{xx}}(G_{1,t}^* A_2 G_{2,t}^* A_1 \mathrm{Im}\, G_{1,t} A_1 G_{2,t} A_2 G_{1,t})_{\boldsymbol{yy}} \\ &+ (G_{1,t} A_1 G_{2,t} A_2 \mathrm{Im}\, G_{1,t} A_2 G_{2,t}^* A_1 G_{1,t}^*)_{\boldsymbol{xx}} (\mathrm{Im}\, G_{1,t})_{\boldsymbol{yy}}\Big] \mathrm{d}t \\ &+ \frac{1}{n\eta_{2,t}^2}(G_{1,t} A_1 \mathrm{Im}\, G_{2,t} A_1 G_{1,t}^*)_{\boldsymbol{xx}} (G_{1,t}^* A_2 \mathrm{Im}\, G_{2,t} A_2 G_{1,t})_{\boldsymbol{yy}} \mathrm{d}t. \end{aligned} \tag{4.58}$$

Note that (the first two lines of) the quadratic variation process in (4.58) contains chains with more than three resolvents, we thus rely on the following *reduction inequalities* to shorten these chains.

Lemma 4.7 *(Reduction inequalities). For any $Q_1, Q_2, Q_3, R_1, R_2 \in \mathbb{C}^{2n\times 2n}$ it holds*

$$\begin{aligned} &\big|(Q_1 G(w_1) Q_2 G(w_2) Q_3)_{\boldsymbol{x}\boldsymbol{y}}\big| \\ &\quad\lesssim \sqrt{N}(Q_1|G(w_1)|Q_1^*)^{1/2}_{\boldsymbol{x}\boldsymbol{x}}(Q_3|G(w_2)|Q_3^*)^{1/2}_{\boldsymbol{x}\boldsymbol{x}}\langle |G(w_1)|Q_2|G(w_2)|Q_2^*\rangle^{1/2} \\ &\langle |G(w_1)|R_1 G(w_2) R_2 |G(w_1)| R_2^* G(w_2)^* R_1^*\rangle \\ &\quad\lesssim N\langle |G(w_1)|R_1|G(w_2)|R_1^*\rangle\langle |G(w_1)|R_2^*|G(w_2)|R_2\rangle, \end{aligned} \tag{4.59}$$

Note that the reduction inequalities above involve $|G|$ as well. For this reason in the following we need to show that if a bound holds for certain products of resolvent then the same bound also holds if one (or more) G is replaced by $|G|$ (see (4.61) below).

To estimate (4.58), for $t \le \tau_2$, we use (4.59) in the form

$$(G_{1,t}^* A_2 G_{2,t}^* A_1 \operatorname{Im} G_{1,t} A_1 G_{2,t} A_2 G_{1,t})_{\boldsymbol{y}\boldsymbol{y}} \lesssim n (G_{1,t}^*|G_{2,t}|G_{1,t})_{\boldsymbol{y}\boldsymbol{y}}\langle \operatorname{Im} G_{1,t}|G_{2,t}|\rangle \lesssim \frac{n^{1+2\xi}}{\eta_{*,t}\gamma_t^2}. \tag{4.60}$$

We point out that in the last inequality of (4.60) we used that for $t \le \tau_2$ we have

$$(G_{1,t}^*|G_{2,t}|G_{1,t})_{\boldsymbol{y}\boldsymbol{y}} \lesssim \frac{n^{2\xi}\log n}{\eta_{*,t}\gamma_t}, \qquad \langle \operatorname{Im} G_{1,t}|G_{2,t}|\rangle \lesssim \frac{\log n}{\gamma_t}. \tag{4.61}$$

This follows by [35, Appendix A] (see also [26, Section 5]) and the fact that the same bound holds for the quantities with $|G|$ replaced with G (see e.g. [35, Proof of Eq. (5.32) in Appendix A] and [32, Lemma 4.10]). By (4.58) and the BDG inequality we thus obtain

$$\sup_{0\le r\le t}\left|\int_0^{r\wedge\tau_2} \mathrm{d}\widehat{\mathcal{E}}_s\right| \lesssim n^{3\xi/2}\sqrt{\int_0^{t\wedge\tau_2}\frac{1}{\eta_{*,s}\gamma_s^2}\left(\frac{\rho_{1,s}}{\eta_{1,s}^2}+\frac{\rho_{2,s}}{\eta_{2,s}^2}\right)\mathrm{d}s} \lesssim \frac{n^{3\xi/2}}{\eta_{*,t\wedge\tau_2}\gamma_{t\wedge\tau_2}}, \tag{4.62}$$

with very high probability, where we used that the third line of (4.58), by the definition of τ_2, is bounded by $n^{4\xi}/(n\eta_{*,t}^4\gamma_t^2) \le 1/(\eta_{*,t}^3\gamma_t^2)$ (recall that $\xi \le \epsilon/10$).

We now explain how to estimate the remaining terms in (4.33). By spectral decomposition, Proposition 4.3, and (4.39), we have

$$\begin{aligned} \big|\langle G_{1,t}A_1 G_{2,t} A_2 G_{1,t}\rangle\big| + \big|\langle M_{121,t}^{A_1,A_2}\rangle\big| &\lesssim \frac{1}{\eta_{1,t}}\langle \operatorname{Im} G_{1,t}A_1|G_{2,t}|A_1^*\rangle^{1/2}\langle \operatorname{Im} G_{1,t}A_2|G_{2,t}|A_2^*\rangle^{1/2} \\ &\quad + \big|\langle M_{121,t}^{A_1,A_2}\rangle\big| \lesssim \frac{1}{\eta_{*,t}\gamma_t}, \end{aligned}$$

with very high probability. We can thus estimate the terms in (EI) of (4.33) by

$$\int_0^{t\wedge\tau_2} \frac{1}{\eta_{*,s}\gamma_s}\cdot\frac{\rho_{1,s}}{\eta_{1,s}}\,\mathrm{d}s \lesssim \frac{1}{\eta_{*,t\wedge\tau_2}\gamma_{t\wedge\tau_2}}, \tag{4.63}$$

where we also used that $\|M_{11,t}^A\| \lesssim (\rho_{1,t}/\eta_{1,t})\|A\|$.

Then, by Proposition 4.3, we estimate the terms in (EII) of (4.33) by

$$\int_0^{t\wedge\tau_2} \frac{1}{\sqrt{n\eta_{*,s}}\gamma_s}\cdot\frac{n^{2\xi}}{\eta_{*,s}\gamma_s}\,\mathrm{d}s \lesssim \frac{n^{2\xi}}{\sqrt{n\eta_{*,t\wedge\tau_2}}}\cdot\frac{1}{\eta_{*,t\wedge\tau_2}\gamma_{t\wedge\tau_2}}. \tag{4.64}$$

We now consider the terms in (EIII) of (4.33). For these terms, by (4.59), we estimate (we just consider one of those term for concreteness)

$$\begin{aligned}\left|(G_{1,t}^2A_1G_{2,t}A_2G_{1,t})_{\boldsymbol{xy}}\right| &\lesssim \frac{1}{\eta_{1,t}}(\operatorname{Im}G_{1,t})_{\boldsymbol{xx}}^{1/2}(G_{1,t}^*A_2G_{2,t}^*A_1\operatorname{Im}G_{1,t}A_1G_{2,t}A_2G_{1,t})_{\boldsymbol{yy}}^{1/2}\\ &\lesssim \frac{n^{1/2+\xi}\sqrt{\rho_{1,t}}}{\eta_{1,t}\eta_{*,t}\gamma_t},\end{aligned}$$

where in the last inequality we used (4.60). We can thus estimate the terms in (EIII) of (4.33) by

$$\int_0^{t} \frac{1}{n\eta_{*,s}}\cdot\frac{n^{1/2+\xi}\sqrt{\rho_{1,t}}}{\eta_{1,t}\eta_{*,t}\gamma_t}\,\mathrm{d}s \lesssim \frac{1}{\eta_{*,t}\gamma_t}. \tag{4.65}$$

From now on, without loss of generality, for simplicity we assume that $A_1, A_2 \in \{E_+, E_-\}$ (see (4.50)).

Combining (4.62)–(4.65) and using (4.18) to estimate the initial condition, from (4.33), using the short–hand notation $\Upsilon_t^{\pm,\pm} := Y_{121,t}(E_\pm, E_\pm)$, we obtain

$$\begin{aligned}\Upsilon_{t\wedge\tau}^{+,+} &= \int_0^{t\wedge\tau_2}\left[2\langle M_{12,s}^{E_+}\rangle\Upsilon_s^{+,+} - \langle M_{21,s}^{E_+}E_-\rangle\Upsilon_s^{+,-} - \langle M_{12,s}^{E_+}E_-\rangle\Upsilon_s^{-,+}\right]\mathrm{d}s + \mathcal{O}\left(\frac{n^{3\xi/2}}{\eta_{*,t}\gamma_t}\right),\\ \Upsilon_{t\wedge\tau}^{+,-} &= \int_0^{t\wedge\tau_2}\left[\left(\langle M_{12,s}^{E_+}\rangle - \langle M_{21,s}^{E_-}E_-\rangle\right)\Upsilon_s^{+,-} + \langle M_{21,s}^{E_-}\rangle\Upsilon_s^{+,+} - \langle M_{12,s}^{E_+}E_-\rangle\Upsilon_s^{-,-}\right]\mathrm{d}s\\ &\quad + \mathcal{O}\left(\frac{n^{3\xi/2}}{\eta_{*,t}\gamma_t}\right),\\ \Upsilon_{t\wedge\tau}^{-,+} &= \int_0^{t\wedge\tau_2}\left[\left(\langle M_{21,s}^{E_+}\rangle - \langle M_{12,s}^{E_-}E_-\rangle\right)\Upsilon_s^{-,+} + \langle M_{12,s}^{E_-}\rangle\Upsilon_s^{+,+} - \langle M_{21,s}^{E_+}E_-\rangle\Upsilon_s^{-,-}\right]\mathrm{d}s\\ &\quad + \mathcal{O}\left(\frac{n^{3\xi/2}}{\eta_{*,t}\gamma_t}\right),\end{aligned}$$

$$\Upsilon^{-,-}_{t\wedge\tau} = \int_0^{t\wedge\tau_2} \big[-2\langle M^{E_-}_{12,s}E_-\rangle\Upsilon^{-,-}_s + \langle M^{E_-}_{12,s}E_+\rangle\Upsilon^{+,-}_s + \langle M^{E_-}_{21,s}E_+\rangle\Upsilon^{-,+}_s\big]\,\mathrm{d}s$$
$$+\mathcal{O}\left(\frac{n^{3\xi/2}}{\eta_{*,t}\gamma_t}\right). \tag{4.66}$$

We are now ready to conclude the estimate of $Y_{121,t}$. We present only the main steps and omit some details, since this proof is analogous to the one presented in (4.51)–(4.56) and to the one in [35, Eqs. (5.43)–(5.46) and Eqs. (5.58)–(5.59)]. From (4.33), combining (4.58)–(4.65), in order to conclude the proof we can now proceed as in (4.51)–(4.56) and apply [35, Lemma 5.6] to a linear system of the form

$$\widehat{\mathcal{Y}}_{t\wedge\tau_2} = \int_0^{t\wedge\tau_2} \widehat{\mathcal{M}}_s\mathcal{Y}_s\,\mathrm{d}s + \int_0^{t\wedge\tau_2}\big[\widehat{\mathcal{F}}_s\,\mathrm{d}s + \mathrm{d}\widehat{\mathcal{E}}_s\big], \tag{4.67}$$

with

$$\left(\int_0^{t\wedge\tau_2}\|\widehat{\mathcal{F}}_s\|\,\mathrm{d}s\right)^2 + \int_0^{t\wedge\tau_2}\|\widehat{\mathcal{C}}_s\|\,\mathrm{d}s \le \frac{n^{3\xi}}{n\gamma_{t\wedge\tau_2}\eta^2_{*,t\wedge\tau_2}} =: \alpha_{t\wedge\tau_2}, \tag{4.68}$$

where $\widehat{\mathcal{C}}_t$ denotes the 4×4 covariance matrix of the martingale $\widehat{\mathcal{E}}_t$. Here we also used the notations

$$\widehat{\mathcal{Y}}_t := \begin{pmatrix}\Upsilon^{+,+}_t\\ \Upsilon^{+,-}_t\\ \Upsilon^{-,+}_t\\ \Upsilon^{-,-}_t\end{pmatrix}, \qquad \widehat{\mathcal{M}}_t := \begin{pmatrix} 2a_{+,t} & ib_t & -ib_t & 0\\ ib_t & a_{+,t}-a_{-,t} & 0 & -ib_t\\ -ib_t & 0 & a_{+,t}-a_{-,t} & ib_t\\ 0 & -ib_t & ib_t & 2a_{-,t},\end{pmatrix} \tag{4.69}$$

where

$$a_{+,t} := \langle M^{E_+}_{12,t}\rangle, \qquad a_{-,t} := \langle M^{E_-}_{12,t}E_-\rangle,$$

and b_t is real and defined in [35, Eq. (A.14)]. Note that the matrix $\widehat{\mathcal{M}}_t$ is exactly the same as the one [35, Eq. (5.44)], and that $\alpha_{t\wedge\tau_2}$ defined in (4.68) is different compared to the one defined in (4.54). We can thus use again the matrix Gronwall inequality from [35, Lemma 5.6] for α_t defined in (4.68) and $f_t := 2\langle M^{E_+}_{12,t}\rangle$ (note that here we redefined f_t, i.e. there is an additional factor of 2 compared to (4.54)). By (4.55), we thus conclude

$$Y_{121,t\wedge\tau_2} = \|\mathcal{Y}_{t\wedge\tau_2}\| \lesssim \frac{n^{3\xi/2}\log n}{\eta_{*,t\wedge\tau_2}\gamma_{t\wedge\tau_2}}.$$

We point out that to obtain this inequality we also used

$$\int_0^t \frac{1}{\eta_{*,s}\gamma_s} \cdot \frac{1}{\gamma_s} \cdot \frac{\eta_{*,s}^2}{\eta_{*,t}^2}\, \mathrm{d}s \lesssim \frac{\log n}{\eta_{*,t}\gamma_t}.$$

Conclusion: In the previous two parts we proved that

$$Y_{12,t\wedge\tau_2} \lesssim \frac{n^\xi}{\sqrt{n\gamma_{t\wedge\tau_2}}\eta_{*,t\wedge\tau_2}}, \qquad Y_{121,t\wedge\tau_2} \lesssim \frac{n^{3\xi/2}\log n}{\eta_{*,t\wedge\tau_2}\gamma_{t\wedge\tau_2}}. \tag{4.70}$$

These two bounds together prove the desired estimate for the initial conditions $\eta_{i,0}$. A completely analogous proof holds if we consider any other initial condition $\widetilde{\eta}_{i,0} \in [\eta_{i,0}, \omega_*]$, obtaining exactly the same bounds (4.70) (with tilde on all quantities, i.e. $\gamma \to \widetilde{\gamma}$, etc...). This, together with the definition of τ_2 in (4.40) shows that $\tau_2 = T_1^\epsilon \wedge T_2^\epsilon$ with very high probability. □

4.1. Improved local law at the edge of the spectrum

In this section we present the proof of the averaged local law at the edge of the spectrum, including $\operatorname{Im} G$'s. While in the bulk of the spectrum there is no difference between G and $\operatorname{Im} G$, we will see that close to the edge there will be an improvement (proportional to the local density) if we replace one (or more) G by $\operatorname{Im} G$. For brevity, we present this proof only in the averaged case, since this is what gives the optimal bound on the eigenvector overlaps in Theorem 3.6. The result from Proposition 4.8 below is the main technical input to prove Theorem 3.5 (see the beginning of Section 4 for a related explanation).

Proposition 4.8. *Fix small n–independent constants $\tau, \omega_*, \epsilon > 0$, and for $i = 1, 2$ fix spectral parameters $\Lambda_{i,0}$ as in (4.12), with $|z_{i,0}| \leq 10$, $|\eta_{i,0}| \leq \omega_*$. For $t \leq T^*$ let $\Lambda_{i,t}$ be the solution of (4.4) with initial condition $\Lambda_{i,0}$. Let $G_{i,t} := (W_t - \Lambda_{i,t})^{-1}$, and let $\widehat{M}_{12,t}^A$ be the deterministic approximation of $\operatorname{Im} G_{1,t} A \operatorname{Im} G_{2,t}$ from (3.27). Let ℓ from (4.13), and define the control parameter*

$$\begin{aligned} \widehat{\gamma} = \widehat{\gamma}(\eta_1, z_1, \eta_2, z_2) &:= |z_1 - z_2|^2 + \rho_1|\eta_1| + \rho_2|\eta_2| + \left(\frac{\eta_1}{\rho_1}\right)^2 + \left(\frac{\eta_2}{\rho_2}\right)^2, \\ \widehat{\gamma}_t &:= \widehat{\gamma}(\eta_{1,t}, z_{1,t}, \eta_{2,t}, z_{2,t}). \end{aligned} \tag{4.71}$$

Assume that for some small $0 < \xi \leq \epsilon/10$, with very high probability, it holds

$$\left|\left\langle \left(\operatorname{Im} G_{1,0} A_1 \operatorname{Im} G_{2,0} - \widehat{M}_{12,0}^{A_1}\right) A_2 \right\rangle\right| \leq \frac{n^\xi \rho_{1,0}\rho_{2,0}}{\sqrt{n\ell_0 \widehat{\gamma}_0}} \wedge \frac{n^\xi}{n|\eta_{1,0}\eta_{2,0}|} \tag{4.72}$$

uniformly in $\|A_i\| \leq 1$ and for any $|\eta_i| \in [|\eta_{i,0}|, n^{100}]$. Then,

$$\left|\left\langle (\operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} - \widehat{M}_{12,t}^{A_1}) A_2 \right\rangle\right| \leq \frac{n^{2\xi}\rho_{1,t}\rho_{2,t}}{\sqrt{n\ell_t\widehat{\gamma}_t}} \wedge \frac{n^{2\xi}}{n|\eta_{1,t}\eta_{2,t}|}, \tag{4.73}$$

with very high probability uniformly in $\|A_i\| \leq 1$ *and all* $t \leq T^*$ *such that* $n\ell_t \geq n^\epsilon$.

Proof. To keep the presentation simple, we only consider the case $A_1, A_2 \in \{E_+, E_-\}$. By (4.20) we have

$$
\begin{aligned}
&\mathrm{d}\langle (\operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} - M_{12,t}^{A_1}) A_2 \rangle \\
&= \sum_{a,b=1}^{2n} \partial_{ab} \langle \operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} A_2 \rangle \frac{\mathrm{d}B_{ab,t}}{\sqrt{n}} + \langle (\operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} - \widehat{M}_{12,t}^{A_1}) A_2 \rangle \mathrm{d}t \\
&\quad + (LT) + (EI) + (EII) + (EIII),
\end{aligned}
\tag{4.74}
$$

where we defined

$$
\begin{aligned}
(LT) :=\, & 2 \sum_{i \neq j} \langle (G_{1,t}^* A_1 G_{2,t} - M_{\overline{1}2,t}^{A_1}) E_i \rangle \langle \widehat{M}_{21,t}^{A_2} E_j \rangle \mathrm{d}t \\
& + 2 \sum_{i \neq j} \langle (\operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} - \widehat{M}_{12,t}^{A_1}) E_i \rangle \langle M_{\overline{2}1,t}^{A_2} E_j \rangle \mathrm{d}t \\
& - 4 \sum_{i \neq j} \langle \widehat{M}_{12,t}^{A_1} E_i \rangle \langle (\operatorname{Im} G_{2,t} A_2 \operatorname{Im} G_{1,t} - \widehat{M}_{21,t}^{A_2} \rangle) E_j \rangle \mathrm{d}t \\
& - 4 \sum_{i \neq j} \langle (\operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} - \widehat{M}_{12,t}^{A_1} \rangle) E_i \rangle \langle \widehat{M}_{21,t}^{A_2} E_j \rangle \mathrm{d}t \\
(EI) :=\, & 2 \sum_{i \neq j} \langle M_{\overline{1}2,t}^{A_1} E_i \rangle \langle (\operatorname{Im} G_{2,t} A_2 \operatorname{Im} G_{1,t} - \widehat{M}_{21,t}^{A_2}) E_j \rangle \mathrm{d}t \\
& + 2 \sum_{i \neq j} \langle (\operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} - \widehat{M}_{12,t}^{A_1}) E_i \rangle \langle (G_{2,t}^* A_2 G_{1,t} - M_{\overline{2}1,t}^{A_2}) E_j \rangle \mathrm{d}t \\
(EII) :=\, & 2 \sum_{i \neq j} \langle (G_{1,t}^* A_1 G_{2,t} - M_{\overline{1}2,t}^{A_1}) E_i \rangle \langle (\operatorname{Im} G_{2,t} A_2 \operatorname{Im} G_{1,t} - \widehat{M}_{21,t}^{A_2}) E_j \rangle \mathrm{d}t \\
& + 2 \sum_{i \neq j} \langle \widehat{M}_{12,t}^{A_1} E_i \rangle \langle (G_{2,t}^* A_2 G_{1,t} - \widehat{M}_{\overline{2}1,t}^{A_2}) E_j \rangle \mathrm{d}t \\
& - 4 \sum_{i \neq j} \langle (\operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} - \widehat{M}_{12,t}^{A_1} \rangle) E_i \rangle \langle (\operatorname{Im} G_{2,t} A_2 \operatorname{Im} G_{1,t} - \widehat{M}_{21,t}^{A_2} \rangle) E_j \rangle \mathrm{d}t \\
(EIII) :=\, & \langle (G_{1,t} - M_{1,t}) \rangle \langle \operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} A_2 G_{1,t} \rangle \mathrm{d}t \\
& + \langle (G_{2,t} - M_{2,t}) \rangle \langle \operatorname{Im} G_{2,t} A_2 \operatorname{Im} G_{1,t} A_1 G_{2,t} \rangle \mathrm{d}t \\
& + \langle (G_{1,t} - M_{1,t})^* \rangle \langle \operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} A_2 G_{1,t}^* \rangle \mathrm{d}t \\
& + \langle (G_{2,t} - M_{2,t})^* \rangle \langle \operatorname{Im} G_{2,t} A_2 \operatorname{Im} G_{1,t} A_1 G_{2,t}^* \rangle \mathrm{d}t \\
& + \left(\frac{\langle \operatorname{Im} G_{1,t} - \operatorname{Im} M_{1,t} \rangle}{\eta_{1,t}} + \frac{\langle \operatorname{Im} G_{2,t} - \operatorname{Im} M_{2,t} \rangle}{\eta_{2,t}} \right) \langle \operatorname{Im} G_{1,t} A_1 \operatorname{Im} G_{2,t} A_2 \rangle.
\end{aligned}
\tag{4.75}
$$

We point out that here we used that by the chiral symmetry of H^z for any diagonal B_1, B_2 we have $\langle G_1 B_1 \mathrm{Im}\, G_2 B_2\rangle = \mathrm{i}\langle \mathrm{Im}\, G_1 B_1 \mathrm{Im}\, G_2 B_2\rangle$.

We now present the estimate of all the various terms in the rhs. of (4.74). These estimates are very similar to the ones presented in [35, Section 5.1], hence we omit some details. Throughout the proof we use that $\widehat{\gamma}_s \gtrsim \widehat{\gamma}_t$, $\gamma_s \gtrsim \gamma_t$, for $s \le t$, and (4.8) even if not mentioned explicitly. For simplicity of notation we also assume that $\eta_{i,t} > 0$.

Define the stopping time

$$\tau_3 := \inf\left\{ t \ge 0 : \max_{A_1, A_2 \in \{E_+, E_-\}} \left|\langle (\mathrm{Im}\, G_{1,t} A_1 \mathrm{Im}\, G_{2,t} - \widehat{M}^{A_1}_{12,t}) A_2\rangle\right| = n^{2\xi}\widehat{\alpha}_t \right\} \wedge T^*,$$
$$\widehat{\alpha}_t := \frac{1}{n|\eta_{1,t}\eta_{2,t}|} \wedge \frac{\rho_{1,t}\rho_{2,t}}{\sqrt{n\ell_t}\widehat{\gamma}_t}. \tag{4.76}$$

Recall (see (3.21) of Lemma 3.3) that for the deterministic term we have the bound

$$\|\widehat{M}^A_{12,t}\| \lesssim \frac{1}{\widehat{\gamma}_t}. \tag{4.77}$$

We start with the estimate of the stochastic term in (4.74). Its quadratic variation process is bounded by

$$\frac{1}{n^2\eta_{*,t}^2}\langle \mathrm{Im}\, G_{1,t} A_1 \mathrm{Im}\, G_{2,t} A_2 \mathrm{Im}\, G_{1,t} A_2 \mathrm{Im}\, G_{2,t} A_1\rangle \mathrm{d}t \lesssim \left(\frac{\rho_{1,t}}{n^2\eta_{1,t}^3\eta_{2,t}^2} \wedge \frac{\rho_{1,t}^2\rho_{2,t}^2}{n\eta_{*,t}^2\widehat{\gamma}_t^2}\right) \mathrm{d}t, \tag{4.78}$$

where to obtain the second term in the rhs. we used the averaged reduction inequality (4.59) to estimate the trace of the product of four G's in terms of the product of two traces involving two G's each. To estimate the traces involving two resolvents we then used (4.77) and the definition of the stopping time τ_3 in (4.76). By the BDG inequality we thus obtain

$$\sup_{0\le r\le t}\left| \int_0^{r\wedge\tau_3} \sum_{a,b=1}^{2n} \partial_{ab}\langle \mathrm{Im}\, G_{1,s} A_1 \mathrm{Im}\, G_{2,s} A_2\rangle \frac{\mathrm{d}B_{ab,s}}{\sqrt{n}}\right| \lesssim n^{\xi}\widehat{\alpha}_{t\wedge\tau_3}. \tag{4.79}$$

Notice that the second term in the second line of (4.74) can be neglected, since it just amounts to a simple rescaling of a factor $e^t \sim 1$. Next, using that by spectral decomposition we have

$$\left|\langle \mathrm{Im}\, G_{1,t} A_1 \mathrm{Im}\, G_{2,t} A_2 G_{1,t}\rangle\right| \lesssim \frac{1}{\eta_{1,t}}\langle \mathrm{Im}\, G_{1,t} A_1 \mathrm{Im}\, G_{2,t} A_1^*\rangle^{1/2}\langle \mathrm{Im}\, G_{1,t} A_2 \mathrm{Im}\, G_{2,t} A_2^*\rangle^{1/2}, \tag{4.80}$$

and using the single resolvent local law $|\langle G_{i,t} - M_{i,t}\rangle| \le n^{\xi}/(n\eta_{i,t})$, we estimate the term in $(EIII)$ by

$$\int_0^{t\wedge\tau_3} \frac{n^\xi}{n\eta_{*,s}^2}\frac{\rho_{1,s}\rho_{2,s}}{\widehat{\gamma}_s}\,\mathrm{d}s \lesssim \frac{n^\xi}{n\ell_{t\wedge\tau_3}}\cdot\frac{\rho_{1,t\wedge\tau_3}\rho_{2,t\wedge\tau_3}}{\widehat{\gamma}_t} \le \widehat{\alpha}_{t\wedge\tau_3}, \tag{4.81}$$

where in the last inequality we also used that $\xi \le \epsilon/10$.

Next, using the definition of the stopping time τ_3 in (4.76) and the local law

$$\left|\langle(G_{1,t}A_1G_{2,t} - M_{12,t}^{A_1})A_2\rangle\right| \lesssim \frac{n^\xi}{n\eta_{1,t}\eta_{2,t}} \wedge \frac{n^\xi}{\sqrt{n\ell_t}\gamma_t}, \tag{4.82}$$

we estimate the terms in (EII) by

$$\int_0^{t\wedge\tau_3}\left(\frac{n^\xi}{n\eta_{1,s}\eta_{2,s}}\wedge\frac{n^\xi}{\sqrt{n\ell_s}\gamma_s}\right)\left(\frac{n^\xi}{n\eta_{1,s}\eta_{2,s}}\wedge\frac{n^\xi\rho_{1,s}\rho_{2,s}}{\sqrt{n\ell_s}\widehat{\gamma}_s}\right)\mathrm{d}s \lesssim \frac{n^{2\xi}}{n\ell_{t\wedge\tau_3}}\widehat{\alpha}_{t\wedge\tau_3}. \tag{4.83}$$

Furthermore, using (4.77) and (4.82), the terms in (EI) are estimated by

$$\int_0^t \frac{\rho_{1,s}\rho_{2,s}}{\widehat{\gamma}_s}\cdot\left(\frac{n^\xi}{n\eta_{1,s}\eta_{2,s}}\wedge\frac{n^\xi}{\sqrt{n\ell_s}\gamma_s}\right)\mathrm{d}s \lesssim n^\xi\widehat{\alpha}_t. \tag{4.84}$$

We are thus left only with the linear terms (LT) in the rhs. of (4.74). These terms will be Gronwalled, giving the desired bound. This last step is the only place where there is a relevant difference compared to [35, Section 5]. In fact, in [35] when we considered G_1G_2, to deal with the two bad directions for the stability operator E_-, E_+, we studied a 4×4 system of equations, corresponding to the four choices $A_1, A_2 \in \{E_+, E_-\}$. Instead, for $\operatorname{Im} G_1 \operatorname{Im} G_2$, due to additional symmetries of $\operatorname{Im} G$ (see (4.85)) it is enough to study a scalar equation. Indeed, by the chiral symmetry of H^z we have

$$\begin{aligned}\langle \operatorname{Im} G_{1,t}E_-\operatorname{Im} G_{2,t}E_-\rangle &= \langle \operatorname{Im} G_{1,t}\operatorname{Im} G_{2,t}\rangle,\\ \langle \operatorname{Im} G_{1,t}E_-\operatorname{Im} G_{2,t}\rangle &= \langle \operatorname{Im} G_{1,t}\operatorname{Im} G_{2,t}E_-\rangle, \quad \langle \widehat{M}_{12,t}^{E_\pm}E_\mp\rangle = 0.\end{aligned} \tag{4.85}$$

This shows that it is enough to only consider two cases: $(A_1, A_2) = (E_+, E_+)$ and $(A_1, A_2) = (E_+, E_-)$. We first consider the case $(A_1, A_2) = (E_+, E_+)$. By inspecting (4.74), one can easily see that the equation for $(A_1, A_2) = (E_+, E_+)$ is self–contained; in fact, the terms containing (E_+, E_-), which naturally emerge in the rhs. of (4.74), cancel algebraically. More precisely, defining

$$\widehat{Y}_t := \left|\langle(\operatorname{Im} G_{1,t}E_+\operatorname{Im} G_{2,t} - \widehat{M}_{12,t}^{E_+})E_+\rangle\right|,$$

and combining (4.79)–(4.84), we obtain

$$\widehat{Y}_{t\wedge\tau_3} = \widehat{Y}_0 + \frac{1}{2}\int_0^t \left[\langle M_{12,s}^I\rangle + \langle M_{\overline{1}2,s}^I\rangle + \langle M_{1\overline{2},s}^I\rangle + \langle M_{\overline{12},s}^I\rangle\right]\widehat{Y}_s\,\mathrm{d}s + n^\xi\widehat{\alpha}_{t\wedge\tau_3}. \tag{4.86}$$

By a Gronwall inequality, using that by explicit computations (see e.g. [35, Lemma 5.7])

$$\exp\left(\frac{1}{2}\int_s^t \left[\langle M^I_{12,r}\rangle + \langle M^I_{\overline{1}2,r}\rangle + \langle M^I_{1\overline{2},r}\rangle + \langle M^I_{\overline{12},r}\rangle\right] \mathrm{d}r\right) \le \exp\left(\int_s^t \left|\langle M^I_{12,r}\rangle\right| \mathrm{d}r\right) \lesssim \frac{\gamma_s^2}{\gamma_t^2} \sim \frac{\widehat{\gamma}_s^2}{\widehat{\gamma}_t^2}, \tag{4.87}$$

and proceeding similar arguments as in the end of [35, Secion 5.1] starting from Eq. (5.58) therein, we obtain $\widehat{Y}_{t\wedge\tau_3} \le n^\xi \widehat{\alpha}_{t\wedge\tau_3}$. This concludes the proof for $(A_1, A_2) = (E_+, E_+)$. Given the result for $(A_1, A_2) = (E_+, E_+)$, the estimate in the case $(A_1, A_2) = (E_+, E_-)$ is completely analogous obtaining the bound $n^\xi \widehat{\alpha}_{t\wedge\tau_3}$ in this case as well. This concludes the proof showing that $\tau_3 = T^*$ with very high probability. □

5. The zag step: proofs of Theorem 3.4 and Theorem 3.5

In Section 4 (zig step), we proved in Proposition 4.3 that if the local laws from Theorem 3.4 hold on a global scale $\eta \sim 1$, then they also hold down to the optimal local scale $\eta \gg 1/n$, at the cost of adding a Gaussian component. This section (zag step) is devoted to remove this added Gaussian component using a Gronwall argument (see Lemma 5.4 and (5.49) below) and hence prove Theorem 3.4. The proof of Theorem 3.5 near the edge is similar to Theorem 3.4 using (5.75) below to remove the Gaussian component added in Proposition 4.8.

The outline of this section is as follows. We first present the proof of the isotropic local law (3.25) and the averaged law in Section 5.1 and 5.2 respectively. In Sections 5.3, we prove standard local laws without $|z_1 - z_2|$ decay stated in Proposition 5.1 below, which serve as an input to prove Theorem 3.4. The proof of Theorem 3.5 is presented in Section 5.4 using Theorem 3.5 as an input.

Before we prove Theorem 3.4, we introduce the local laws for $G^{z_1}(\mathrm{i}\eta_1)A_1G_2^{z_2}(\mathrm{i}\eta_2)$ without gaining from the $|z_1-z_2|$-decorrelations. The proof of these local laws is standard and thus postponed to Section 5.3.

Proposition 5.1. *Fix small constants $\epsilon, \tau > 0$. The following hold*

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2) - M_{12}^{A_1}\right)_{\boldsymbol{xy}}\right| \prec \frac{\sqrt{\rho^*}}{\sqrt{n}\eta_*^{3/2}}, \tag{5.1}$$

$$\left|\left\langle\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2) - M_{12}^{A_1}\right)A_2\right\rangle\right| \prec \frac{1}{n\eta_*^2}, \tag{5.2}$$

uniformly for any $|z_i| \le 1+\tau$, $\min_{i=1}^2\{|\eta_i|\rho_i\} \ge n^{-1+\epsilon}$, any deterministic matrices $\|A_1\| + \|A_2\| \lesssim 1$ and unit vectors $\boldsymbol{x}, \boldsymbol{y}$. Here $\rho_i := \rho^{z_i}(\mathrm{i}\eta_i)$, $\eta_* := \min_{i=1}^2 |\eta_i|$,

$\rho^* := \max_{i=1}^2 \rho_i$, and $M_{12}^{A_1}$ is given in (3.16). Moreover, recalling $M_{121}^{A_1,A_2}$ defined in (3.19), we also have the following estimate for three resolvents

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2)A_2G^{z_1}(\mathrm{i}\eta_1) - M_{121}^{A_1,A_2}\right)_{\boldsymbol{xy}}\right| \prec \frac{1}{\eta_*^2}. \tag{5.3}$$

Next we aim to improve Proposition 5.1 to Theorem 3.4 gaining from the $|z_1 - z_2|$-decorrelation.

5.1. Proof of the isotropic local law in (3.25)

We start with proving the isotropic local law in (3.25), using Proposition 5.2 below iteratively. In this iteration procedure we gradually replace a small power of η_*/ρ^* with γ. We will focus on the regime that $\gamma \gtrsim \eta_*/\rho^*$, otherwise the desired local law follows directly from Proposition 5.1.

Proposition 5.2. *Fix small constants $\epsilon, \tau > 0$. Assume that for any $z_i \in \mathbb{C}$ and $\eta_i \in \mathbb{R}$ satisfying $|z_i| \leq 1 + \tau$, $\min_{i=1}^2\{|\eta_i|\rho_i\} \geq n^{-1+\epsilon}$, $\gamma \gtrsim \eta_*/\rho^*$, the following hold for some $0 \leq b < 1$,* i.e.,

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2) - M_{12}^{A_1}\right)_{\boldsymbol{xy}}\right| \prec \frac{(\rho^*)^{\frac{1-b}{2}}}{\sqrt{n}(\eta_*)^{\frac{3-b}{2}}\gamma^{\frac{b}{2}}}, \tag{5.4}$$

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2)A_2G^{z_1}(\mathrm{i}\eta_1) - M_{121}^{A_1,A_2}\right)_{\boldsymbol{xy}}\right| \prec \frac{1}{\eta_*^{2-b}\gamma^b}, \tag{5.5}$$

uniformly for any deterministic bounded matrices A_i $(i = 1, 2)$ and unit vectors $\boldsymbol{x}, \boldsymbol{y}$, where γ is defined in (3.23) and $M_{12}^{A_1}$, $M_{121}^{A_1,A_2}$ are defined in (3.16), (3.19). Then the following also hold for a larger $b' = \min\{b + \theta,\ 1\}$ with some $0 < \theta \leq \epsilon/10$,

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2) - M_{12}^{A_1}\right)_{\boldsymbol{xy}}\right| \prec \frac{(\rho^*)^{\frac{1-b'}{2}}}{\sqrt{n}(\eta_*)^{\frac{3-b'}{2}}\gamma^{\frac{b'}{2}}}, \tag{5.6}$$

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2)A_2G^{z_1}(\mathrm{i}\eta_1) - M_{121}^{A_1,A_2}\right)_{\boldsymbol{xy}}\right| \prec \frac{1}{\eta_*^{2-b'}\gamma^{b'}}, \tag{5.7}$$

uniformly for any deterministic bounded matrices A_i $(i = 1, 2)$ and unit vectors $\boldsymbol{x}, \boldsymbol{y}$.

Proof of (3.25). Note that Proposition 5.1 implies that the initial assumptions in (5.4)-(5.5) with $b = 0$ are satisfied. Starting from (5.4)-(5.5) for $b = 0$ as the initial step and iterating Proposition 5.2 for at most $O(\theta^{-1})$ times by gradually increasing b from 0 to 1, we have proved the isotropic local law for $G^{z_1}A_1G^{z_2}$ in (3.25) modulo Proposition 5.2 that we will prove shortly. For a graphical illustration of this iterative procedure we refer to Fig. 2. □

Note that as a by-product, we also obtained the following isotropic version for three resolvents. This Lemma 5.3 will be needed as an auxiliary estimate in the next section to prove the averaged local law in (3.26).

Lemma 5.3. *Fix small constants $\epsilon, \tau > 0$. For any deterministic matrices $\|A_1\|+\|A_2\| \lesssim 1$ and vectors $\|\boldsymbol{x}\|, \|\boldsymbol{y}\| \lesssim 1$, the following holds*

$$\left|\left\langle \boldsymbol{x}, \left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2)A_2G^{z_1}(\mathrm{i}\eta_1) - M_{121}^{A_1,A_2}\right)\boldsymbol{y}\right\rangle\right| \prec \frac{1}{\eta_*\gamma}, \tag{5.8}$$

uniformly for any $|z_i| \leq 1+\tau$ and $\min_{i=1}^2\{|\eta_i|\rho_i\} \geq n^{-1+\epsilon}$.

Before we give the actual proof of Proposition 5.2, we introduce the following short-hand notations and a preliminary lemma for zag step (see Lemma 5.4 below). Define

$$R^{xy} := \left(G_1AG_2 - M_{12}^{A}\right)_{\boldsymbol{xy}}, \qquad S^{xy} := \left(G_1A_1G_2A_2G_1 - M_{121}^{A_1,A_2}\right)_{\boldsymbol{xy}}, \tag{5.9}$$

with $G_i := G^{z_i}(\mathrm{i}\eta_i)$, $i = 1, 2$, and set the following control parameters

$$\mathcal{E}_{\mathfrak{b}} := \frac{(\rho^*)^{\frac{1-\mathfrak{b}}{2}}}{\sqrt{n}(\eta_*)^{\frac{3-\mathfrak{b}}{2}}\gamma^{\frac{\mathfrak{b}}{2}}}, \qquad \widetilde{\mathcal{E}}_{\mathfrak{b}} := \frac{1}{\eta_*^{2-\mathfrak{b}}\gamma^{\mathfrak{b}}}, \qquad \mathfrak{b} \in [0,1]. \tag{5.10}$$

It is easy to check that, for any $0 \leq b \leq b' \leq 1$,

$$\frac{1}{\sqrt{n\gamma}\eta_*} \lesssim \mathcal{E}_{b'} \leq \mathcal{E}_b \leq \frac{\sqrt{\rho^*}}{\sqrt{n}\eta_*^{3/2}}, \qquad \frac{1}{\eta_*\gamma} \lesssim \widetilde{\mathcal{E}}_{b'} \leq \widetilde{\mathcal{E}}_b \leq \frac{1}{\eta_*^2}. \tag{5.11}$$

With the above short-hand notations, we aim to prove, for any $z_i \in \mathbb{C}$ and $\eta \in \mathbb{R}$ satisfying $|z_i| \leq 1+\tau$, $n|\eta_i|\rho_i \gtrsim n^{\epsilon}$, and $\gamma \gtrsim \eta_*/\rho^*$,

$$|R^{xy}| \prec \mathcal{E}_{b'}, \qquad |S^{xy}| \prec \widetilde{\mathcal{E}}_{b'}, \qquad b' = \min\left\{b+\theta,\ 1\right\}, \tag{5.12}$$

assuming that $|R^{xy}| \prec \mathcal{E}_b$ and $|S^{xy}| \prec \widetilde{\mathcal{E}}_b$ for some $0 \leq b < 1$.

The proof of Proposition 5.2 will be a Gronwall argument [zag step] together with Proposition 4.4 [zig step]. The bounds (5.6) and (5.7) are already known from Proposition 4.4 (together with the global laws needed as an initial input that will be stated later in (5.18)–(5.19)) whenever X has a small Gaussian component. We show that this Gaussian component does not change the estimate up to the relevant order by running an Ornstein-Uhlenbeck flow to remove the small Gaussian component. We will work with the time evolution of high moments of R_t^{xy}, S_t^{xy}. The key of the Gronwall argument is to give an estimate on the time derivative of these high moments in terms of themselves, this will be the content of Lemma 5.4 below.

We now setup this flow argument. Given the initial ensemble H^z in (3.2) for fixed $z \in \mathbb{C}$ and X satisfying Assumption 2.1, we consider the following Ornstein-Uhlenbeck matrix flow

$$\mathrm{d}H_t^z = -\frac{1}{2}(H_t^z + Z)\mathrm{d}t + \frac{1}{\sqrt{n}}\mathrm{d}\mathcal{B}_t, \quad Z := \begin{pmatrix} 0 & zI \\ \overline{z}I & 0 \end{pmatrix}, \quad \mathcal{B}_t := \begin{pmatrix} 0 & B_t \\ B_t^* & 0 \end{pmatrix}, \tag{5.13}$$

with initial condition $H_{t=0}^z = H^z$, where B_t is an $n \times n$ matrix with i.i.d. standard complex valued Brownian motion entries. Indeed, the matrix flow H_t^z interpolates between the initial matrix H^z and the same matrix with X replaced with an independent complex Ginibre ensemble, *i.e.*,

$$X_t \stackrel{\mathrm{d}}{=} e^{-\frac{t}{2}}X + \sqrt{1-e^{-t}}\,\mathrm{Gin}(\mathbb{C}). \tag{5.14}$$

The corresponding time-dependent resolvent of $H_t^{z_i}$ $(i = 1, 2)$ is denoted by

$$G_i := G_t^{z_i}(\mathrm{i}\eta_i) = (H_t^{z_i} - \mathrm{i}\eta_i)^{-1}, \qquad i = 1, 2,$$

which we use to define time dependent R_t^{xy}, S_t^{xy} as in (5.9). For notational brevity, we will not keep track of the dependence of t in the notations G_i from now on. Note that the parameters $z_i \in \mathbb{C}$, $\eta_i \in \mathbb{R}$ are fixed and hence the deterministic matrices M_{12}^A and $M_{121}^{A_1,A_2}$ in (5.9) are independent of the time t.

Lemma 5.4. *Fix small constant $\epsilon, \tau > 0$ and any integer $m \geq 1$. For any $|z_i| \leq 1 + \tau$ and $\eta_i \in \mathbb{R}$ satisfying $\min_{i=1}^2\{|\eta_i|\rho_i\} \geq n^{-1+\epsilon}$ and $\gamma \gtrsim \eta_*/\rho^*$, assuming*[7] *that for some $0 \leq b < 1$,*

$$\boldsymbol{E}|R_t^{xy}|^{2m} \prec (\mathcal{E}_b)^{2m}, \qquad R_t^{xy} = \big(G_1AG_2 - M_{12}^A\big)_{\boldsymbol{xy}},$$
$$\boldsymbol{E}|S_t^{xy}|^{2m} \prec (\widetilde{\mathcal{E}}_b)^{2m}, \qquad S_t^{xy} = \big(G_1A_1G_2A_2G_1 - M_{121}^{A_1,A_2}\big)_{\boldsymbol{xy}},$$

uniformly for any $t \geq 0$, any deterministic bounded matrices A_i $(i = 1, 2)$ and unit vectors $\boldsymbol{x}, \boldsymbol{y}$, then the following hold for $b' = \min\{b + \theta,\ 1\}$, $0 < \theta \leq \epsilon/10$, and the prefactor $C_n := 1 + \frac{1}{\sqrt{n}\eta_*^{1+\frac{\epsilon}{2}}}$,

$$\Big|\frac{\mathrm{d}\boldsymbol{E}|R_t^{xy}|^{2m}}{\mathrm{d}t}\Big| \lesssim C_n\Big(\boldsymbol{E}\big|R_t^{xy}\big|^{2m} + O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\Big), \tag{5.15}$$

$$\Big|\frac{\mathrm{d}\boldsymbol{E}\big|\mathcal{S}_t^{xy}\big|^{2m}}{\mathrm{d}t}\Big| \lesssim C_n\Big(\boldsymbol{E}\big|\mathcal{S}_t^{xy}\big|^{2m} + O_\prec\big(\widetilde{\mathcal{E}}_{b'}\big)^{2m}\Big), \tag{5.16}$$

[7] We remark that stochastic domination $\prec$ and upper bounds of arbitrarily high moments of a random variable are equivalent for random variables bounded by some n^C which is always true in our case, *i.e.*, for $|X| \leq n^C$ we have that $X \prec \Psi$ if and only if $\boldsymbol{E}|X|^{2m} \leq n^\xi \Psi^{2m}$ for any $m \geq 1$ and $\xi > 0$. We frequently use this fact in the proof without mentioning it.

uniformly for any $t \geq 0$, any deterministic bounded matrices A_i $(i = 1, 2)$ and unit vectors $\boldsymbol{x}, \boldsymbol{y}$.

The proof of Lemma 5.4 will be postponed till the end of this section. Now we are ready to use Proposition 4.4 (*zig-step*) and Lemma 5.4 (main technical input for *zag-step*) to prove Proposition 5.2.

Proof of Proposition 5.2. Recall that the matrix flows X_t, W_t defined in (4.1)-(4.2) and the characteristic flow $\Lambda(t) := \Lambda_t$ defined in (4.4). The corresponding resolvent is given by $G_t^{z_t}(\mathrm{i}\eta_t) = (H_t^{z_t} - \mathrm{i}\eta_t)^{-1}$, where $H_t^{z_t}$ is defined as in (3.1) with time dependent z_t and X_t. We remark that one should not confuse this notation with $G_t^z(\mathrm{i}\eta) = (H_t^z - \mathrm{i}\eta)^{-1}$ defined below (5.14), where both $\eta \in \mathbb{R}$ and $z \in \mathbb{C}$ are fixed.

Fix any $|z_i| \leq 1+\tau$ and η_i with $n|\eta_i|\rho_i \gtrsim n^\epsilon$ where $\rho_i := \rho^{z_i}(\mathrm{i}\eta_i)$. Choose a small $T \sim 1$. By Lemma 4.2, there exist initial conditions $\Lambda_i(0)$ satisfying $\mathrm{dist}(\mathrm{i}\eta_i(0), \mathrm{supp}(\rho_i(0))) \gtrsim 1$ such that at time $t = T$, $z_i(T) = z_i$ and $\eta_i(T) = \eta_i$. We also remark that from Lemma 4.1, the initial conditions $\eta_i(0), z_i(0), \rho_i(0)$ have the following properties:

$$\frac{\eta_i(0)}{\rho_i(0)} \gtrsim T, \qquad \rho_i(0) \sim \rho_i(T), \qquad z_i(0) = z_i(T)(1 + O(T)). \tag{5.17}$$

We next state the following global laws, *i.e.*, for any spectral parameters with $\mathrm{dist}(\mathrm{i}\eta_i, \mathrm{supp}(\rho^{z_i})) \sim 1$,

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2) - M_{12}^{A_1}\right)_{\boldsymbol{xy}}\right| \prec \frac{1}{\sqrt{n}}, \tag{5.18}$$

$$\left|\left(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2)A_2G^{z_1}(\mathrm{i}\eta_1) - M_{121}^{A_1,A_2}\right)_{\boldsymbol{xy}}\right| \prec 1. \tag{5.19}$$

The proof of these global laws is standard and similar to [31, Proposition 4.1] (see also [26, Appendix B]) with the only exception that in the current case we need to consider the 2×2 block structure due to the chiral symmetry, but this can be done with very minor changes in the proof, and so omitted. These global laws directly imply that the initial condition (4.18) of Proposition 4.4 at $t = 0$ is satisfied, hence from this proposition we obtain the desired local laws at $t = T \sim 1$, *i.e.*,

$$|R_T^{xy}| \prec \frac{1}{\sqrt{n\gamma}\eta_*}, \qquad |S_T^{xy}| \prec \frac{1}{\eta_*\gamma}, \qquad \eta_i(T) = \eta_i, \quad z_i(T) = z_i. \tag{5.20}$$

This completes the zig step and it proves Theorem 3.4 for X being a Gaussian divisible matrix, albeit with a large Gaussian component, *i.e.*, $X_T \overset{\mathrm{d}}{=} e^{-\frac{T}{2}}X + \sqrt{1 - e^{-T}}\mathrm{Gin}(\mathbb{C})$, with $T \sim 1$.

Next we aim to use Lemma 5.4 to remove the Gaussian component $T \sim 1$. However the differential inequalities in (5.15)-(5.16) are only effective in the Gronwall argument when the pre–factor C_n is bounded, which means that η_* needs to be somewhat large.

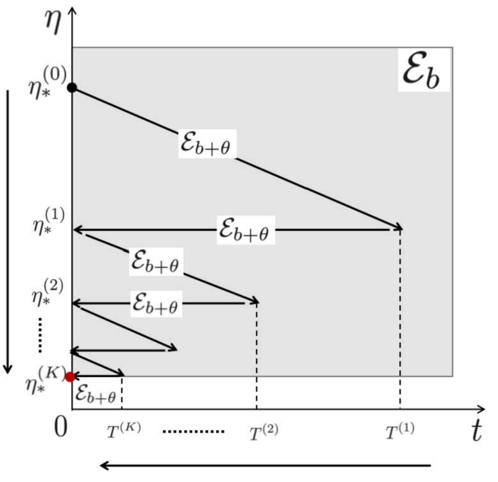

Fig. 1. The shaded square indicates the region in the (η, t) plane where the local law holds with an error bound $\mathcal{E}_b$, as given in assumption (5.4). Starting from the global law at the initial black dot ($\eta_*^{(0)} \sim 1$), the slanted line downward indicates the zig step using Proposition 4.4, while the horizontal line leftward indicates the zag step using Lemma 5.4 and a Gronwall argument. Iterating finitely many zig-zags until reaching the final red point ($\eta_*^{(K)} \sim \eta_*$, $K = O(\epsilon^{-1})$) yields the improved local law with an improved error $\mathcal{E}_{b+\theta}$ for any $\min_{i=1}^2 |\eta_i|\rho_i \gtrsim n^{-1+\epsilon}$. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

So we will need an iterative procedure to handle the smaller η_* regime using a smaller Gaussian component, hence smaller T, so that the larger prefactor C_n is compensated by the shorter integration time.

In the first step, we assume that $|\eta_*| \gtrsim n^{-1/2+\epsilon/2}$, then the prefactor $C_n \le C$ for some constant $C > 0$. Using Gronwall's inequality for $T \sim 1$, we have

$$\boldsymbol{E}|R_0^{xy}|^{2m} \le e^{\int_0^T C\mathrm{d}t}\boldsymbol{E}|R_T^{xy}|^{2m} + \int_0^T e^{\int_s^T C\mathrm{d}t} O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\mathrm{d}s = O_\prec\big(\mathcal{E}_{b'}\big)^{2m}, \tag{5.21}$$

where we also used (5.20) and (5.11). One can handle S_0^{xy} in a similar way. This hence proves (5.12) for any $|z_i| \le 1+\tau$ and $|\eta_*| \gtrsim n^{-1/2+\epsilon/2}$, *i.e.*,

$$\big|R_0^{xy}\big| \prec \mathcal{E}_{b'}, \qquad \big|S_0^{xy}\big| \prec \widetilde{\mathcal{E}}_{b'}, \qquad\qquad \forall |\eta_i| \gtrsim n^{-1/2+\epsilon/2}. \tag{5.22}$$

This completes the zag step.

To extend the above results to smaller $|\eta_*| \lesssim n^{-1/2+\epsilon/2}$, we use zig-zag iteratively as follows (see the illustration in Fig. 1). In each iteration step $k \ge 1$, we use the superscript (k) to indicate the step number, *e.g.*, we denote the spectral parameters by $\eta_i^{(k)} \in \mathbb{R}$, $z_i^{(k)} \in \mathbb{C}$ and the time by $T^{(k)} \in \mathbb{R}^+$. Recall that our final goal is to prove (5.12) for any fixed $|z_i| \le 1+\tau$ and $n|\eta_i|\rho_i \gtrsim n^\epsilon$. We also recall from Lemma 4.1 that the size of ρ_i does not change along the characteristic flow in (4.4) by more than a constant factor. Note that (3.10) and $n|\eta_i|\rho \gtrsim n^\epsilon$ imply that $|\eta_i| \gtrsim n^{-1+\epsilon}$ and $n^{-1/2+\epsilon/2} \lesssim \rho_i \lesssim 1$. Then we set $\rho_* = \min_{i=1}^2 \rho_i \sim n^{-c_*}$ with $0 \le c_* \le 1/2 - \epsilon/2$. In general for any $k \ge 1$, we define the scales recursively as follows:

$$\eta_*^{(k)} := n^{-a_k}, \qquad a_k = \frac{\frac{1}{2} + a_{k-1} - c_*}{1+\frac{\epsilon}{2}}, \qquad a_1 = \frac{1}{2} - \frac{\epsilon}{2}. \tag{5.23}$$

We will stop the iterations if $\eta_*^{(k)}$ reaches $\eta_* = \min_{i=1}^2 |\eta_i| \gtrsim n^{-1+\epsilon}$, so $a_k \ge 1-\epsilon$. The number of iteration steps is denoted by K. It is straightforward to check that a_k is increasing in k at least up to 1 using that $0 \le c_* \le 1/2 - \epsilon/2$.

In the k-th step, we fix $|\eta_i^{(k)}| \gtrsim \eta_*^{(k)} = n^{-a_k}$ and $|z_i^{(k)}| \le 1+\tau$ with $\rho_i^{(k)} \sim \rho_i$. Assume that in the previous step, we have proved (5.22) for any $|\eta_i^{(k-1)}| \gtrsim \eta_*^{(k-1)} = n^{-a_{k-1}}$ and $|z_i^{(k-1)}| \le 1+\tau$ with $\rho_i^{(k-1)} \sim \rho_i$. Recall the characteristic flow $\Lambda(t) = \Lambda_t$ in (4.4) and choose $T^{(k)} \sim \eta_*^{(k-1)}/\rho_* = n^{-a_{k-1}+c_*}$. From Lemma 4.1, there exist initial conditions $|z_i(0)| = |z_i^{(k)}| + O(T^{(k)})$ and $|\eta_i(0)| \gtrsim \rho_* T^{(k)} = \eta_*^{(k-1)} = n^{-a_{k-1}}$ such that $\eta_i(T^{(k)}) = \eta_i^{(k)}$ and $z_i(T^{(k)}) = z_i^{(k)}$. The result from the previous step in (5.22) ensures that the assumption (4.18) of Proposition 4.4 (with b' playing the role of b) is satisfied for any $|\eta_i^{(k-1)}| \gtrsim \eta_*^{(k-1)} = n^{-a_{k-1}}$. Thus Proposition 4.4 implies that for $t = T^{(k)} \sim n^{-a_{k-1}+c_*}$,

$$\left|R^{xy}_{T^{(k)}}\right| \prec \mathcal{E}_{b'}, \qquad \left|S^{xy}_{T^{(k)}}\right| \prec \widetilde{\mathcal{E}}_{b'}, \qquad z_i(T^{(k)}) = z_i^{(k)}, \quad \eta_i(T^{(k)}) = \eta_i^{(k)}, \tag{5.24}$$

with $|z_i^{(k)}| \le 1+\tau$ and $|\eta_i^{(k)}| \gtrsim \eta_*^{(k)} = n^{-a_k}$ such that $\rho_i^{(k)} \sim \rho_i$. Then we apply Gronwall's inequality to (5.15)-(5.16) similarly to (5.21) to remove the Gaussian component $T^{(k)} \sim n^{-a_{k-1}+c_*}$. Note that the prefactor $C_n = 1 + \frac{1}{\sqrt{n}(\eta_*^{(k)})^{1+\frac{\epsilon}{2}}}$ satisfies

$$e^{\int_0^{T^{(k)}} C_n \mathrm{d}t} \lesssim 1 + T^{(k)} C_n \lesssim 1 + n^{-a_{k-1}+c_*}\Big(1 + \frac{1}{n^{1/2 - a_k(1+\frac{\epsilon}{2})}}\Big)\Big) \lesssim 1, \tag{5.25}$$

using (5.23) and that $c_* \le 1/2 - \epsilon/2 \le a_k$ for any $k \ge 1$. Hence we extend (5.22) down to the regime $|z_i^{(k)}| \le 1+\tau$, $|\eta_i^{(k)}| \gtrsim \eta_*^{(k)} = n^{-a_k}$, $\rho_i^{(k)} \sim \rho_i$ in the k-th iteration step.

We iterate the above process and stop at K-th step when $\eta_*^{(K)}$ reaches $\eta_* = \min_{i=1}^2 |\eta_i|$. The number of iterations K indeed depends on c_* and η_* but it always remains n-independent. For example, if both z_1 and z_2 are in the bulk $|z_i| \le 1-\tau$ and $\eta_* = n^{-1+\epsilon}$, hence $c_* = 0$, then we only need to iterate twice using (5.23). The same also applies if both are at the edge $|z_i| = 1 + O(n^{-1/2})$ and $\eta_* = n^{-3/4+\epsilon}$, in which case $c_* = 1/4 - \epsilon/4$. However, if z_1 is in the bulk and z_2 is at the edge with $c_* = 1/4 - \epsilon/4$ and $\eta_* = n^{-1+\epsilon}$, then we need to iterate for three times. The worst situation is when $|z_1| = 1-\tau$ and $|z_2| = 1+\tau$ with $c_* = 1/2 - \epsilon/2$ and $\eta_* = n^{-1+\epsilon}$. Then from (5.23), we know that

$$a_k - a_{k-1} = \frac{a_{k-1} + \frac{\epsilon}{2}}{1+\frac{\epsilon}{2}} - a_{k-1} = \frac{\epsilon}{2}\frac{1-a_{k-1}}{1+\frac{\epsilon}{2}} \lesssim \epsilon^2,$$

using that $a_k \le 1-\epsilon$. This implies that we need to iterate for at most $O(\epsilon^{-2})$ times to extend (5.22) down to the final regime $n|\eta_i|\rho_i \gtrsim n^{\epsilon}$. $\square$

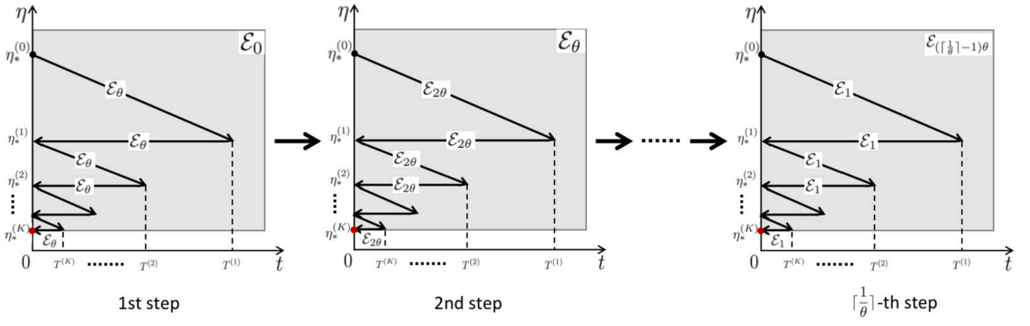

Fig. 2. In each step we use the same zig-zag as explained in Fig. 1 to improve the local law obtained in the previous step (with the error bound $\mathcal{E}_{(k-1)\theta}$ indicated on the upper right corner of the shaded square) to the next order bound $\mathcal{E}_{k\theta}$ (indicated along the zigzag lines). Starting from Theorem 5.1 as the initial input, we iterate this process for $\lceil\frac{1}{\theta}\rceil$ times to improve the a priori error bound $\mathcal{E}_0$ to the desired $\mathcal{E}_1$.

To summarize the proof of the isotropic local law, we use Fig. 2 to illustrate both the zig-zag iteration used in the proof of Proposition 5.2 and the b-iterations used to prove (3.25).

Finally, we end this section with the proof of Lemma 5.4.

Proof of Lemma 5.4. To simplify the notations, we assume that $|\eta_1| \sim |\eta_2| \sim \eta$ and $\rho_1 \sim \rho_2 \sim \rho$ for some η, ρ. The proof for general η_i and ρ_i is essentially the same with η, ρ below replaced by η_*, ρ^*.

We first apply Itô's formula to $|R_t^{xy}|^{2m} = (R_t^{xy})^m(\overline{R_t^{xy}})^m$ and perform the cumulant expansions (see *e.g.*, [55, Lemma 7.1]) as below. Without loss of generality, we may assume in the following that all the cumulants of the normalized entries of $\sqrt{n}X$ are equal to one. We thus obtain

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{E}|R_t^{xy}|^{2m}}{\mathrm{d}t} &= \sum_{k=3}^{K_0}\frac{1}{k!}\mathcal{T}_k + O_\prec\big(n^{-100m}\big)\\ &= \sum_{k=3}^{K_0}\frac{1}{k!}\left(\sum_{l=1}^{k}\sum_{s_1+\cdots+s_l=k;s_i\geq 1}\mathcal{T}_k^{(s_1,\cdots,s_l)}\right) + O_\prec\big(n^{-100m}\big),\end{aligned} \tag{5.26}$$

with the k-th order cumulant expansion term given by

$$\mathcal{T}_k^{(s_1,\cdots,s_l)} := \frac{1}{n^{\frac{k}{2}}}\sum_{a=1}^{n}\sum_{B=n+1}^{2n}\boldsymbol{E}\Big[\big(\partial^{s_1}R_t^{xy}\big)\cdots\big(\partial^{s_l}R_t^{xy}\big)\big(R_t^{xy}\big)^{2m-l}\Big], \qquad s_i \geq 1, \tag{5.27}$$

where ∂^{s_i} denotes the s_i-th directional derivative either in the direction h_{aB} or in h_{aB}, with $(h_{\mathfrak{u}\mathfrak{v}})_{\mathfrak{u},\mathfrak{v}\in[2n]} := H_t^z$ given by (5.13). Note that for $a, b \in [n]$ and $A, B \in [n+1, 2n]$, the corresponding entries $\{h_{ab}\}$ and $\{h_{AB}\}$ all equal to zero. For notational brevity, we will not distinguish between ∂h_{aB} and ∂h_{Ba} and use the notation ∂ to indicate either

$\partial/\partial h_{aB}$ or $\partial/\partial h_{Ba}$. From now on, with a slight abuse of notation R_t^{xy} indicates either $R_t^{xy} \in \mathbb{C}$ itself or its complex conjugate $\overline{R_t^{xy}}$). Here we stopped the cumulant expansions in (5.26) at a sufficiently high power $K_0 = O(m)$, and the last error term $O_\prec(n^{-100m})$ was estimated using the moment condition in (2.1) and that $\|G\| \le |\eta|^{-1}$. This argument is quite standard and already used in many previous works, *e.g.,* below [35, Eq. (7.6)], so we omit the details.

By a direct computation, we obtain that

$$\frac{\partial R_t^{xy}}{\partial h_{aB}} = -\big((G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}} + (G_1AG_2)_{\boldsymbol{x}\boldsymbol{e}_a}(G_2)_{\boldsymbol{e}_B\boldsymbol{y}}\big). \tag{5.28}$$

In general for any $s \ge 1$, each partial derivative $\partial^s R_t^{xy}$ contains a product of s Green function entries $\prod_{i=1}^s (G)_{\boldsymbol{u}_i,\boldsymbol{v}_i}$ and one factor $(G_1AG_2)_{\boldsymbol{u}_{s+1}\boldsymbol{v}_{s+1}}$ with $\boldsymbol{u}_i, \boldsymbol{v}_i$ $(1 \le i \le s+1)$ chosen from $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{e}_a, \boldsymbol{e}_B\}$. Moreover, the choice of $\boldsymbol{x}$ and $\boldsymbol{y}$ only appears once, while both a and B appear s times. For instance, we list below several examples of $\partial^2 R_t^{xy}$ and $\partial^3 R_t^{xy}$, *i.e.,*

$$\partial^2 R_t^{xy}: \qquad (G_1)_{\boldsymbol{x}\boldsymbol{e}_B}(G_1)_{aa}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}, \quad (G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1AG_2)_{BB}(G_2)_{\boldsymbol{e}_a\boldsymbol{y}}; \tag{5.29}$$

$$\partial^3 R_t^{xy}: \qquad (G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1)_{aa}(G_1)_{BB}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}, \quad (G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1)_{aa}(G_1AG_2)_{BB}(G_2)_{\boldsymbol{e}_B\boldsymbol{y}}. \tag{5.30}$$

Moreover, we introduce the following estimates on the products of Green function entries. Note that using (3.3) and (3.20), we have

$$\sum_{\mathrm{j}=1}^{2n} |(M)_{\boldsymbol{u}\boldsymbol{e}_{\mathrm{j}}}|^k \lesssim 1 + \sqrt{n}\mathbb{1}_{k=1}, \qquad \sum_{\mathrm{j}=1}^{2n} |(M_{12}^A)_{\boldsymbol{u}\boldsymbol{e}_{\mathrm{j}}}|^k \lesssim \frac{1}{\gamma^k} + \frac{\sqrt{n}}{\gamma}\mathbb{1}_{k=1}, \tag{5.31}$$

for any $k \ge 1$ and any bounded deterministic vector $\boldsymbol{u} \in \mathbb{C}^n$. Note that for $k = 1$ we also used the Cauchy-Schwarz inequality. As a consequence of (5.31), combining with the isotropic local law in (3.11) and the assumption in (5.4), we have

$$\begin{aligned} &\sum_{\mathrm{j}=1}^{2n} \big|(G)_{\boldsymbol{u}\boldsymbol{e}_{\mathrm{j}}}\big|^k \prec n\Big(\sqrt{\frac{\rho}{n\eta}}\Big)^k + 1 + \sqrt{n}\mathbb{1}_{k=1}, \\ &\sum_{\mathrm{j}=1}^{2n} \big|(G_1AG_2)_{\boldsymbol{u}\boldsymbol{e}_{\mathrm{j}}}\big|^k \prec n\mathcal{E}_b^k + \frac{1}{\gamma^k} + \frac{\sqrt{n}}{\gamma}\mathbb{1}_{k=1}, \end{aligned} \tag{5.32}$$

for any $k \ge 1$ and any bounded deterministic vector $\boldsymbol{u} \in \mathbb{C}^n$. We remark that the above estimates hold simultaneously for all $t \ge 0$. This follows by a standard grid argument using the Lipschitz continuity of $G_t = (H_t - \mathrm{i}\eta)^{-1}$ for $t \in [0, T_0]$ with $T_0 = 100\log n$, while for $t \ge T_0$, G_t is almost the resolvent of a Ginibre ensemble up to an error $O_\prec(n^{-10})$. Without specific mentioning, the estimates in the following hold uniformly for all t.

Now we are ready to look at the most critical third order terms given by (5.27), *i.e.*,

$$\begin{aligned}\mathcal{T}_3^{(3)} =&\mathcal{T}_3^{(1,1,1)} + \mathcal{T}_3^{(1,2)} + \mathcal{T}_3^{(3)}\\ =&\frac{1}{n^{3/2}}\sum_{a,B}\boldsymbol{E}\Big[\big(\partial R_t^{xy}\big)^3\big(R_t^{xy}\big)^{2m-3} + \partial R_t^{xy}\big(\partial^2 R_t^{xy}\big)\big(R_t^{xy}\big)^{2m-2} + \big(\partial^3 R_t^{xy}\big)\big(R_t^{xy}\big)^{2m-1}\Big].\end{aligned} \tag{5.33}$$

Using (5.28) and (5.32) for $k=3$, the first part $\mathcal{T}_3^{(1,1,1)}$ in (5.33) is bounded by

$$\begin{aligned}\big|\mathcal{T}_3^{(1,1,1)}\big| \lesssim&\frac{1}{n^{3/2}}\sum_{a,B}\boldsymbol{E}\Big[\Big(\Big|(G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\Big|^3 + \Big|(G_1AG_2)_{\boldsymbol{x}\boldsymbol{e}_a}G_{\boldsymbol{e}_B\boldsymbol{y}}\Big|^3\Big)\big|R_t^{xy}\big|^{2m-3}\Big]\\ \prec&\Big(\frac{\rho^{3/2}}{n\eta^{3/2}}+\frac{1}{\sqrt{n}}\Big)\Big(\mathcal{E}_b^3+\frac{1}{n}\gamma^{-3}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-3}\\ \lesssim&\Big(1+\frac{1}{n\eta^{3/2+\frac{\epsilon}{2}}}\Big)\Big(\boldsymbol{E}\big|R_t^{xy}\big|^{2m}+O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\Big),\end{aligned} \tag{5.34}$$

where in the last step we used Young's inequality and the following bounds from (5.10)-(5.11):

$$\frac{\mathcal{E}_b}{\mathcal{E}_{b'}} = \Big(\frac{\rho\gamma}{\eta}\Big)^{\frac{b'-b}{2}} \lesssim \Big(\frac{1}{\rho\eta}\Big)^{\frac{\theta}{2}}, \qquad \frac{\gamma^{-1}}{\mathcal{E}_{b'}} \lesssim \frac{\sqrt{n}\eta}{\sqrt{\gamma}} \lesssim \sqrt{n\eta\rho}, \tag{5.35}$$

with $b'=\min\big\{b+\theta,\ 1\big\}$, $0<\theta\le\epsilon/10$, using that $\gamma\lesssim\rho^{-2}$ from (3.23) and $\gamma\gtrsim\eta/\rho$.

We next consider the second part $\mathcal{T}_3^{(1,2)}$ in (5.33). Recall the second derivative terms for instance in (5.29). One of the most critical term (containing the least off-diagonal entries of G) is given by

$$(*) := \frac{1}{n^{3/2}}\sum_{a,B}\boldsymbol{E}\Big[(G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}(G_1)_{\boldsymbol{x}\boldsymbol{e}_B}(G_1)_{aa}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\Big)\big|R_t^{xy}\big|^{2m-2}\Big]. \tag{5.36}$$

Using (3.11) and by a Cauchy-Schwarz inequality, such term is bounded by

$$\begin{aligned}|(*)| \prec&\frac{\rho^{3/2}}{n\sqrt{\eta}}\boldsymbol{E}\Big[\sqrt{\sum_B\big|(G_1)_{\boldsymbol{x}\boldsymbol{e}_B}\big|^2\sum_B\big|(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\big|^4}\big|R_t^{xy}\big|^{2m-2}\Big]\\ \prec&\rho^2\Big(\frac{\mathcal{E}_b^2}{\sqrt{n\eta}}+\frac{1}{n\eta\gamma^2}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-2}\\ \lesssim&\Big(1+\frac{1}{\sqrt{n}\eta^{1+\frac{\epsilon}{2}}}\Big)\Big(\boldsymbol{E}\big|R_t^{xy}\big|^{2m}+O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\Big),\end{aligned} \tag{5.37}$$

where we used (5.32) for $k=2,4$ in the second step and (5.35) in the last step. The other terms in $\mathcal{T}_3^{(1,2)}$ can be bounded similarly as in (5.37) even with less efforts. Then we conclude that

$$\left|\mathcal{T}_3^{(1,2)}\right| \lesssim \Big(1+\frac{1}{\sqrt{n}\eta^{1+\frac{\epsilon}{2}}}\Big)\Big(\boldsymbol{E}\big|R_t^{xy}\big|^{2m}+O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\Big). \tag{5.38}$$

We finally consider the last part $\mathcal{T}_3^{(3)}$ in (5.33). Recall the third derivative terms in (5.30). We only look at one of the worst terms (with the least off-diagonal entries of G), *i.e.*,

$$\begin{aligned}(**) :=&\Big|\frac{1}{n^{3/2}}\sum_{a,B}\boldsymbol{E}\Big[(G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1)_{aa}(G_1)_{BB}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\big(R_t^{xy}\big)^{2m-1}\Big]\Big|\\ \prec&\Big|\frac{1}{n^{3/2}}\sum_{a,B}\boldsymbol{E}\Big[(G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1-M_1)_{aa}(G_1-M_1)_{BB}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\big(R_t^{xy}\big)^{2m-1}\Big]\Big|\\ &+\Big(\frac{\rho^{3/2}}{\sqrt{n\eta}}\mathcal{E}_b+\frac{\rho^2}{\sqrt{n}\gamma}+\frac{\rho^2}{n\eta\gamma}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-1},\end{aligned} \tag{5.39}$$

where we used the following isotropic bounds

$$\Big|\frac{1}{\sqrt{n}}\sum_a (G)_{\boldsymbol{x}\boldsymbol{e}_a}\Big| \prec 1+\sqrt{\frac{\rho}{n\eta}}, \qquad \Big|\frac{1}{\sqrt{n}}\sum_B (G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\Big| \prec \frac{1}{\gamma}+\mathcal{E}_b, \tag{5.40}$$

from (3.11), the assumption (5.4) and (5.31). Using (5.32) for $k=1$, we obtain that

$$\begin{aligned}|(**)| \prec&\Big(\frac{\rho^{3/2}}{n\eta^{3/2}}\mathcal{E}_b+\frac{\rho^{3/2}}{\sqrt{n\eta}}\mathcal{E}_b+\frac{\rho^2}{\sqrt{n}\gamma}+\frac{\rho^2}{n\eta\gamma}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-1}\\ \lesssim&\Big(1+\frac{1}{\sqrt{n}\eta^{1+\frac{\epsilon}{2}}}\Big)\Big(\boldsymbol{E}\big|R_t^{xy}\big|^{2m}+O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\Big),\end{aligned} \tag{5.41}$$

where we also used (5.35) and Young's inequality in the last step. The other terms in $\mathcal{T}_3^{(3)}$ can be handled similarly as above even with less efforts. Hence we obtain the same upper bound for $\mathcal{T}_3^{(3)}$ as in (5.38).

To sum up (5.34)-(5.41), all the third order terms in (5.33) are bounded by

$$\big|\mathcal{T}_3\big| \lesssim\Big(1+\frac{1}{\sqrt{n}\eta^{1+\frac{\epsilon}{2}}}\Big)\Big(\boldsymbol{E}\big|R_t^{xy}\big|^{2m}+O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\Big). \tag{5.42}$$

Note that the higher order terms in the cumulant expansion in (5.27) with $k\geq 4$ can be bounded similarly using the differentiation rule (5.28) and the bound (5.32) with much less efforts, since we gain an additional $n^{-1/2}$ from the moment condition in (2.1), so we omit the details. Therefore, we conclude

$$\Big|\frac{\mathrm{d}\boldsymbol{E}|R_t^{xy}|^{2m}}{\mathrm{d}t}\Big| \lesssim\Big(1+\frac{1}{\sqrt{n}\eta^{1+\frac{\epsilon}{2}}}\Big)\Big(\boldsymbol{E}\big|R_t^{xy}\big|^{2m}+O_\prec\big(\mathcal{E}_{b'}\big)^{2m}\Big). \tag{5.43}$$

This proves the first inequality in (5.15).

The proof of the second inequality (5.16) is similar to (5.15), actually with less efforts, so we only sketch the modifications. We perform similar arguments on S_t^{xy} as in (5.26)-(5.43). The corresponding differentiation rule in (5.28) is then replaced with

$$\begin{aligned}\frac{\partial S_t^{xy}}{\partial h_{aB}} = &-\big((G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1A_1G_1A_2G_2)_{\boldsymbol{e}_B\boldsymbol{y}} + (G_1A_1G_2)_{\boldsymbol{x}\boldsymbol{e}_a}(G_2A_2G_1)_{\boldsymbol{e}_B\boldsymbol{y}} \\ &+ (G_1A_1G_2AG_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1)_{\boldsymbol{e}_B\boldsymbol{y}}\big).\end{aligned} \tag{5.44}$$

Similarly to (5.32), we obtain from the assumption (5.5) and the deterministic bound (3.22) that

$$\sum_{\mathrm{j}=1}^{2n}\big|(G_1A_1G_2A_2G_1)_{\boldsymbol{u}\boldsymbol{e}_{\mathrm{j}}}\big|^k \prec n\widetilde{\mathcal{E}}_b^k + \frac{1}{(\eta\gamma)^k} + \frac{\sqrt{n}}{\eta\gamma}\mathbb{1}_{k=1}. \tag{5.45}$$

Moreover, we have the analogue of (5.35), *i.e.*, for $b' = \min\big\{b+\theta,\ 1\big\}$ with $0<\theta\le\epsilon/10$,

$$\frac{\widetilde{\mathcal{E}}_b}{\widetilde{\mathcal{E}}_{b'}} = \Big(\frac{\gamma}{\eta}\Big)^{b'-b} \lesssim \Big(\frac{1}{\rho^2\eta}\Big)^{\theta}, \qquad \frac{1}{\eta\gamma\mathcal{E}_{b'}} \lesssim 1. \tag{5.46}$$

Using (5.32), (5.44)-(5.46), and repeating similar arguments above, we obtain the analog of (5.43)

$$\frac{\mathrm{d}\boldsymbol{E}\big|\mathcal{S}_t^{xy}\big|^{2m}}{\mathrm{d}t} \lesssim \Big(1+\frac{1}{\sqrt{n}\eta^{1+\frac{\epsilon}{2}}}\Big)\Big(\boldsymbol{E}\big|\mathcal{S}_t^{xy}\big|^{2m} + O_\prec\big(\widetilde{\mathcal{E}}_{b'}\big)^{2m}\Big). \tag{5.47}$$

This proves the second inequality (5.16) and completes the proof of Lemma 5.4. $\square$

5.2. Proof of the averaged local law in (3.26)

The proof of the averaged local law in (3.26) is similar to the proof of (3.25) in the previous subsection, with the difference that now we additionally use (3.25) and Lemma 5.3 as technical inputs. Note that we focus on the regime that $\gamma \ge \eta_*/\rho^*$, while for the complementary regime the desired local law follows directly from (5.2).

We first establish a similar differentiation inequality as Lemma 5.4. Define, for brevity, the notations

$$R_t := \langle\big(G_1A_1G_2 - M_{12}^{A_1}\big)A_2\rangle, \qquad G_i = G_t^{z_i}(\mathrm{i}\eta_i) = (H_t^{z_i} - \mathrm{i}\eta_i)^{-1}, \qquad i=1,2. \tag{5.48}$$

For any $|z_i| \le 1+\tau$ and $\eta_i\in\mathbb{R}$ satisfying $\ell = \min_{i=1}^2\{|\eta_i|\rho_i\} \ge n^{-1+\epsilon}$ and $\gamma \gtrsim \eta_*/\rho^*$, we claim

$$\frac{\mathrm{d}\boldsymbol{E}|R_t|^{2m}}{\mathrm{d}t} \lesssim \Big(1+\frac{1}{\sqrt{n}\eta_*}\Big)\left(\boldsymbol{E}\big|R_t\big|^{2m} + O_\prec\Big(\frac{1}{\sqrt{n\ell}\gamma}\Big)^{2m}\right). \tag{5.49}$$

Assume (5.49) is proved. Then the desired local law (3.26) follows from using similar zig-zag arguments as in the proof of Proposition 5.2 above. The only difference is to replace Proposition 4.4 with Proposition 4.3, and replace the global law in (5.18) with its averaged version from [29, Theorem 5.2], *i.e.*,

$$\Big|\big\langle\big(G^{z_1}(\mathrm{i}\eta_1)A_1G^{z_2}(\mathrm{i}\eta_2)-M_{12}^{A_1}\big)A_2\big\rangle\Big|\prec\frac{1}{n},\qquad \operatorname{dist}(\mathrm{i}\eta,\operatorname{supp}(\rho^z))\sim 1. \tag{5.50}$$

This finishes the proof of (3.26), hence the proof of Theorem 3.4, modulo the differentiation inequality in (5.49) that we will prove below. □

Proof of (5.49). For notational simplicity, we assume $|\eta_1|\sim|\eta_2|\sim\eta$ and $\rho_1\sim\rho_2\sim\rho$. The proof for general η_i and ρ_i is essentially the same with η, ρ replaced by η_*, ρ^* in the following. We apply Itô's formula to $|R_t|^{2m}$ and perform the cumulant expansions as in (5.26)-(5.27). As explained below (5.27), for notational brevity, we will use ∂ to indicate either $\partial/\partial h_{aB}$ or $\partial/\partial h_{Ba}$, and with a slight abuse of notation R_t indicates either the complex number R_t or $\overline{R_t}$. By a direct computation, we obtain that

$$\frac{\partial R_t}{\partial h_{aB}}=-\frac{1}{n}\Big((G_1A_1G_2A_2G_1)_{Ba}+(G_2A_2G_1A_1G_2)_{Ba}\Big)=O_\prec\Big(\frac{1}{n\eta\gamma}\Big), \tag{5.51}$$

where we also used (3.22) and (5.8) in the last step. In general for $s\ge 2$, each partial derivative $\partial^s R_t$ contains $s+2$ Green functions. The resulting terms have two forms: 1) a product of $s-1$ Green function entries $\prod_{i=1}^{s-1}(G)_{\boldsymbol{u}_i,\boldsymbol{v}_i}$ and one factor $(G_1A_1G_2A_2G_1)_{\boldsymbol{u}_s\boldsymbol{v}_s}$; 2) a product of $s-2$ Green function entries $\prod_{i=1}^{s-2}(G)_{\boldsymbol{u}_i,\boldsymbol{v}_i}$, one factor $(G_1A_1G_2)_{\boldsymbol{u}_{s-1}\boldsymbol{v}_{s-1}}$, and another factor $(G_2A_2G_1)_{\boldsymbol{u}_s\boldsymbol{v}_s}$, with $\boldsymbol{u}_i,\boldsymbol{v}_i$ $(1\le i\le s)$ chosen from $\{\boldsymbol{e}_a,\boldsymbol{e}_B\}$. Note that both choices of $\boldsymbol{e}_a$ and $\boldsymbol{e}_B$ appear exactly s times. For instance, we list below several examples of ∂^2R_t and ∂^3R_t, *i.e.*,

$$\partial^2R_t:\qquad \frac{1}{n}(G_1A_1G_2A_2G_1)_{aa}(G_1)_{BB},\qquad \frac{1}{n}(G_1A_1G_2)_{aa}(G_2A_2G_1)_{BB}; \tag{5.52}$$

$$\partial^3R_t:\qquad \frac{1}{n}(G_1A_1G_2A_2G_1)_{aB}(G_1)_{aa}(G_1)_{BB},\qquad \frac{1}{n}(G_1A_1G_2A_2G_1)_{aa}(G_1)_{BB}(G_1)_{aB}, \tag{5.53}$$

$$\frac{1}{n}(G_1A_1G_2)_{aa}(G_2A_2G_1)_{BB}(G_1)_{aB},\qquad \frac{1}{n}(G_1A_1G_2)_{aB}(G_2A_2G_1)_{aa}(G_1)_{BB}. \tag{5.54}$$

Note that from the 1G local law (3.11) and (3.3), we have

$$\max_{a=1}^{n}\max_{B=n}^{2n}\Big\{|G_{aa}|,|G_{BB}|,|G_{aB}|,|G_{Ba}|\Big\}\lesssim\rho+\mathbb{1}_{B=a+n}, \tag{5.55}$$

for $n\eta\rho\gtrsim n^\epsilon$. Further using 2G local law (3.25) and (3.20), we have

$$\max_{1\leq \mathrm{i},\mathrm{j}\leq 2n}\Big\{\Big|(G_1A_1G_2)_{\mathrm{ij}}\Big|,\Big|(G_2A_2G_1)_{\mathrm{ij}}\Big|\Big\}\prec\frac{1}{\gamma}+\frac{1}{\sqrt{n\gamma\eta}},\tag{5.56}$$

while the 3G bound (5.8) and (3.22) imply that

$$\max_{1\leq \mathrm{i},\mathrm{j}\leq 2n}\Big\{\Big|(G_1A_1G_2A_2G_1)_{\mathrm{ij}}\Big|,\Big|(G_2A_2G_1A_1G_2)_{\mathrm{ij}}\Big|\Big\}\prec\frac{1}{\eta\gamma}.\tag{5.57}$$

Therefore, we obtain that

$$\Big|\frac{\partial^s R_t}{\partial h_{aB}^{s_1}\partial h_{Ba}^{s_2}}\Big|\prec\frac{\rho^{s-1}}{n\eta\gamma}+\frac{\rho^{s-2}}{n\gamma^2}+\frac{\rho^{s-2}}{n^2\gamma\eta^2}\lesssim\frac{\rho^s}{n\eta\rho\gamma},\qquad s=s_1+s_2\geq 1,\tag{5.58}$$

where we also used that $\gamma\gtrsim\frac{\eta}{\rho}$ and $n\eta\rho\gtrsim n^{\epsilon}$. Thus using (5.58), the k-th order term $\mathcal{T}_k^{(s_1,\cdots,s_l)}$ with $s_1+\cdots+s_l=k$ given as in (5.27) is bounded by

$$\begin{aligned}\big|\mathcal{T}_k^{(s_1,\cdots,s_l)}\big|\prec&\frac{\rho^k}{n^{\frac{k-4}{2}}}\Big(\frac{1}{n\eta\rho\gamma}\Big)^l\boldsymbol{E}\big|R_t\big|^{2m-l}\\ \lesssim&\frac{\rho^k}{n^{\frac{k-4}{2}}}\Big(\frac{1}{\sqrt{n\eta\rho}}\Big)^l\Big(\boldsymbol{E}\big|R_t\big|^{2m}+O_\prec\Big(\frac{1}{\sqrt{n\eta\rho}\gamma}\Big)^{2m}\Big).\end{aligned}\tag{5.59}$$

Hence for any $k\geq 4$, we obtain that

$$\big|\mathcal{T}_k^{(s_1,\cdots,s_l)}\big|\lesssim\boldsymbol{E}\big|R_t\big|^{2m}+O_\prec\Big(\frac{1}{\sqrt{n\eta\rho}\gamma}\Big)^{2m},\tag{5.60}$$

where we also used $n\eta\rho\gg 1$ and $\rho\lesssim 1$. For the third order terms with $k=3$ and $l\geq 2$, we also have

$$\big|\mathcal{T}_k^{(s_1,\cdots,s_l)}\big|\lesssim\frac{\rho^2}{\sqrt{n}\eta}\Big(\boldsymbol{E}\big|R_t\big|^{2m}+O_\prec\Big(\frac{1}{\sqrt{n\eta\rho}\gamma}\Big)^{2m}\Big).\tag{5.61}$$

We next focus on the only remaining term $\mathcal{T}_3^{(3)}$ with $k=3$ and $l=1$. Recall the third derivative terms in (5.53). By a direct computation, $\mathcal{T}_3^{(3)}$ is bounded by

$$\begin{aligned}\big|\mathcal{T}_3^{(3)}\big|=&\frac{\rho^2}{n^{5/2}}\sum_{a,B}\boldsymbol{E}\Big[\Big(\big|(G_1AG_2A^*G_1)_{Ba}\big|+\big|(G_2AG_1A^*G_2)_{Ba}\big|\Big)\big|R_t\big|^{2m-1}\Big]\\&+O_\prec\Big(\frac{\rho^{3/2}}{n\eta^{3/2}\gamma}\boldsymbol{E}\big|R_t\big|^{2m-1}\Big)+O_\prec\Big(\frac{\rho}{n\gamma^2}\boldsymbol{E}\big|R_t\big|^{2m-1}\Big).\end{aligned}\tag{5.62}$$

To bound the error terms in the last line above, we used the estimates (5.55)-(5.57), together with that $|G_{aB}|\prec\sqrt{\frac{\rho}{n\eta}}+\mathbb{1}_{a+n=B}$. Using the Cauchy-Schwarz inequality and Ward identity, the leading term on first line above is bounded by

$$\frac{\rho^2}{n^{5/2}} \sum_{a,B} \boldsymbol{E}\Big[\Big(\big|(G_1 A G_2 A^* G_1)_{Ba}\big| + \big|(G_2 A G_1 A^* G_2)_{Ba}\big|\Big)\big|R_t\big|^{2m-1}\Big]$$

$$\lesssim \frac{\rho^2}{n^{3/2}} \boldsymbol{E}\Big[\sqrt{\frac{\langle \operatorname{Im} G_1 A G_2 A^* \operatorname{Im} G_1 A G_2^* A^*\rangle}{\eta_1^2}}\big|R_t\big|^{2m-1}\Big] \prec \frac{\rho^2}{n^{3/2}\eta^2\sqrt{\gamma}} \boldsymbol{E}\big|R_t\big|^{2m-1}, \quad (5.63)$$

where in the last step, we also used that

$$\langle G_1 A G_2 A^* G_1 A G_2 A^*\rangle = \frac{1}{n}\sum_{\mathrm{i},\mathrm{j}=1}^{2n} (G_1 A G_2 A^*)_{\mathrm{ij}} (G_1 A G_2 A^*)_{\mathrm{ji}} \prec \frac{1}{\eta^2\gamma} + \frac{1}{\gamma^2} \lesssim \frac{1}{\eta^2\gamma}, \quad (5.64)$$

which follows from (3.25), (5.31) for $k = 2$, and that $\gamma \gtrsim \frac{\eta}{\rho} \gtrsim \eta$. Therefore we obtain from (5.62)

$$\begin{aligned} \big|\mathcal{T}_3^{(3)}\big| &\prec \Big(\frac{\rho^2}{n^{3/2}\eta^2\sqrt{\gamma}} + \frac{\rho^{3/2}}{n\eta^{3/2}\gamma} + \frac{\rho}{n\gamma^2}\Big)\boldsymbol{E}\big|R_t\big|^{2m-1} \\ &\lesssim \Big(1 + \frac{1}{\sqrt{n\eta}}\Big)\left(\boldsymbol{E}\big|R_t\big|^{2m} + O_\prec\Big(\frac{1}{\sqrt{n\eta\rho\gamma}}\Big)^{2m}\right), \end{aligned} \quad (5.65)$$

where we also used that $\frac{\eta}{\rho} \lesssim \gamma \lesssim \rho^{-2}$ from (3.23) and $n\eta\rho \gtrsim n^\epsilon$. Combining (5.65) with (5.60) and (5.61), we have proved the differentiation inequality in (5.49). □

5.3. Proof of Proposition 5.1 without $|z_1 - z_2|$-decorrelation

In this part, we aim to prove the local laws for $G^{z_1}(\mathrm{i}\eta_1) A G^{z_2}(\mathrm{i}\eta_2)$ without gaining from the $|z_1 - z_2|$-decorrelations. The proof is similar to the proofs in the previous subsections with less efforts, so we only sketch it.

Proof of Proposition 5.1. Note that the averaged local law (5.1) has already been proved in [29, Theorem 5.2]. Moreover, the last statement (5.3) follows directly from using the bound (3.22) and that

$$\Big|\big(G^{z_1}(\mathrm{i}\eta_1) A_1 G^{z_2}(\mathrm{i}\eta_2) A_2 G^{z_1}(\mathrm{i}\eta_1)\big)_{\boldsymbol{xy}}\Big| \lesssim \sqrt{\frac{\big(G^{z_1} A_1 \operatorname{Im} G^{z_2} A_1^* (G^{z_1})^*\big)_{\boldsymbol{xx}} \big(\operatorname{Im} G^{z_1}\big)_{\boldsymbol{yy}}}{\eta_1\eta_2}} \lesssim \frac{1}{\eta_*^2} \quad (5.66)$$

by Schwarz inequality and Ward identity. So we will only consider the isotropic local law (5.2). As explained in the proof of Proposition 5.2, it suffices to derive a similar differentiation inequality as in Lemma 5.4, *i.e.*,

$$\frac{\mathrm{d}\boldsymbol{E}|R_t^{\boldsymbol{xy}}|^{2m}}{\mathrm{d}t} \lesssim \Big(1 + \frac{1}{\sqrt{n\eta_*}}\Big)\left(\boldsymbol{E}\big|R_t^{\boldsymbol{xy}}\big|^{2m} + O_\prec\Big(\frac{\sqrt{\rho}}{\sqrt{n}\eta^{3/2}}\Big)^{2m}\right). \quad (5.67)$$

The proof uses the same strategy as Lemma 5.4. Actually it is much easier since we do not want to gain from $|z_1 - z_2|$-decay. Again we also assume $|\eta_1| \sim |\eta_2| \sim \eta$ and $\rho_1 \sim \rho_2 \sim \rho$ to simplify the notations.

We repeat the same cumulant expansions as in (5.26)-(5.27). Recall the differentiation rules in (5.28)-(5.30) and the third order terms in (5.33). From [35, Lemma D.2] we know

$$\big|(G_1A_1G_2)_{\boldsymbol{uv}}\big| \prec \frac{\rho}{\eta}, \qquad \big|(G_1A_1G_2A_2G_1)_{\boldsymbol{uv}}\big| \prec \frac{\rho}{\eta^2}, \tag{5.68}$$

for any deterministic unit vectors $\boldsymbol{u}, \boldsymbol{v}$. Then (5.32) is now replaced with the following

$$\sum_{\mathfrak{j}=1}^{2n} \big|(G)_{\boldsymbol{u}\boldsymbol{e}_{\mathfrak{j}}}\big|^k \prec n\Big(\sqrt{\frac{\rho}{n\eta}}\Big)^k + 1 + \sqrt{n}\mathbb{1}_{k=1}, \qquad \sum_{\mathfrak{i}=1}^{2n} \Big|(G_1AG_2)_{\boldsymbol{e}_{\mathfrak{i}}\boldsymbol{v}}\Big|^k \prec \frac{\rho^{k-1}}{\eta^{k+1}} + \sqrt{\frac{n\rho}{\eta^3}}\mathbb{1}_{k=1}, \tag{5.69}$$

where the second inequality follows from (5.68) and the Ward identity.

Using (5.69) for $k=3$, the first term $\mathcal{T}_3^{(1,1,1)}$ in (5.33) is bounded by

$$\begin{aligned} \big|\mathcal{T}_3^{(1,1,1)}\big| \lesssim & \frac{1}{n^{3/2}} \sum_{a,B} \boldsymbol{E}\Big[\Big(\Big|(G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\Big|^3 + \Big|(G_1AG_2)_{\boldsymbol{x}\boldsymbol{e}_a}G_{\boldsymbol{e}_B\boldsymbol{y}}\Big|^3\Big)\big|R_t^{xy}\big|^{2m-3}\Big] \\ \prec & \Big(\frac{\rho^{7/2}}{n^2\eta^{11/2}} + \frac{\rho^2}{n^{3/2}\eta^4}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-3}. \end{aligned} \tag{5.70}$$

For the second term $\mathcal{T}_3^{(1,2)}$ in (5.33), it then suffices to check the worst term in (5.36), *i.e.*,

$$\begin{aligned} (*) =& \frac{1}{n^{3/2}} \sum_{a,B} \boldsymbol{E}\Big[(G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}(G_1)_{\boldsymbol{x}\boldsymbol{e}_B}(G_1 - M_1)_{aa}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\big|R_t^{xy}\big|^{2m-2}\Big] \\ &+ O_\prec\Big(\frac{\rho^2}{n\eta^3}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-2} \prec \Big(\frac{\rho^2}{n^{3/2}\eta^4} + \frac{\rho^2}{n\eta^3}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-2}, \end{aligned} \tag{5.71}$$

where we used the first isotropic bound in (5.40), (5.68), and (5.69). For the last term $\mathcal{T}_3^{(3)}$ in (5.33), we focus on the worst term as discussed in (5.39). Using (5.68), (5.69) and the first isotropic bound in (5.40), such term is bounded by

$$\begin{aligned} (**) =& \frac{1}{n^{3/2}} \sum_{a,B} \boldsymbol{E}\Big[(G_1)_{\boldsymbol{x}\boldsymbol{e}_a}(G_1 - M_1)_{aa}(G_1)_{BB}(G_1AG_2)_{\boldsymbol{e}_B\boldsymbol{y}}\big(R_t^{xy}\big)^{2m-1}\Big] \\ &+ O_\prec\Big(\frac{\rho^{5/2}}{\sqrt{n}\eta^{3/2}}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-1} \prec \Big(\frac{\rho^{3/2}}{n\eta^{5/2}} + \frac{\rho^{5/2}}{\sqrt{n}\eta^{3/2}}\Big)\boldsymbol{E}\big|R_t^{xy}\big|^{2m-1}. \end{aligned} \tag{5.72}$$

Suming up and using Young's inequality, all the third order terms in (5.33) are bounded by

$$|\mathcal{T}_3| \lesssim \Big(1+\frac{1}{\sqrt{n\eta}}\Big)\left(\boldsymbol{E}\big|R_t^{xy}\big|^{2m}+O_\prec\Big(\frac{\sqrt{\rho}}{\sqrt{n}\eta^{3/2}}\Big)^{2m}\right). \tag{5.73}$$

Moreover, all the higher order terms ($k \geq 4$) in (5.27) can be bounded similarly with less efforts, since we gain additional $n^{-1/2}$ from the moment condition in (2.1). Hence we conclude with (5.67) and finish the proof of Proposition 5.1. □

5.4. Proof of Theorem 3.5: local law for $\operatorname{Im} G^{z_1} A \operatorname{Im} G^{z_2}$

In this section, we will prove Theorem 3.5, showing an improvement close to the edge of the spectrum when the resolvent G is replaced by $\operatorname{Im} G$, using Proposition 4.8 and Theorem 3.4, Lemma 5.3 as technical inputs.

Proof of Theorem 3.5. From (3.10) we know that $\rho_i \gtrsim \eta_i^{1/3}$ for $|z_i| \leq 1+n^{-1/2+\tau}$ and $\ell := \min_{i=1}^2\{|\eta_i|\rho_i\} \gtrsim n^{-1+\epsilon}$ with $0<\tau<\epsilon/10$. This further implies that $\rho_* = \min_{i=1}^2 \rho_i \gtrsim n^{-1/4-\epsilon/4}$. To prove Theorem 3.5, we will focus on the regime $|z_1-z_2|^2 \gtrsim \rho_1\eta_1+\rho_2\eta_2$, which also implies that $\gamma \gtrsim \frac{\eta_*}{\rho^*}$ from (3.23). For the complementary regime $|z_1-z_2|^2 \lesssim \rho_1\eta_1+\rho_2\eta_2$, the desired local law follows directly from Proposition 5.1 without gaining from the $|z_1-z_2|$-decay.

We define the following short-hand notations

$$\widehat{R_t} := \langle(\operatorname{Im} G_1 A_1 \operatorname{Im} G_2 - \widehat{M}_{12}^A)A_2\rangle, \qquad G_i = G_t^{z_i}(\mathrm{i}\eta_i) = (G_t^{z_i}-\mathrm{i}\eta_i)^{-1}, \quad i=1,2, \tag{5.74}$$

with the time dependent matrix H_t^z defined in (5.13). Then we claim the analogue of Lemma 5.4, *i.e.*,

$$\frac{\mathrm{d}\boldsymbol{E}|\widehat{R_t}|^{2m}}{\mathrm{d}t} = C_n'\left(\boldsymbol{E}\big|\widehat{R_t}\big|^{2m}+O_\prec\Big(\frac{1}{\sqrt{n\ell}}\frac{\rho_1\rho_2}{\widehat{\gamma}}\Big)^{2m}\right), \qquad C_n' := 1+\frac{1}{\sqrt{n\eta_*\rho_*}}, \tag{5.75}$$

for any $|z_i| \leq 1+\tau$, $n\ell \gtrsim n^\epsilon$, $\gamma \gtrsim \frac{\eta_*}{\rho^*}$, and $|\rho_i| \gtrsim n^{-c_*}$ with $0 \leq c_* \leq 1/4-\epsilon/4$.

We will prove (5.75) below. Assuming it now, similarly to the zig-zag arguments used in the proof of Proposition 5.2, we use (5.75) and Proposition 4.8 iteratively to prove (3.27). In the first step, for any fixed $|z_i^{(1)}| \leq 1+\tau$, $|\eta_i^{(1)}| \gtrsim \eta_*^{(1)} := n^{-\frac14}$ with $\rho_i^{(1)} \sim \rho_i$, we choose a small constant $T^{(1)} \sim 1$ so that $\operatorname{dist}(\mathrm{i}\eta_i(0), \operatorname{supp}(\rho_i(0))) \gtrsim 1$ from Lemma 4.2. Then the global law (5.50) and Proposition 4.8 yield the local law (3.27) at time $t=T^{(1)}$. Using (5.75), $\rho_* \gtrsim n^{-1/4}$, and similar arguments as in (5.21), the Gaussian component $T^{(1)} \sim 1$ can be removed, *i.e.*,

$$\big|\widehat{R_0}\big| \prec \frac{1}{\sqrt{n\ell}}\frac{\rho_1\rho_2}{\widehat{\gamma}}, \qquad |\eta_i| \gtrsim \eta_*^{(1)} = n^{-\frac14}. \tag{5.76}$$

In general for any $k \geq 1$, we define the scales recursively by

$$\eta_*^{(k)} := n^{-a_k}, \qquad a_k = a_{k-1}+\frac12-2c_*, \qquad a_1 = \frac14, \tag{5.77}$$

recalling that $\rho_* = n^{-c_*}$ with $0 \le c_* \le 1/4-\epsilon/4$. Clearly a_k is increasing and $a_k - a_{k-1} \ge \epsilon/2$. Replacing Proposition 4.4 with Proposition 4.8, the k-th iteration step is similar to the k-th step explained in the proof of Proposition 5.2. Using the new scales defined in (5.77), we obtain the analogue of (5.25), *i.e.*,

$$e^{\int_0^{T^{(k)}} C_n' \mathrm{d}t} \lesssim T^{(k)} C_n' \lesssim n^{-a_{k-1}+c_*}\Big(1 + \frac{1}{n^{1/2-a_k-c_*}})\Big) \lesssim 1, \tag{5.78}$$

where we used that $0 \le c_* \le 1/4 - \epsilon/4$. We stop the iterations when $\eta_*^{(k)}$ reaches $\eta_* = \min_{i=1}^2 |\eta_i| \gtrsim n^{-1+\epsilon}$. In the worst case when z_1 is in the bulk and z_2 is at the edge with $c_* = n^{-1/4-\epsilon/4}$ and $\eta_* = n^{-1+\epsilon}$, the number of iterations needed is at most $O(1/\epsilon)$. This hence proves Theorem 3.5 modulo the proof of (5.75) that we will do below. □

Proof of (5.75). We apply Itô's formula to $|\widehat{R_t}|^{2m}$ and perform the cumulant expansions as in (5.26). Note that from (5.8) we have

$$\frac{\partial \widehat{R_t}}{\partial_{aB}} = -\frac{1}{2n\mathrm{i}}\Big((G_1A_1\mathrm{Im}\,G_2A_2G_1)_{Ba} - (G_1^*A_1\mathrm{Im}\,G_2A_2G_1^*)_{Ba} \\ + (G_2A_2\mathrm{Im}\,G_1A_1G_2)_{Ba} - (G_2^*A_2\mathrm{Im}\,G_1A_1G_2^*)_{Ba}\Big) = O_\prec\Big(\frac{1}{n\eta_*\gamma}\Big). \tag{5.79}$$

We remark that, even though some of $\mathrm{Im}\,G_i$ are preserved after taking derivatives, we do not need to gain extra smallness from $\mathrm{Im}\,G_i$ for these isotropic terms. In general for any $s \ge 1$, the partial derivative $\partial^s \widehat{R_t}$ has a similar form as described in (5.51)-(5.53), with certain G's replaced with $\mathrm{Im}\,G$'s. Thus the same bound as in (5.58) also applies to $\partial^s \widehat{R_t}$, *i.e.*,

$$\Big|\frac{\partial^s \widehat{R_t}}{\partial h_{aB}^{s_1}\partial h_{Ba}^{s_2}}\Big| \prec \frac{(\rho^*)^{s-1}}{n\eta_*\gamma}, \qquad s = s_1 + s_2 \ge 1, \tag{5.80}$$

where we also used $n\ell \gtrsim n^\epsilon$ and $\gamma \gtrsim \frac{\eta_*}{\rho^*}$. Using (5.80), the corresponding k-th order cumulant expansion term as in (5.27), denoted by $\widehat{\mathcal{T}}_k^{(s_1,\cdots,s_l)}$ with $s_1 + \cdots + s_l = k$, is bounded by

$$\big|\widehat{\mathcal{T}}_k^{(s_1,\cdots,s_l)}\big| \lesssim \frac{(\rho^*)^k}{n^{\frac{k-4}{2}}}\Big(\frac{1}{n\eta_*\rho^*\gamma}\Big)^l \boldsymbol{E}\big|\widehat{R_t}\big|^{2m-l} \\ \lesssim \frac{1}{n^{\frac{k-4}{2}}(\rho_*)^l}\Big(\frac{1}{\sqrt{n\ell}}\Big)^l\Big(\boldsymbol{E}\big|\widehat{R_t}\big|^{2m} + \Big(\frac{1}{\sqrt{n\ell}}\frac{\rho_1\rho_2}{\widehat{\gamma}}\Big)^{2m}\Big), \tag{5.81}$$

where we also used that $n\eta_*\rho^* \ge n\ell \gtrsim n^\epsilon$, $\rho^* \lesssim 1$, and $k \ge l$. Using that $\rho_* \gtrsim n^{-1/4}$, the above naive estimates are already good enough to prove (5.75), for the terms with $5 \le k \le 7$ and $l \le 2k-8$, as well as all the high order terms with $k \ge 8$. For the remaining terms, we need the finer estimates below.

If there exists at least one s_i such that $s_i = 1$, then, using (5.79), (5.80), the Cauchy-Schwarz inequality and the Ward identity, we obtain

$$\begin{aligned}\left|\widehat{\mathcal{T}}_k^{(s_1,\cdots,s_l)}\right| \prec & \frac{(\rho^*)^{k-1}}{n^{\frac{k+2}{2}}}\Big(\frac{1}{n\eta_*\rho^*\gamma}\Big)^{l-1}\sum_{a,B}\boldsymbol{E}\Big[\Big(\big|(G_1A_1\operatorname{Im}G_2A_2G_1)_{Ba}\big| \\ & + \big|(G_2A_1\operatorname{Im}G_1A_2G_2)_{Ba}\big|\Big)\big|\widehat{R}_t\big|^{2m-l}\Big] \\ \le & \frac{(\rho^*)^{k-1}}{n^{\frac{k-1}{2}}\eta_*}\Big(\frac{1}{n\eta_*\rho^*\gamma}\Big)^{l-1}\boldsymbol{E}\Big[\sqrt{\langle\operatorname{Im}G_1A_1\operatorname{Im}G_2A_2\operatorname{Im}G_1A_2^*\operatorname{Im}G_2A_1^*\rangle}\big|\widehat{R}_t\big|^{2m-l}\Big] \\ \lesssim & \frac{(\rho^*)^{k-1}}{n^{\frac{k-2}{2}}\eta_*}\Big(\frac{1}{n\eta_*\rho^*\gamma}\Big)^{l-1}\boldsymbol{E}\Big[\langle\operatorname{Im}G_1A_1\operatorname{Im}G_2A_2\rangle|\widehat{R}_t|^{2m-l}\Big],\end{aligned} \tag{5.82}$$

where in the last step we also used the following inequality (from Eq. (4.59))

$$\langle\operatorname{Im}G_1A_1\operatorname{Im}G_2A_2\operatorname{Im}G_1A_2^*\operatorname{Im}G_2A_1^*\rangle \le n\langle\operatorname{Im}G_1A_1\operatorname{Im}G_2A_2\rangle^2. \tag{5.83}$$

Recall from (5.74) and (3.21) that

$$|\langle\operatorname{Im}G_1A_1\operatorname{Im}G_2A_2\rangle| \prec |\widehat{R}_t| + \frac{\rho_1\rho_2}{\widehat{\gamma}}.$$

Thus we conclude that, if there exists at least one $s_i = 1$, then

$$\begin{aligned}\left|\widehat{\mathcal{T}}_k^{(s_1,\cdots,s_l)}\right| \prec & \frac{(\rho^*)^{k-1}}{n^{\frac{k-2}{2}}\eta_*}\Big(\frac{1}{n\eta_*\rho^*\gamma}\Big)^{l-1}\Big(\boldsymbol{E}|\widehat{R}_t|^{2m-l+1} + \frac{\rho_1\rho_2}{\widehat{\gamma}}\boldsymbol{E}|\widehat{R}_t|^{2m-l}\Big) \\ \lesssim & \frac{1}{n^{\frac{k-2}{2}}\eta_*(\rho_*)^{l-1}}\Big(\frac{1}{\sqrt{n\ell}}\Big)^{l-1}\left(\boldsymbol{E}\big|\widehat{R}_t\big|^{2m} + \Big(\frac{1}{\sqrt{n\ell}}\frac{\rho_1\rho_2}{\widehat{\gamma}}\Big)^{2m}\right).\end{aligned} \tag{5.84}$$

We remark that if there exist at least two indices such that $s_i = s_j = 1$, then using the same strategy, we gain a bit more, *i.e.*,

$$\begin{aligned}\left|\widehat{\mathcal{T}}_k^{(s_1,\cdots,s_l)}\right| \le & \frac{(\rho^*)^{k-2}}{n^{\frac{k}{2}}(\eta_*)^2}\Big(\frac{1}{n\eta_*\rho^*\gamma}\Big)^{l-2}\boldsymbol{E}\Big[\langle\operatorname{Im}G_1A_1\operatorname{Im}G_2A_2\rangle^2|\widehat{R}_t|^{2m-l}\Big] \\ \lesssim & \frac{1}{n^{\frac{k}{2}}(\eta_*)^2(\rho_*)^{l-2}}\Big(\frac{1}{\sqrt{n\ell}}\Big)^{l-2}\left(\boldsymbol{E}\big|\widehat{R}_t\big|^{2m} + O_\prec\Big(\frac{1}{\sqrt{n\ell}}\frac{\rho_1\rho_2}{\widehat{\gamma}}\Big)^{2m}\right).\end{aligned} \tag{5.85}$$

Using that $\rho_* \gtrsim n^{-1/4}$, it is straightforward to check that the above finer bounds (5.84) and (5.85) are good enough to prove (5.75), for terms with $5 \le k \le 7$, $l \le 2k-8$, as well as the third and fourth order terms: $\widehat{\mathcal{T}}_3^{(1,1,1)}$, $\widehat{\mathcal{T}}_3^{(1,2)}$, $\widehat{\mathcal{T}}_4^{(1,1,1,1)}$, $\widehat{\mathcal{T}}_4^{(1,1,2)}$, $\widehat{\mathcal{T}}_4^{(1,3)}$. It then suffices to check the remaining terms, *i.e.*, $\widehat{\mathcal{T}}_3^{(3)}$, $\widehat{\mathcal{T}}_4^{(4)}$, $\widehat{\mathcal{T}}_4^{(2,2)}$.

We start with the third order term $\widehat{\mathcal{T}}_3^{(3)}$ with $k = 3$ and $l = 1$. Recall the third derivative terms as in (5.53). Except for the terms with $(G_1A_1\operatorname{Im}G_2A_2G_1)_{aB}$ that can be bounded similarly using (5.84), the most critical and representative terms are given by

$$T_I := \frac{1}{n^{5/2}} \sum_{a,B} \boldsymbol{E}\Big[(G_1A_1G_2A_2G_1)_{aa}(G_1)_{BB}(G_1)_{aB}\big(\widehat{R_t}\big)^{2m-1}\Big], \tag{5.86}$$

$$T_{II} := \frac{1}{n^{5/2}} \sum_{a,B} \boldsymbol{E}\Big[(G_1A_1G_2)_{aa}(G_2A_2G_1)_{BB}(G_1)_{aB}\big(\widehat{R_t}\big)^{2m-1}\Big], \tag{5.87}$$

and all the other terms can be bounded similarly. Using the isotropic local law (3.11) and (5.31),

$$\begin{aligned}
|T_I| &\prec \frac{1}{n^{5/2}} \sum_{a,B} \boldsymbol{E}\Big[(G_1A_1G_2A_2G_1 - M_{121})_{aa}(G_1 - M_1)_{BB}(G_1 - M_1)_{aB}\big(\widehat{R_t}\big)^{2m-1}\Big] \\
&\quad + \frac{\rho^*}{n\eta_*\gamma}\boldsymbol{E}\big|\widehat{R_t}\big|^{2m-1} \\
&\prec \Big(\frac{\rho^*}{n^{3/2}(\eta_*)^2\gamma} + \frac{\rho^*}{n\eta_*\gamma}\Big)\boldsymbol{E}\big|\widehat{R_t}\big|^{2m-1}; \qquad (5.88)\\
|T_{II}| &\prec \frac{1}{n^{5/2}} \sum_{a,B} \boldsymbol{E}\Big[(G_1A_1G_2 - M_{12})_{aa}(G_2A_2G_1 - M_{12})_{BB}(G_1 - M_1)_{aB}\big(\widehat{R_t}\big)^{2m-1}\Big] \\
&\quad + \frac{1}{(n\gamma)^{3/2}\eta_*}\boldsymbol{E}\big|\widehat{R_t}\big|^{2m-1} \\
&\prec \left(\frac{\sqrt{\rho^*}}{n^2(\eta_*)^{5/2}\gamma} + \frac{1}{(n\gamma)^{3/2}\eta_*}\right)\boldsymbol{E}\big|\widehat{R_t}\big|^{2m-1}, \qquad (5.89)
\end{aligned}$$

where we also used (3.20), (3.22) and (5.31) to bound the M-terms. Therefore, we obtain that

$$\begin{aligned}
\big|\widehat{\mathcal{T}}_3^{(3)}\big| &\prec \Big(\frac{\rho^*}{n^{3/2}(\eta_*)^2\gamma} + \frac{\rho^*}{n\eta_*\gamma} + \frac{\sqrt{\rho^*}}{n^2(\eta_*)^{5/2}\gamma} + \frac{1}{(n\gamma)^{3/2}\eta_*}\Big)\boldsymbol{E}\big|\widehat{R_t}\big|^{2m-1} \\
&\lesssim \frac{1}{\sqrt{n\ell}}\Big(1 + \frac{1}{\sqrt{n}\eta_*\rho_*}\Big)\left(\boldsymbol{E}\big|\widehat{R_t}\big|^{2m} + O_\prec\Big(\frac{1}{\sqrt{n\ell}}\frac{\rho_1\rho_2}{\widehat{\gamma}}\Big)^{2m}\right),
\end{aligned} \tag{5.90}$$

using that $\widehat{\gamma}/\gamma \lesssim 1$, $\gamma \gtrsim \frac{\eta_*}{\rho^*}$ and $n\ell \gg 1$. One could obtain a similar bound for $\widehat{\mathcal{T}}_4^{(4)}$ and $\widehat{\mathcal{T}}_4^{(2,2)}$ using (3.25) and (5.8) easily, since we gain additional $n^{-1/2}$ from the fourth order cumulants. This finishes the proof of (5.75). $\square$

Appendix A. Proof of several lemmas and estimates in Section 4

We start with proving Lemma 3.3, whose proof relies on the following Lemma A.1. Recall the structure of the M matrix from (3.3) and the notations (1.10). We use $m_j =$

$m^{z_j}(\mathrm{i}\eta_j)$ and $u_j = u^{z_j}(\mathrm{i}\eta_j)$, $j = 1, 2$. It was proven in [30, Appendix B] that the two–body stability operator $\mathcal{B}_{12}$ defined in (3.17) has two non-trivial (small) eigenvalues

$$\beta_\pm = \beta_\pm(z_1, \eta_1, z_2, \eta_2) := 1 - u_1 u_2 \mathrm{Re}\,[z_1\overline{z_2}] \pm \sqrt{s}, \qquad s := m_1^2 m_2^2 - u_1^2 u_2^2 (\mathrm{Im}\,[z_1\overline{z_2}])^2, \tag{A.1}$$

with the standard complex square root taken in the branch cut $\mathbb{C} \setminus (-\infty, 0]$, while the remaining eigenvalues are trivially equal to one. The corresponding (right) eigenvectors of $\beta\pm$ are given by

$$\begin{aligned} &\mathcal{B}_{12}[R_\pm] = \beta_\pm R_\pm, \\ &R_\pm := \begin{pmatrix} -u_1 u_2 \mathrm{Re}\, z_1 \bar{z}_2 \pm \sqrt{s} & z_1 u_1 m_2 + \frac{m_1^2 z_2 u_2 m_2}{\mathrm{i} u_1 u_2 \mathrm{Im}\, z_1 \bar{z}_2 \mp \sqrt{s}} \\ \bar{z}_2 u_2 m_1 + \frac{m_2^2 \bar{z}_1 u_1 m_1}{\mathrm{i} u_1 u_2 \mathrm{Im}\, z_1 \bar{z}_2 \mp \sqrt{s}} & \frac{m_1 m_2}{\mathrm{i} u_1 u_2 \mathrm{Im}\, z_1 \bar{z}_2 \mp \sqrt{s}} (-u_1 u_2 \mathrm{Re}\, z_1 \bar{z}_2 \pm \sqrt{s}) \end{pmatrix}, \end{aligned} \tag{A.2}$$

and $\mathcal{B}_{12}[F^{(*)}] = F^{(*)}$. Note that $\mathcal{B}_{12}$ is not self-adjoint, and its adjoint operator is given by

$$\mathcal{B}_{12}^* := 1 - \mathcal{S}[(M^{z_1})^* \cdot (M^{z_2})^*], \tag{A.3}$$

with corresponding (left) eigenvectors given by

$$\mathcal{B}_{12}^*[L_\pm^*] = \overline{\beta_\pm} L_\pm^*, \qquad L_\pm := \begin{pmatrix} \frac{\mathrm{i} u_1 u_2 \mathrm{Im}\, z_1 \bar{z}_2 \mp \sqrt{s}}{m_1 m_2} & 0 \\ 0 & 1 \end{pmatrix}. \tag{A.4}$$

Lemma A.1. *Fix any $|z_i| < 10$ and $0 < |\eta_i| < 10$ $(i = 1, 2)$. On the imaginary axis, both the product $\beta_+\beta_-$ and sum $\beta_+ + \beta_-$ are positive and*

$$0 < \beta_+\beta_- \sim \widehat{\gamma} = |z_1 - z_2|^2 + \rho_1|\eta_1| + \rho_2|\eta_2| + \left(\frac{\eta_1}{\rho_1}\right)^2 + \left(\frac{\eta_2}{\rho_2}\right)^2, \tag{A.5}$$

$$0 < \beta_+ + \beta_- \sim |z_1 - z_2|^2 + \rho_1^2 + \rho_2^2 + \frac{|\eta_1|}{\rho_1} + \frac{|\eta_2|}{\rho_2}. \tag{A.6}$$

Furthermore, $\beta_+ + \beta_- - \beta_+\beta_-$ is also positive and

$$0 < \beta_+ + \beta_- - \beta_+\beta_- \sim \rho_1^2 + \rho_2^2 + \frac{\eta_1}{\rho_1} + \frac{\eta_2}{\rho_2}. \tag{A.7}$$

Moreover, $\beta_+ - \beta_-$ is either real-valued or purely imaginary and

$$0 < \mathrm{Re}\,(\beta_+ - \beta_-) \lesssim \rho_1\rho_2, \qquad 0 < \mathrm{Im}\,(\beta_+ - \beta_-) \lesssim |z_1 - z_2|. \tag{A.8}$$

Finally, we have

$$\min\{|\beta_\pm|\} \gtrsim \gamma = \frac{|z_1 - z_2|^2 + \rho_1|\eta_1| + \rho_2|\eta_2| + \left(\frac{\eta_1}{\rho_1}\right)^2 + \left(\frac{\eta_2}{\rho_2}\right)^2}{|z_1 - z_2| + \rho_1^2 + \frac{|\eta_1|}{\rho_1} + \rho_2^2 + \frac{|\eta_2|}{\rho_2}}. \tag{A.9}$$

Proof of Lemma A.1. We start with proving the upper bound of $\beta_+\beta_-$ in (A.5). Using that $2\mathrm{Re}\,(z_1\overline{z_2}) = |z_1|^2+|z_2|^2-|z_1-z_2|^2$ and that $m^2 = -u+u^2|z|^2$ from (3.4), we have

$$\begin{aligned}\beta_+\beta_- =&1-2u_1u_2\mathrm{Re}\,(z_1\overline{z_2})+u_1^2u_2^2|z_1|^2|z_2|^2-m_1^2m_2^2\\ =&1-u_1u_2\big(|z_1|^2+|z_2|^2-|z_1-z_2|^2\big)+u_1^2u_2^2|z_1|^2|z_2|^2-m_1^2m_2^2\\ =&1-u_1u_2\big(1+|z_1|^2(1-u_1)+|z_2|^2(1-u_2)-|z_1-z_2|^2\big)\\ =&u_1u_2|z_1-z_2|^2-m_1^2\frac{u_2}{u_1}(1-u_1)-m_2^2\frac{u_1}{u_2}(1-u_2)+(1-u_1)(1-u_2).\end{aligned} \tag{A.10}$$

From (3.10) we know $\rho \gtrsim |\eta|$ and hence

$$1-u=\frac{|\eta|}{|\eta|+|m|}\sim\frac{|\eta|}{\rho}.$$

Thus we conclude that

$$\beta_+\beta_-\sim|z_1-z_2|^2+\rho_1|\eta_1|+\rho_2|\eta_2|+\Big|\frac{\eta_1\eta_2}{\rho_1\rho_2}\Big|. \tag{A.11}$$

This proves the upper bound in (A.5) for any z by a simple Cauchy-Schwarz inequality. To obtain the comparable lower bound in (A.5), we split the discussion into two cases. For $|z|<1$, it follows directly from (A.11) and that $\rho^3\gtrsim|\eta|$ from (3.10). For the complementary regime $|z|>1$, since $\rho^3\lesssim|\eta|$, the desired lower bound in (A.5) was already obtained in Appendix A (Lemma A.1) of the gumbel paper.

We continue to prove the remaining estimates. Similarly to (A.10) we have

$$\begin{aligned}\beta_++\beta_-=2-2u_1u_2\mathrm{Re}\,[z_1\overline{z_2}] =&u_1u_2|z_1-z_2|^2-\frac{u_2}{u_1}m_1^2-\frac{u_1}{u_2}m_2^2+1-u_1+1-u_2\\ \sim&|z_1-z_2|^2+\rho_1^2+\frac{|\eta_1|}{\rho_1}+\rho_2^2+\frac{|\eta_2|}{\rho_2}.\end{aligned} \tag{A.12}$$

This proves (A.6). Combining (A.10) with (A.12), we have

$$\beta_++\beta_--\beta_+\beta_-=1-u_1u_2-u_2m_1^2-u_1m_2^2\sim\rho_1^2+\frac{|\eta_1|}{\rho_1}+\rho_2^2+\frac{|\eta_2|}{\rho_2},$$

which proves (A.7). Similarly, we have $\beta_+-\beta_-=2\sqrt{m_1^2m_2^2-u_1^2u_2^2(\mathrm{Im}\,[z_1\overline{z_2}])^2}$ in the standard branch cut $\mathbb{C}\setminus(-\infty,0]$. Note that $m_1^2m_2^2-u_1^2u_2^2(\mathrm{Im}\,[z_1\overline{z_2}])^2$ is real valued, which implies that $\beta_+-\beta_-$ is either real or purely imaginary. More precisely, we have

$$\begin{aligned}\mathrm{Re}\,(\beta_+-\beta_-) =&2\mathrm{Re}\,\Big(\sqrt{m_1^2m_2^2-u_1^2u_2^2(\mathrm{Im}\,[z_1\overline{z_2}])^2}\Big)\le 2\rho_1\rho_2,\\ \mathrm{Im}\,(\beta_+-\beta_-) =&2\mathrm{Im}\,\Big(\sqrt{m_1^2m_2^2-u_1^2u_2^2(\mathrm{Im}\,[z_1\overline{z_2}])^2}\Big)\le 2u_1u_2|\mathrm{Im}\,[z_1\overline{z_2}]|\lesssim|z_1-z_2|,\end{aligned} \tag{A.13}$$

which prove (A.8).

Finally it still remains to prove (A.9). Combining (A.7) with (A.5), we obtain that

$$\frac{1}{\beta_-}+\frac{1}{\beta_+}-1=\frac{\beta_++\beta_--\beta_+\beta_-}{\beta_+\beta_-}\lesssim\frac{\rho_1^2+\frac{|\eta_1|}{\rho_1}+\rho_2^2+\frac{|\eta_2|}{\rho_2}}{|z_1-z_2|^2+\rho_1|\eta_1|+\rho_2|\eta_2|+\left(\frac{\eta_1}{\rho_1}\right)^2+\left(\frac{\eta_2}{\rho_2}\right)^2}. \tag{A.14}$$

Similarly, combining (A.8) with (A.5), we obtain that

$$\left|\frac{1}{\beta_-}-\frac{1}{\beta_+}\right|=\left|\frac{\beta_+-\beta_-}{\beta_+\beta_-}\right|\lesssim\frac{\max\{\rho_1\rho_2,|z_1-z_2|\}}{|z_1-z_2|^2+\rho_1|\eta_1|+\rho_2|\eta_2|+\left(\frac{\eta_1}{\rho_1}\right)^2+\left(\frac{\eta_2}{\rho_2}\right)^2}. \tag{A.15}$$

Therefore, we obtain that

$$\left|\frac{1}{\beta_\pm}\right|\lesssim\left|\frac{1}{\beta_-}+\frac{1}{\beta_+}\right|+\left|\frac{1}{\beta_-}-\frac{1}{\beta_+}\right|\lesssim\frac{|z_1-z_2|+\rho_1^2+\frac{|\eta_1|}{\rho_1}+\rho_2^2+\frac{|\eta_2|}{\rho_2}}{|z_1-z_2|^2+\rho_1|\eta_1|+\rho_2|\eta_2|+\left(\frac{\eta_1}{\rho_1}\right)^2+\left(\frac{\eta_2}{\rho_2}\right)^2}, \tag{A.16}$$

which finishes the proof of Lemma A.1. □

Proof of Lemma 3.3. Recall from (3.16) that

$$M_{12}^A=\mathcal{B}_{12}^{-1}\Big[M^{z_1}(\mathrm{i}\eta_1)AM^{z_2}(\mathrm{i}\eta_2)\Big]. \tag{A.17}$$

Then (3.20) directly follows from (A.9), *i.e.*,

$$\|M_{12}^A\|\lesssim\|\mathcal{B}_{12}^{-1}\|\le\frac{1}{\min\{|\beta_\pm|\}}\lesssim\frac{1}{\gamma}. \tag{A.18}$$

Moreover, $\widehat{M}_{12}^A$ is a fourfold linear combination of $M_{12}^A(\pm\eta_1,z_1,\pm\eta_2,z_2)$ using the identity $\operatorname{Im}G=\frac{1}{2\mathrm{i}}(G-G^*)$ twice. To prove (3.21), it suffices to show that

$$\left|\left\langle\mathcal{B}_{12}^{-1}\Big[M^{z_1}(\mathrm{i}\eta_1)A_1M^{z_2}(\mathrm{i}\eta_2)\Big]A_2\right\rangle-\left\langle\mathcal{B}_{12}^{-1}\Big[M^{z_1}(\mathrm{i}\eta_1)A_1M^{z_2}(-\mathrm{i}\eta_2)\Big]A_2\right\rangle\right|\lesssim\frac{\rho_1\rho_2}{\widehat{\gamma}}. \tag{A.19}$$

We will focus on the worst case when A_1,A_2 are along the bad directions $L_\pm$. Note that $L_\pm$ are linear combinations of $E_\pm$. It then reduces to prove (A.19) for $A_1,A_2\in\{E_+,E_-\}$. By a direct computation from (A.4), we obtain

$$\left(\mathcal{B}_{12}^*\right)^{-1}[E_\pm]=\begin{pmatrix}\frac{1-z_1u_1\bar{z}_2u_2\pm m_1m_2}{\beta_-\beta_+} & 0\\ 0 & \frac{m_1m_2\pm(1-\bar{z}_1u_1z_2u_2)}{\beta_-\beta_+}\end{pmatrix}. \tag{A.20}$$

Hence using (A.20) and (3.3), we obtain that

$$\begin{aligned}\left\langle \mathcal{B}_{12}^{-1}\left[M^{z_1}(\mathrm{i}\eta_1)E_\pm M^{z_2}(\mathrm{i}\eta_2)\right]E_\pm\right\rangle &= \left\langle \left(\mathcal{B}_{12}^*\right)^{-1}[E_\pm], M^{z_1}(\mathrm{i}\eta_1)E_\pm M^{z_2}(\mathrm{i}\eta_2)\right\rangle \\ &= \pm\,\frac{m_1^2m_2^2 \pm m_1m_2 + \mathrm{Re}\,[z_1\overline{z_2}]u_1u_2 - |z_1|^2|z_2|^2u_1^2u_2^2}{\beta_+\beta_-};\end{aligned} \tag{A.21}$$

$$\begin{aligned}\left\langle \mathcal{B}_{12}^{-1}\left[M^{z_1}(\mathrm{i}\eta_1)E_- M^{z_2}(\mathrm{i}\eta_2)\right]E_+\right\rangle &= -\left\langle \mathcal{B}_{12}^{-1}\left[M^{z_1}(\mathrm{i}\eta_1)E_+ M^{z_2}(\mathrm{i}\eta_2)\right]E_-\right\rangle \\ &= \frac{-\mathrm{i}\mathrm{Im}\,[z_1\overline{z_2}]u_1u_2}{\beta_+\beta_-}.\end{aligned} \tag{A.22}$$

Using that $u(-\mathrm{i}\eta) = u(\mathrm{i}\eta)$ and $m(-\mathrm{i}\eta) = -m(\mathrm{i}\eta)$, we obtain the following cancellation

$$\left|\left\langle \mathcal{B}_{12}^{-1}\left[M^{z_1}(\mathrm{i}\eta_1)A_1M^{z_2}(\mathrm{i}\eta_2)\right]A_2\right\rangle - \left\langle \mathcal{B}_{12}^{-1}\left[M^{z_1}(\mathrm{i}\eta_1)A_1M^{z_2}(-\mathrm{i}\eta_2)\right]A_2\right\rangle\right| \lesssim \frac{\rho_1\rho_2}{\beta_+\beta_-} \lesssim \frac{\rho_1\rho_2}{\widehat{\gamma}}, \tag{A.23}$$

for any $A_1, A_2 \in \{E_+, E_-\}$, which proves (3.21).

We next prove the last statement (3.22). Recall from (3.19) and (3.6) that

$$M_{121}^{A_1,A_2} = \mathcal{B}_{11}^{-1}\left[M_1A_1M_{21}^{A_2} + M_1\Big(\langle M_{12}^{A_1}E_+\rangle E_+ - \langle M_{12}^{A_1}E_-\rangle E_-\Big)M_{21}^{A_2}\right]. \tag{A.24}$$

From [29, Eq. (3.8)], we know $\mathcal{B}_{11}^*$ has two small eigenvalues

$$\beta_* := \beta_-(z,\eta,z,\eta) \sim \frac{\eta}{\mathrm{Im}\, m}, \qquad \beta := \beta_+(z,\eta,z,\eta) \sim \frac{\eta}{\mathrm{Im}\, m} + (\mathrm{Im}\, m)^2, \tag{A.25}$$

with the left eigenvectors given by E_- and E_+, respectively. It then suffices to test the above matrix against these two directions $\{E_+, E_-\}$, *i.e.*,

$$\begin{aligned}&\left\langle M_{121}^{A_1,A_2}E_\pm\right\rangle \\ &= \frac{\langle M_{12}^{A_1}E_+\rangle\langle E_\pm M_1E_+M_{21}^{A_2}\rangle - \langle M_{12}^{A_1}E_-\rangle\langle E_\pm M_1E_-M_{21}^{A_2}\rangle + \langle E_\pm M_1A_1M_{21}^{A_2}\rangle}{\beta_\pm(z_1,\eta_1,z_1,\eta_1)}.\end{aligned} \tag{A.26}$$

Note that using (A.25) and (A.18) the last term above is easily bounded by

$$\left|\frac{1}{\beta_{(*)}}\langle E_\pm M_1A_1M_{21}^{A_2}\rangle\right| \lesssim \frac{\rho_1}{\eta_1\gamma}. \tag{A.27}$$

For the remaining terms, using (A.17), we further write

$$\begin{aligned}&\langle M_{12}^{A_1}E_+\rangle\langle E_\pm M_1E_+M_{21}^{A_2}\rangle - \langle M_{12}^{A_1}E_-\rangle\langle E_\pm M_1E_-M_{21}^{A_2}\rangle \\ =&\big\langle(\mathcal{B}_{12}^*)^{-1}[E_+^*], M_1A_1M_2\big\rangle\big\langle(\mathcal{B}_{21}^*)^{-1}[E_+^*M_1^*E_\pm^*], M_2A_2M_1\big\rangle \\ &- \big\langle(\mathcal{B}_{12}^*)^{-1}[E_-^*], M_1A_1M_2\big\rangle\big\langle(\mathcal{B}_{21}^*)^{-1}[E_-^*M_1^*E_\pm^*], M_2A_2M_1\big\rangle.\end{aligned} \tag{A.28}$$

Using $M_1 = m_1E_+ - z_1u_1F - \overline{z_1}u_1F^*$, the matrices of the form $E_\pm^* M_1^* E_\pm^*$ can be written as linear combinations of $E_\pm$, F and F^*. Then we introduce the following estimates. To simplify the notations, we may assume $\rho \sim \rho_1 \sim \rho_2$ and $\eta \sim \eta_1 \sim \eta_2$. The same upper bound estimates below also apply to general η_i and ρ_i, with ρ and η replaced by ρ^* and η_*.

$$\left(\mathcal{B}_{12}^*\right)^{-1}[E_\pm] = \begin{pmatrix} \frac{1-z_1\bar{z}_2 \pm m_1m_2 + O(\eta/\rho)}{\beta_-\beta_+} & 0 \\ 0 & \frac{m_1m_2 \pm (1-\bar{z}_1z_2+O(\eta/\rho))}{\beta_-\beta_+} \end{pmatrix}, \tag{A.29}$$

$$\left(\mathcal{B}_{12}^*\right)^{-1}[F] = \begin{pmatrix} \frac{m_2(\overline{z_1}-\overline{z_2})+O(\eta)}{\beta_-\beta_+} & 1 \\ 0 & \frac{m_1(\overline{z_2}-\overline{z_1})+O(\eta)}{\beta_-\beta_+} \end{pmatrix}, \tag{A.30}$$

$$\left(\mathcal{B}_{12}^*\right)^{-1}[F^*] = \begin{pmatrix} \frac{m_1(z_2-z_1)+O(\eta)}{\beta_-\beta_+} & 0 \\ 1 & \frac{m_2(z_1-z_2)+O(\eta)}{\beta_-\beta_+} \end{pmatrix}, \tag{A.31}$$

which follow directly from (A.3)-(A.4) and that $u = 1 + O(\eta/\rho)$. By direct computations using (A.29)-(A.31) and (A.25), we obtain that

$$\begin{aligned} &\left| \frac{\langle M_{12}^{A_1}E_+\rangle\langle E_+M_1E_+M_{21}^{A_2}\rangle - \langle M_{12}^{A_1}E_-\rangle\langle E_+M_1E_-M_{21}^{A_2}\rangle}{\beta} \right| \lesssim \frac{1}{\rho^2}\frac{\rho^5+\rho|z_1-z_2|^2}{(\widehat{\gamma})^2} \lesssim \frac{\rho^2}{\eta\widehat{\gamma}}, \\ &\left| \frac{\langle M_{12}^{A_1}E_+\rangle\langle E_-M_1E_+M_{21}^{A_2}\rangle - \langle M_{12}^{A_1}E_-\rangle\langle E_-M_1E_-M_{21}^{A_2}\rangle}{\beta_*} \right| \lesssim \frac{\rho}{\eta}\frac{\rho|z_1-z_2|^2}{(\widehat{\gamma})^2} \lesssim \frac{\rho^2}{\eta\widehat{\gamma}}, \end{aligned} \tag{A.32}$$

for $|z| \le 1$, where we also used $\rho^3 \gtrsim \eta$, $\widehat{\gamma} = \beta_-\beta_+ \gtrsim |z_1-z_2|^2 + \rho\eta$, and

$$m^2 = |z|^2 - 1 + O(\rho\eta), \qquad 2\mathrm{Re}\,(z_1\overline{z_2}) = |z_1|^2 + |z_2|^2 - |z_1-z_2|^2, \qquad |\mathrm{Im}\,(z_1\overline{z_2})| \sim |z_1-z_2|.$$

Plugging the above estimates and (A.27) into (A.26) we obtain that

$$\left\langle M_{121}^{A_1,A_2}E_\pm \right\rangle \lesssim \frac{(\rho^*)^2}{\eta_*\widehat{\gamma}} + \frac{\rho_1}{\eta_1\gamma} \lesssim \frac{1}{\eta_*\gamma}. \tag{A.33}$$

It hence proves (3.22) for $|z| \le 1$. The proof for the complementary regime $1 < |z| \le 10$ is similar and actually easier using that $\rho^3 \le \eta$, so we omit the details. We hence finish the proof of Lemma 3.3. $\square$

Proof of (4.27) for $\eta_1\eta_2 < 0$. Note that for $|z_i| \le 1 - c$, from (A.1) and (A.4) we know that, $\beta_+ \sim 1$, $\beta_- \sim \gamma$, and L_- is the only bad direction. Recall from [30, Eq. (B.6)] that

$$E_\pm = \big(1 + O(|z_1-z_2|)\big)L_{\pm\sigma} + O(|z_1-z_2|)L_{\mp\sigma}, \qquad \sigma := \mathrm{sgn}(\eta_1\eta_2). \tag{A.34}$$

Decomposing the matrix A as $A = \langle AE_+\rangle E_+ + \langle AE_-\rangle E_- + A_c$, we obtain that

$$\begin{aligned}
\big|\langle M_{12}^A E_{-\sigma}\rangle\big| &\lesssim \big|\langle M_{12}^{E_+} E_{-\sigma}\rangle\big| + \big|\langle M_{12}^{E_-} E_{-\sigma}\rangle\big| + \big|\langle M_{12}^{A_c} E_{-\sigma}\rangle\big| \lesssim \big|\langle M_{12}^{E_{-\sigma}} E_{-\sigma}\rangle\big| + \frac{|z_1 - z_2|}{\gamma},\\
\big|\langle M_{12}^A E_{\sigma}\rangle\big| &\lesssim \big|\langle M_{12}^{E_+} E_{\sigma}\rangle\big| + \big|\langle M_{12}^{E_-} E_{\sigma}\rangle\big| + \big|\langle M_{12}^{A_c} E_{\sigma}\rangle\big| \lesssim \frac{|z_1 - z_2|}{\gamma}.
\end{aligned} \tag{A.35}$$

This proves (4.27) for $\eta_1\eta_2 < 0$. Similar upper bounds also hold for $\eta_1\eta_2 > 0$ with switched signs. □

Proof of Lemma 4.6. The first relation for M_{12} was proved in Eq. (A.26) in the Appendix A.5 of [30]. The second relation for M_{121} then follows by a so-called meta argument [36,32], using recursive Dyson equations for chains of product of resolvents, see e.g. [32, Lemma D.1]. □

Data availability

No data was used for the research described in the article.

References

[1] A. Adhikari, J. Huang, Dyson Brownian motion for general β and potential at the edge, Probab. Theory Relat. Fields 178 (2020) 893–950.
[2] I. Afanasiev, M. Shcherbina, T. Shcherbina, Universality of the second correlation function of the deformed Ginibre ensemble, arXiv:2405.00617, 2024.
[3] O.H. Ajanki, L. Erdős, T. Krüger, Universality for general Wigner-type matrices, Probab. Theory Relat. Fields 169 (2017) 667–727.
[4] O.H. Ajanki, L. Erdős, T. Krüger, Stability of the matrix Dyson equation and random matrices with correlations, Probab. Theory Relat. Fields 173 (2019) 293–373.
[5] G. Akemann, Y.P. Förster, M. Kieburg, Universal eigenvector correlations in quaternionic Ginibre ensembles, J. Phys. A 53 (14) (2020) 145201.
[6] G. Akemann, R. Tribe, A. Tsareas, O. Zaboronski, On the determinantal structure of conditional overlaps for the complex Ginibre ensemble, Random Matrices: Theory Appl. 9 (4) (2020) 2050015.
[7] G. Akemann, S.S. Byun, K. Noda, Pfaffian structure of the eigenvector overlap for the symplectic Ginibre ensemble, Ann. Henri Poincaré (2025).
[8] J. Alt, L. Erdős, T. Krüger, Local inhomogeneous circular law, Ann. Appl. Probab. 28 (1) (2018) 148–203.
[9] J. Alt, L. Erdős, T. Krüger, Spectral radius of random matrices with independent entries, Probab. Math. Phys. 2 (2) (2021) 221–280.
[10] J. Banks, J.G. Vargas, A. Kulkarni, N. Srivastava, Overlaps, eigenvalue gaps, and pseudospectrum under real Ginibre and absolutely continuous perturbations, Ann. Inst. Henri Poincaré Probab. Stat. 60 (4) (2024) 2736.
[11] J. Banks, A. Kulkarni, S. Mukherjee, N. Srivastava, Gaussian regularization of the pseudospectrum and Davies' conjecture, Commun. Pure Appl. Math. 74 (10) (2021) 2114–2131.
[12] S. Belinschi, M.A. Nowak, R. Speicher, W. Tarnowski, Squared eigenvalue condition numbers and eigenvector correlations from the single ring theorem, J. Phys. A 50 (10) (2017) 105204.
[13] F. Benaych-Georges, O. Zeitouni, Eigenvectors of non normal random matrices, Electron. Commun. Probab. 23 (70) (2018) 1–12.
[14] P. Bourgade, Extreme gaps between eigenvalues of Wigner matrices, J. Eur. Math. Soc. 24 (8) (2021) 2823–2873.
[15] P. Bourgade, G. Dubach, The distribution of overlaps between eigenvectors of Ginibre matrices, Probab. Theory Relat. Fields 177 (2020) 397–464.

[16] P. Bourgade, H.-T. Yau, J. Yin, Local circular law for random matrices, Probab. Theory Relat. Fields 159 (2014) 545–595.
[17] P. Bourgade, H.-T. Yau, J. Yin, The local circular law II: the edge case, Probab. Theory Relat. Fields 159 (2014) 619–660.
[18] P. Bourgade, G. Cipolloni, J. Huang, Fluctuations for non-Hermitian dynamics, arXiv:2409.02902, 2024.
[19] Z. Burda, J. Grela, M.A. Nowak, W. Tarnowski, P. Warchoł, Dysonian dynamics of the Ginibre ensemble, Phys. Rev. Lett. 113 (10) (2014) 104102.
[20] A. Campbell, G. Cipolloni, L. Erdős, H.C. Ji, On the spectral edge of non-Hermitian random matrices, arXiv:2404.17512, 2024, Accepted to Ann. Probab.
[21] J.T. Chalker, B. Mehlig, Eigenvector statistics in non-Hermitian random matrix ensembles, Phys. Rev. Lett. 81 (16) (1998) 3367.
[22] G. Cipolloni, J. Kudler-Flam, Entanglement entropy of non-Hermitian eigenstates and the Ginibre ensemble, Phys. Rev. Lett. 130 (1) (2023) 010401.
[23] G. Cipolloni, J. Kudler-Flam, Non-Hermitian Hamiltonians violate the eigenstate thermalization hypothesis, Phys. Rev. B 109 (2) (2024) L020201.
[24] G. Cipolloni, L. Erdős, D. Schröder, Edge universality for non-Hermitian random matrices, Probab. Theory Relat. Fields 179 (2021) 1–28.
[25] G. Cipolloni, L. Erdős, D. Schröder, Fluctuation around the circular law for random matrices with real entries, Electron. J. Probab. 26 (2021) 1–61.
[26] G. Cipolloni, L. Erdős, D. Schröder, Optimal multi-resolvent local laws for Wigner matrices, Electron. J. Probab. 27 (2022) 1–38.
[27] G. Cipolloni, L. Erdős, D. Schröder, On the condition number of the shifted real Ginibre ensemble, SIAM J. Matrix Anal. Appl. 43 (3) (2022) 1469–1487.
[28] G. Cipolloni, L. Erdős, J. Henheik, Eigenstate thermalisation at the edge for Wigner matrices, arXiv:2309.05488, 2023, Accepted to Ann. Inst. Henri Poincaré Probab. Stat.
[29] G. Cipolloni, L. Erdős, D. Schröder, Central limit theorem for linear eigenvalue statistics of non-Hermitian random matrices, Commun. Pure Appl. Math. 76 (2023) 946–1034.
[30] G. Cipolloni, L. Erdős, D. Schröder, Mesoscopic central limit theorem for non-Hermitian random matrices, Probab. Theory Relat. Fields 188 (2023) 1131–1182.
[31] G. Cipolloni, L. Erdős, J. Henheik, O. Kolupaiev, Eigenvector decorrelation for random matrices, arXiv:2410.10718, 2024.
[32] G. Cipolloni, L. Erdős, J. Henheik, D. Schröder, Optimal lower bound on eigenvector overlaps for non-Hermitian random matrices, J. Funct. Anal. 287 (4) (2024) 110495.
[33] G. Cipolloni, L. Erdős, H.C. Ji, Non-Hermitian spectral universality at critical points, Probab. Theory Relat. Fields (2025).
[34] G. Cipolloni, L. Erdős, Y. Xu, Precise asymptotics for the spectral radius of a large random matrix, J. Math. Phys. 65 (6) (2024) 063302.
[35] G. Cipolloni, L. Erdős, Y. Xu, Universality of extremal eigenvalues of large random matrices, arXiv: 2312.08325.
[36] N. Cook, W. Hachem, J. Najim, D. Renfrew, Non-Hermitian random matrices with a variance profile (I): deterministic equivalents and limiting ESDs, Electron. J. Probab. 23 (110) (2018).
[37] N. Crawford, R. Rosenthal, Eigenvector correlations in the complex Ginibre ensemble, Ann. Appl. Probab. 32 (4) (2022) 2706–2754.
[38] M.J. Crumpton, T.R. Würfel, Spectral density of complex eigenvalues and associated mean eigenvector self-overlaps at the edge of elliptic Ginibre ensembles, arXiv:2405.02103, 2024.
[39] M.J. Crumpton, Y.V. Fyodorov, T.R. Würfel, Mean eigenvector self-overlap in the real and complex elliptic Ginibre ensembles at strong and weak non-Hermiticity, Ann. Henri Poincaré (2025).
[40] G. Dubach, On eigenvector statistics in the spherical and truncated unitary ensembles, Electron. J. Probab. 26 (124) (2021).
[41] G. Dubach, Explicit formulas concerning eigenvectors of weakly non-unitary matrices, Electron. Commun. Probab. 28 (6) (2023).
[42] S. Dubova, K. Yang, Bulk universality for complex eigenvalues of real non-symmetric random matrices with iid entries, Ann. Henri Poincaré (2025).
[43] S. Dubova, K. Yang, H.T. Yau, J. Yin, Gaussian statistics for left and right eigenvectors of complex non-Hermitian matrices, arXiv:2403.19644, 2024.
[44] L. Erdős, H.C. Ji, Wegner estimate and upper bound on the eigenvalue condition number of non-Hermitian random matrices, Commun. Pure Appl. Math. 77 (9) (2024) 3785–3840.

[45] S. Esaki, M. Katori, S. Yabuoku, Eigenvalues, eigenvector-overlaps, and regularized Fuglede-Kadison determinant of the non-Hermitian matrix-valued Brownian motion, arXiv:2306.00300, 2023.
[46] Y.V. Fyodorov, On statistics of bi-orthogonal eigenvectors in real and complex Ginibre ensembles: combining partial Schur decomposition with supersymmetry, Commun. Math. Phys. 363 (2) (2018) 579–603.
[47] Y.V. Fyodorov, B. Mehlig, Statistics of resonances and nonorthogonal eigenfunctions in a model for single-channel chaotic scattering, Phys. Rev. E 66 (4) (2002) 045202.
[48] Y.V. Fyodorov, M. Osman, Eigenfunction non-orthogonality factors and the shape of CPA-like dips in a single-channel reflection from lossy chaotic cavities, J. Phys. A 55 (22) (2022) 224013.
[49] Y.V. Fyodorov, D.V. Savin, Statistics of resonance width shifts as a signature of eigenfunction nonorthogonality, Phys. Rev. Lett. 108 (18) (2012) 184101.
[50] Y.V. Fyodorov, W. Tarnowski, Condition numbers for real eigenvalues in the real elliptic Gaussian ensemble, Ann. Henri Poincaré 22 (1) (2021) 309–330.
[51] Y.V. Fyodorov, E. Gudowska-Nowak, M.A. Nowak, W. Tarnowski, Fluctuation-dissipation relation for non-Hermitian Langevin dynamics, Phys. Rev. Lett. 134 (2025) 087102.
[52] S. Ghosh, M. Kulkarni, S. Roy, Eigenvector correlations across the localization transition in non-Hermitian power-law banded random matrices, Phys. Rev. B 108 (6) (2023) L060201.
[53] V.L. Girko, Circular law, Teor. Veroâtn. Primen. 29 (1984) 669–679.
[54] J. Grela, P. Warchoł, Full Dysonian dynamics of the complex Ginibre ensemble, J. Phys. A 51 (42) (2018) 425203.
[55] Y. He, A. Knowles, Mesoscopic eigenvalue statistics of Wigner matrices, Ann. Appl. Probab. 27 (3) (2017) 1510–1550.
[56] J. Huang, B. Landon, Rigidity and a mesoscopic central limit theorem for Dyson Brownian motion for general and potentials, Probab. Theory Relat. Fields 175 (2019) 209–253.
[57] V. Jain, A. Sah, M. Sawhney, On the real Davies' conjecture, Ann. Probab. 49 (6) (2021) 3011–3031.
[58] R.A. Janik, W. Nörenberg, M.A. Nowak, G. Papp, I. Zahed, Correlations of eigenvectors for non-Hermitian random-matrix models, Phys. Rev. E 60 (3) (1999) 2699.
[59] D.Z. Liu, L. Zhang, Critical edge statistics for deformed GinUEs, arXiv:2311.13227, 2023.
[60] D.Z. Liu, L. Zhang, Repeated erfc statistics for deformed GinUEs, Ann. Henri Poincaré (2025).
[61] A. Maltsev, M. Osman, Bulk universality for complex non-Hermitian matrices with independent and identically distributed entries, Probab. Theory Relat. Fields (2024).
[62] Bernhard Mehlig, John T. Chalker, Statistical properties of eigenvectors in non-Hermitian Gaussian random matrix ensembles, J. Math. Phys. 41 (5) (2000) 3233–3256.
[63] S. Morimoto, M. Katori, T. Shirai, Generalized eigenspaces and pseudospectra of nonnormal and defective matrix-valued dynamical systems, arXiv:2411.06472, 2024.
[64] K. Noda, Determinantal structure of the overlaps for induced spherical unitary ensemble, Random Matrices: Theory Appl. 14 (3) (2025) 2550012.
[65] M. Osman, Bulk universality for real matrices with independent and identically distributed entries, arXiv preprint, Electron. J. Probab. 30 (2025) 1–66.
[66] M. Osman, Least non-zero singular value and the distribution of eigenvectors of non-Hermitian random matrices, J. Math. Phys. 66 (8) (2025).
[67] M. Osman, Universality for diagonal eigenvector overlaps of non-Hermitian random matrices, arXiv: 2409.16144, 2024.
[68] S.S. Roy, S. Bandyopadhyay, R.C. de Almeida, P. Hauke, Unveiling eigenstate thermalization for non-Hermitian systems, Phys. Rev. Lett. 134 (18) (2025) 180405.
[69] A. Sankar, D.A. Spielman, S. Teng, Smoothed analysis of the condition numbers and growth factors of matrices, SIAM J. Matrix Anal. Appl. 28 (2) (2006) 446–476.
[70] W. Tarnowski, Condition numbers for real eigenvalues of real elliptic ensemble: weak non-normality at the edge, J. Phys. A 57 (2024) 255204.
[71] K. Tikhomirov, Invertibility via distance for noncentered random matrices with continuous distributions, Random Struct. Algorithms 57 (2) (2020) 526–562.
[72] M. Walters, S. Starr, A note on mixed matrix moments for the complex Ginibre ensemble, J. Math. Phys. 56 (1) (2015) 013301.
[73] T.R. Würfel, M.J. Crumpton, Y.V. Fyodorov, Mean left-right eigenvector self-overlap in the real Ginibre ensemble, Random Matrices: Theory Appl. 13 (4) (2024) 2450017.
[74] J. Yin, The local circular law III: general case, Probab. Theory Relat. Fields 160 (2014) 679–732.
[75] L. Zhang, Bulk universality for deformed GinUEs, arXiv:2403.16120, 2024.
[76] L. Zhang, Mean eigenvector self-overlap in deformed complex Ginibre ensemble, arXiv:2407.09163, 2024.